DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics

Luca Stanca

No. 78 - July 2004

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
http://dipeco.economia.unimib.it

The effects of attendance on academic performance: panel data evidence for Introductory Microeconomics

Luca Stanca*
Economics Department
University of Milan Bicocca

July 2004

Abstract

This paper presents new evidence on the effects of attendance on academic performance. We exploit a large panel data set for Introductory Microeconomics students to explicitly take into account the effect of unobservable factors correlated with attendance, such as ability, effort and motivation. We find that neither proxy variables nor instrumental variables provide a viable solution to the omitted variable bias. Panel estimators indicate that attendance has a positive and significant impact on performance. Lecture and classes have a similar effect on performance individually, although their impact cannot be identified separately. Overall, the results indicate that, after controlling for unobservable student characteristics, teaching has an important independent effect on learning.

JEL Classification: A22, I21.

*University of Milan Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. Telephone: #39 2 6448 6655. E-mail: luca.stanca@unimib.it

1 Introduction

It is commonly assumed that university students benefit from attending lectures. This assumption, however, needs to be tested, as developments in information technology are increasingly calling for a reassessment of the traditional approach to university education, largely based on physical attendance of lectures and classes, and a number of alternative "weightless" educational models, based on distance learning, are being introduced. Nevertheless, as pointed out by Romer (1993), until recently there was relatively little evidence about attendance and its effects on student learning.[1]

In the past decade, a number of studies have examined the relationship between students' attendance (or absenteeism) and academic performance, generally finding that attendance does matter for academic achievement (see e.g. Durden and Ellis (1995), Devadoss and Foltz (1996), Chan et al. (1997), Marburger (2001) Rodgers (2001), Bratti and Staffolani (2002), Dolton et al. (2003), Kirby and McElroy (2003)). This kind of evidence has led some authors to call for measures to increase student attendance and even to consider the possibility of making attendance mandatory in some undergraduate courses.[2]

The main problem in assessing the effects of attendance on academic performance is that attendance levels are not exogenous, given that students *choose* whether to attend lectures and classes, and that this choice is affected by unobservable individual characteristics, such as ability, effort and motivation, that are also likely to determine performance: better students, who are more able, work harder or are more motivated, tend to have higher attendance levels, other things being equal. This implies that estimates of the impact of attendance on academic performance are likely to be subject to omitted variable bias.

Most existing studies either brush aside this problem or attempt to disentangle the impact of attendance on performance from unobservable ability and motivational factors by including in the set of regressors proxies of capability (students' grade-point-averages, scores on college entry exams, etc.), effort (homework-assignment completion) and motivation (students' self reported interest in the course). However, such indicators are generally an imperfect measure of ability and motivation. As a consequence, OLS esti-

[1] "Even though teaching is a very large part of what we do, we know very little about many aspects of instruction and learning" (Romer, 1993, p. 214).

[2] See Romer (1993) and the following discussion in Brauer et al. (1994).

mates of the returns to attendance obtained from specifications that include appropriate control variables are still likely to be biased and inconsistent, to the extent that they incorrectly attribute to attendance the effect of the component of ability and motivation not captured by the controls.

One possible solution would be to find appropriate instruments for attendance. However, it is generally quite difficult to find variables correlated with attendance but uncorrelated with unobservable ability, effort and motivation. An alternative route, followed in this paper, is to exploit the variability of attendance and performance in the time dimension, if a panel data set is available. This allows to take into account time-invariant unobservable factors that affect both attendance and performance, and therefore to eliminate the omitted variable bias that characterizes estimates of the effect of attendance on performance based on cross-sectional data.

For the analysis presented in this study, we collected observations on the performance of 766 Introductory Microeconomics students on several tests, and their attendance levels at lectures and classes covering the material examined on those tests. We also have information on proxies for ability (high school grade, grade point average, exam speed, and proficiency in calculus), effort (number of study hours) and motivation (subject and teacher evaluation), candidate instruments for attendance and a number of other individual characteristics. We can therefore compare the results obtained with three approaches: OLS controlling for unobservable factors with proxy variables; instrumental variables (2SLS) for attendance; panel estimators (random effects and fixed effects).

We find that both OLS and IV estimates of the effects of attendance on performance are positive and significant. However, neither proxy variables nor instrumental variables provide a viable solution to the omitted variable bias: proxy variables do not capture all the correlation between the regressor of interest and the omitted factors, while candidate instrumental variables are found to be correlated with the error term. When we eliminate the omitted variable bias, using a fixed effect estimator, the point estimate for attendance is about half the size of the OLS and IV estimates, but the effect on performance remains positive and significant. We also find that lecture and classes have a similar effect on performance individually, although their impact cannot be identified separately. Overall, the results indicate that teaching is a key factor for student learning.

The remainder of the paper is structured as follows. Section 2 reviews the empirical literature on student attendance and academic performance.

Sections 3 and 4 describe the data set and the econometric methodology, respectively. Section 5 presents the results of the empirical analysis. Section 6 concludes with a discussion of the main findings and the implications of the analysis.

2 Literature review

In a widely cited study, Romer (1993) reported evidence on absenteeism in undergraduate economics courses at three major US universities, finding an average attendance rate of about 67 per cent. The paper also presented regression results, based on a sample of 195 Intermediate Macroeconomics students, indicating a positive and significant relationship between student attendance and exam performance. This result was found to be qualitatively robust to the inclusion among the explanatory variables of students' grade point average and the fraction of problem sets completed.[3] On the basis of these findings, Romer suggested that measures aimed at increasing attendance, including making attendance mandatory, could be considered.[4]

Prior to Romer, Schmidt (1983) had investigated student time allocation in a sample of 216 macroeconomic principles students, finding that time spent in lectures and discussion sections has a positive and significant effect on exam performance, even after controlling for hours of study. Park and Kerr (1990) had found an inverse relationship between students' attendance and their course grades in a money and banking course over a four-year period, even after controlling for the effect of unobservable motivation by means of students' self-reported hours of study and their perceived value of the course.[5]

Following the controversial conclusions of Romer (1993), in the past decade a number of empirical studies in the economic education literature

[3]In order to control for the effects of motivation, Romer also examined the results obtained by restricting his sample to students who had completed all the problem sets assigned during the semester.

[4]"I believe that the results here both about the extent of absenteeism and its relation to performance are suggestive enough to warrant experimenting with making class attendance mandatory in some ungergraduate lecture courses." (Romer, 1993, p. 173).

[5]See McConnell and Lamphear (1969), Paden and Moyer (1969), Buckles and McMahon (1971), Browne, et al. (1991) for early studies finding no significant impact of attendance on academic performance. See also Siegfried and Fels (1979) for a comprehensive survey on research on teaching college economics.

have examined the relationship between student attendance and academic performance. Durden and Ellis (1995) investigate the link between overall course grade and self-reported attendance levels in a sample of 346 principles of economics students over three semesters. Their results, based on OLS controlling for ability and motivational factors (GPA, college-entrance exam scores, having had a course in calculus) indicate that attendance matters for academic performance. In particular, whereas low levels of absenteeism have little effect on the eventual outcome, excessive absenteeism has a large and significant effect.

Devadoss and Foltz (1996) examine attendance in a sample of about 400 agricultural economics students at four large U.S. universities. They find that, even after controlling for both prior grade point average and the degree of motivation, on average students who attended all classes achieved a full letter grade higher than students who attended no more than 50 per cent of the same classes. A positive and significant relationship between attendance and academic performance is also found by Chan et al. (1997) in a sample of 71 Principles of Finance students.

More recently, Marburger (2001) investigates the relationship between absenteeism and exam performance in a sample of 60 students of a principles of microeconomics course. In this study, information on student attendance at each class during the semester is matched with records of the class meeting when the material corresponding to each question was covered. The results indicate that students who miss class on a given date are significantly more likely to respond incorrectly to questions relating to material covered that day than students who were present. Rodgers (2001) finds that attendance has a small but statistically significant effect on performance in a sample of 167 introductory statistics course. Kirby and McElroy (2003) study the determinants of levels of attendance at lectures and classes and the relationship with exam performance in a sample of 368 first year economics students, finding that hours worked and travel time are the main determinants of class attendance, and that the latter, in turn, has a positive and diminishing marginal effect on grade.

Among studies who reach less robust conclusions about the positive effect of attendance on performance, Bratti and Staffolani (2002) argue that estimates of student performance regressions that omit study hours might be biased, given that hours of study are a significant determinant of lecture attendance. Using a sample of 371 first-year Economics students they find that the positive and significant effect of lecture attendance on performance is not robust to the inclusion of the number of hours of study. Dolton et al. (2003), applying stochastic frontier techniques to a large sample of Spanish students, find tha both formal study and self study are significant determinants of exam scores but that the former may be up to four times more important than the latter. However, they also find that self study time may be insignificant if ability bias is corrected for.

All of these studies, with the exception of Marburger (2001) and Rodgers (2001), are based on cross-sectional data sets. As a consequence, as observed by Romer (1993), the possibility that the estimated relationship between attendance and exam performance reflects the impact of omitted factors rather than a true effect cannot be ruled out. In the following we thus report results obtained using panel data on Introductory Microeconomics students to estimate the *net* effect of attendance on academic performance.

3 Data

The data for this study were collected by conducting a survey of 766 students attending the Introductory Microeconomic course at the University of Milan in the academic years 2001 to 2004.[6] The course, taught over twelve weeks in the spring semester, is taken by all students in the first-year of study of the Economics degree. The exam is based on four mid-term tests, administered every three weeks, covering equal fractions of the course and carrying the same weight for the final grade. Questionnaires were distributed to students with each of the four test papers, and compiled before starting the tests. This produced four independently pooled panels (one for every year), each with a cross-section of about two hundred students observed over four tests, resulting in a potential balanced panel of 3064 observations ($N = 766$ times $T = 4$). The number of observations for the actual (unbalanced) panel is 2913 ($\overline{T} = 3.8$), due to inconplete questionnaires and a number of students dropping out before the end of the course.[7]

Summary statistics for the main variables are reported in table 1. Academic performance is measured by students' test score (SCO). The actual

[6] There are 7 parallel sections of the Introductory Microeconomic course offered to about 1,500 first year students. All sections have the same content (syllabus and textbook), even though the number and format of intermediate tests may differ.

[7] Note that only the three main variables of interest (test score, lecture attendance and class attendance) are time varying, whereas all other variables are time invariant.

test score ranges potentially from -36 to + 36, as it results from summing up the outcome of 24 independent true/false questions, with 1.5 marks for correct answers, -1.5 marks for wrong answers, and 0 for no answer, so that random guessing implies a zero expected score. Test scores were then rescaled to the range (-100, +100), to make them more easily interpretable and comparable with the results reported in the literature. In our sample, the average rescaled test score is 58.7, with a range between -41.7 and 100.[8] The figures on lecture (LEC) and class (CLA) attendance are estimated percentages, as reported by the students, out of a variable number of two-hour lectures and classes for each of the three-week course units.[9] On average, the students in the sample attended about two-thirds of classes (67.4) and a slightly higher percentage of lectures (70.8).[10]

Ability is proxied by four main indicators, based on students' past performance in both high school and university: *high school grade* (HSG), ranging between 60 and 100, is the leaving certificate score; *grade point average* (GPA) is the average mark on exams passed before taking Introductory Microeconomics;[11] *exams per annum* (EPA), is a measure of speed in completing course work, defined as the number of exams passed divided by the years of registration; *calculus* is a dummy variable equal to 1 if the student has passed the first-year calculus exam.[12] The means of both GPA and EPA appear to be quite low (76.9 and 2.1, respectively). Effort is measured by the average number of *hours of study* per week (in addition to attendance) for Introductory Microeconomics (SSH), ranging between 1 and 35 around a mean value of 10.9. Motivation is measured by two indicators: *subject evaluation* (SEV) and *teacher evaluation* (TEV), self-reported assessments defined on a 0 to

[8] The pass mark for the rescaled score is 50, given that in the Italian university system the exam pass score is 18.

[9] There were generally 8 lectures and 3 classes for each course unit.

[10] The figures for lectures are very similar to those reported by Romer (1993) and Rodgers (2001), while substantially higher than those reported by Kirby and McElroy (2003). The higher attendance rate for classes relative to lectures is the opposite of what was found by both Rodgers (2001) and Kirby and McElroy (2003).

[11] The actual GPA, defined on a 18-30 scale, was rescaled to a 60-100 scale to make it more easily comparable.

[12] The interpretation of the last two variables might be considered difficult for first-year students, who represent the majority of the sample. However, they can be considered informative as they provide information on performance in the first semester for first-year students, and overall performance for the remaing group of students in higher year of registration.

100 scale (average values of 74.7 per cent and 80.9 per cent, respectively).

Additional quantitative variables include *travel time* to reach university (in minutes), *age*, and *year of registration*. A number of dummy variables provide information on student characteristics, such as gender (1=*female*), *foreign language* (1=non-native speaker), *work* (1=worked while taking the course), *web* (1=internet available at home), and *live away from home*. Further information on the background of students is provided by cathegorical variables referring to *high school type*, *education* and *occupation* for both father and mother, and *province of residence*.[13]

4 Methodology

We are interested in estimating the parameters characterizing the relationship between teaching and learning. We assume that learning is the output of an educational production function that reflects the match between two types of factors: academic input and student input.[14] Academic input broadly refers to teaching (lectures, classes, seminars, tutorials, office hours, etc.). Student input is assumed to reflect a number of individual factors, among which the three main ones are ability, effort, and motivation. Assuming linearity, the relationship can be described as

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{1}$$

where y_i is learning for individual i, x_1 is academic input, x_2 is student input, and ε_i is an error term reflecting all other factors that affect learning, with $i = 1,...,N$.

We measure learning by academic performance (test score) and teaching by lecture and class attendance. It is more difficult to find an appropriate measure for student input, given that factors such as ability, effort and motivation are not directly observable. This would not be a problem for the estimation of β_1 if student input and attendance were uncorrelated. However, ability, effort and motivation are all likely to be positively correlated

[13]Tables 6-10 in the data appendix provide descriptive statistics for quantitative indicators of performance, attendance, ability, effort and motivation, by sub-groups defined according to students' characteristics.

[14]See e.g. Pritchett and Filmer (1999), Lazear (2001), Todd and Wolpin (2003) and Coates (2003).

with attendance: students who are more able, work harder or are more motivated, tend to have higher attendance levels. As a consequence, the OLS estimator of β_1 in equation (1), omitting x_2, would be biased and inconsistent, as it would attribute to attendance an effect that is actually due to unobservable student characteristics. In short, we face a classic example of omitted variable bias.

One possible solution is to find appropriate *proxy variables* for student input. This implies estimating

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i}^* + \varepsilon_i \tag{2}$$

where we assume that $x_{2i} = \gamma_0 + \gamma_1 x_{2i}^* + \nu_i$ describes the relationship between the unobservable factors and the proxy variables. Note that, in order to obtain a consistent estimator for β_1, x_{1i} and ν_i must be uncorrelated: the proxy variables must capture all of the correlation between the unobserved factors (student input) and the regressor of interest (attendance). In the following we use *high school grade*, *grade point average*, *exams per annum* and *calculus* as proxies for ability, *hours of study* as a proxy for effort, and *subject and teacher evaluation* as proxies for motivation.[15]

If there are no proxy variables available, or the ones available are not suitable because they do not capture all the correlation between the regressor of interest and the omitted factors, an alternative solution is to find appropriate *instrumental variables* for attendance. The instruments would allow to net out the correlation of student input with attendance, so that $\widehat{\beta}_1^{IV}$ would measure its net effect on academic performance. Note, however, that the consistence of the IV estimator relies on the assumption of instrument validity, which is often difficult to maintain in practice. In addition, even if the assumption of instrument validity is satisfied, the instruments can be weakly related to the endogenous variables, resulting in imprecise estimates.

In the following we consider estimates of equation (2), with and without the inclusion of the proxy variables, obtained by two-stage least squares, using *travel time*, *work* and *web* as instruments for attendance.[16] The choice of the instruments is based on the assumption that longer travel time, being

[15]Note that subject and teacher assessment provide information about the "match" between academic and student inputs. They are therefore a measure of the suitability of the student for the subject, which is what we refer to by the term *motivation*.

[16]We report Davidson-McKinnon (1993) endogeneity tests of the null hypothesis that attendance is uncorrelated with the error term, so that OLS is a consistent estimator, under the maintained assumption that the IV estimator is consistent. We also report

a working student and having internet at home should be negatively related to attendance, while not having a direct impact on performance. Note, in particular, that in the case of work, this assumption can be maintained given that (in the complete specification) we are controlling for the number of hours of study.

A third possibility is to exploit the time dimension of the data set, assuming that the omitted variables do not change over time, to eliminate the effect of unobservable factors using a *panel estimator* in the following specification:

$$y_{it} = \beta_1 x_{1it} + \beta_2 x_{2i} + \alpha_i + \eta_{it} \tag{3}$$

where η_{it} is the idiosyncratic error component, i.i.d.$(0, \sigma_\eta^2)$, uncorrelated with $(x_{1it}, x_{2i}, \alpha_i)$, and α_i is i.i.d.$(0, \sigma_\alpha^2)$, potentially correlated with x_{1it} and x_{2i}.

The fixed effect (FE) estimator is based on the assumption that α_i is correlated with the explanatory variables (or that it represents fixed constants), and is obtained as OLS on the data transformed in deviations from individual means:

$$y_{it} - \overline{y}_{i.} = (x_{1it} - \overline{x}_{1i.})\, \beta_1 + (\eta_{it} - \overline{\eta}_{i.}) \tag{4}$$

This estimator is consistent even in the presence of unobservable effects correlated with the regressors, provided η_{it} and x_{1t} are uncorrelated at all leads and lags. However, the fixed effects estimator wipes out all time invariant regressors and is not efficient. We therefore also consider the random effects estimator (RE), based on quasi-deviations from individual means:

$$y_{it} - \theta\overline{y}_{i.} = (x_{1it} - \theta\overline{x}_{i.})\, \beta_1 + (x_{2i} - \theta\overline{x}_{2i.})\, \beta_2 + (\alpha_i - \theta\alpha_i) + (\eta_{it} - \theta\eta_{i.}) \tag{5}$$

where $\theta = 1 - \left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + T\sigma_\alpha^2} \right)^{\frac{1}{2}}$ is a measure of the weight of the between component in the total variability of the error term. This estimator is inconsistent in the presence of unobservable effects correlated with the regressors. However, it is efficient (as it is the GLS estimator) and it allows to estimate the parameters of time-invariant regressors.[17]

Sargan tests of overidentifying restrictions distributed as χ^2 with two degrees of freedom under the null hypothesis of instrument validity.

[17] We report Hausman tests of the null hypothesis that the individual-specific component of the error term (α_i) is uncorrelated with the regressors, based on the comparison of the estimates obtained for the fixed and random effects models. We also report Breusch-Pagan Lagrange Multiplier tests of the hypothesis of constant variance of the individual-specific component of the error term (α_i), i.e. a test of the pooled (OLS) model against the alternative of the random effect model.

All the specifications estimated below include time fixed effects captured by year-test specific dummies to allow for intercept heterogeneity in the sixteen cross sections (four tests in each of the four years): $\lambda_t = \sum_{j=1}^{4} \sum_{k=1}^{4} \delta_{jk} D_{jk}$, where D_{jk} is a time dummies for year j and test k, and δ_{jk} is the corresponding parameter. We also control for individual characteristics such as year of registration, gender, foreign language, and live away from home, and include sets of dummy variables for high school type, parental education and occupation, and province of origin.

5 Results

This section presents the estimation results. We start by estimating equation (1) by OLS, and examine the impact on the estimated coefficient for attendance of controlling for unobservable factors such as ability, effort and motivation. We then consider the results obtained with IV for the same set of specifications. Next, we present estimates obtained for panel data estimators (random effects and fixed effects). Finally, we examine the respective effects of lecture and class attendance on performance.

Table 2 reports OLS estimates of alternative specifications of the relationship between academic performance and attendance. All specifications produce a coefficient estimate for attendance that is positive and statistically significant at the one per cent level. In the basic univariate specification (column 1), the point estimate indicates that one additional percentage point of lecture attendance corresponds to a 0.09 percent improvement in performance. As reported in column 2, the addition of a set of controls for individual characteristics does not affect the estimated coefficient for attendance. In this specification, *year of enrollment* and *live away from home* are negatively and significantly associated to performance.

Next, we consider how controlling for unobservable factors, such as ability, effort and motivation, affects the estimated coefficient for attendance. Adding either the set of ability proxies (column 3) or the set of effort and motivation indicators (column 4), the estimated coefficient for lecture attendance falls to 0.076 and 0.087, respectively.[18] Adding both sets of indi-

[18]Note that, to the extent that GPA reflects the effect of attendance in other courses, and that attendance is positively correlated across courses, the inclusion of this variable could lead to underestimate the effect of attendance on performance in Introductory Microeconomics.

cators (column 5) leaves the coefficient virtually unchanged (0.073) relative to the specification that includes only ability controls. These results, consistent with the findings in Romer (1993), suggest that ability is positively related to both attendance and performance, so that in estimating the effect of attendance on performance it is crucial to take into account the effect of unobserved ability. Controlling for effort and motivation, instead, does note seem to have a major impact on the estimated coefficient for attendance.[19]

Focusing on the complete specification (column 5), all the ability indicators have a positive and significant coefficient. One additional percentage point of HSG or GPA corresponds to 0.20 and 0.22 percent improvements in test score, respectively. The point estimates for exam speed and calculus are also quantitatively large: one additional exam per annum is associated to a 1.19 per cent higher test score, and students who have passed calculus have a test score 3.42 percentage points higher than the others.[20] The indicators of motivation and effort also have the expected sign: one additional hour of study per week produces a 0.14 percentage point increase in performance, although the coefficient is only marginally significant. Subject evaluation has a positive coefficient (0.07) significant at the 10 per cent level, whereas teacher evaluation has a significant and larger coefficient: one additional percentage point in teacher evaluation corresponds to a 0.16 percentage point increase in test score.

Looking at the other controls in column 5, an additional year of registration has a significant negative impact on test score of 2.6 percentage points: the older a student, in terms of academic career, the worse his/her performance. Speaking a foreign language and living away from home both have very large negative and statistically significant effects on test score (-5.03 and -4.45, respectively). The coefficient on the female dummy, on the other hand, is negative but not statistically significant, indicating that gender does not have a significant effect on performance, consistently with the results in Williams et al. (1992) and Durden and Ellis (1995).[21].

The results in table 2 indicate that controlling for ability, effort and mo-

[19]This could be interpreted as indicating either that effort and motivation are not correlated with attendance, or that student and teacher evaluation and hours of study are not good proxies for motivation and effort.

[20]This finding is consistent with the results in Brasfield et al. (1992) and Durden and Ellis (1995).

[21]Other studies, however, report significant gender-related differences in performance (see e.g. Sigfried (1979), Lumsden and Scott (1987)).

tivation by means of proxy variables lowers the estimated coefficient for attendance from 0.09 to 0.073. This result could be interpreted positively, as in Romer (1993), as a sign that "an important part of the relationship reflects a genuine effect of attendance". An alternative, more plausible interpretation is that, despite the introduction of a set of control variables, the relationship still reflects the impact of omitted factors correlated with regressors: to the extent that, despite the control factors, there are still unobservable fixed effects correlated with attendance, $\widehat{\beta}_{OLS}$ remains biased and inconsistent (likely to be over-estimated). We thus turn to estimators that, under specific assumptions, are immune from the omitted variable bias.

Table 3 presents the results obtained estimating alternative specifications of the relationship between attendance and performance by instrumental variables (2SLS), using *travel time*, *work* and *web* as instruments for attendance. The results indicate that the estimated coefficient for attendance is very sensitive to the set of controls included in the specification. In particular, the inclusion of ability proxies (column 3) determines a large drop in the estimated coefficient (from 0.15 in the basic univariate specification to 0.07 in the full specification) and a negative impact on its significance. In the complete specification (column 5) the estimated coefficient for attendance falls to 0.065 and is not statistically significant. The high sensitivity of the results to the set of controls suggests that the instruments are not valid. This is confirmed by the results of the test for overidentifying restrictions (presented in the last two rows of table 3), that reject the hypothesis of instrument validity for all models (except, marginally, for the full specification in column 5).[22]

Given that IV estimation does not provide a solution to the omitted variable bias, we now turn to estimates obtained by exploiting the panel structure of the data set. Table 4 presents estimates of the fixed effect model (column 1) and the random effect model for alternative specifications (columns 2-4). The coefficient for lectures attended is positive and statistically significant in all models reported. The random effect estimates are similar to the OLS estimates, indicating that the weight of the between component in the error term is small relative to that of the within component. In particular, in the full specification (column 4) the RE model indicates that attending an extra

[22]This also implies that the assumptions on which the Davidson-McKinnon test (whose results do not reject the null hypothesis of exogeneity for attendance) is based are not met: the instruments are not truly exogenous, so that the IV estimator is not consistent.

one percent of lectures increases test score by 0.07 points.[23]

The RE estimates for all other regressors are quite similar to the ones obtained with OLS. All the ability indicators have positive and significant coefficients. Hours of study per week have a positive and marginally significant effect on performance. Both subject and teacher evaluation have positive coefficients, although only the latter is statistically significant. Year of registration, living away from home and foreign language all have large negative and significant coefficients, although the latter is only marginally statistically significant in the full specification.

The fixed effect model (column 1) produces an estimated coefficient for attendance that is positive and statistically significant at the 5 per cent level. The point estimate (0.039) is about half the size of the OLS and RE estimates, suggesting that there is indeed positive correlation between unobserved effects and time varying regressors, even after controlling for ability, effort, motivation and other individual characteristics. This is confirmed by the Hausman test statistic (29.28) that strongly rejects the null hypothesis of unobservable characteristics uncorrelated with attendance (p-value=0.01). This result is quite important, as it indicates that the inclusion of proxy variables is not sufficient to capture all the correlation between the regressor of interest and unobservable ability, effort and motivation.

Besides statistical significance, is the estimated effect of attendance on performance quantitatively relevant? Given that each two-hour lecture is equivalent to 12.5 per cent of total attendance, the 0.04 estimate in the fixed effect model implies that missing one lecture is associated to about a half percentage point drop in test score. This also implies that an average student who has not missed any lectures obtains a test score 1.2 percentage points higher than a student who has the average attendance level (70.8 per cent). It is interesting to observe that the return to each hour of self-study is substantially lower than that to each hour spent attending lectures.[24]

Summing up, the results for the panel estimators provide three main indications: first, proxy variables are not sufficient to control for omitted variable bias; second, even after eliminating the omitted variable bias, using a fixed effect estimator, attendance has a positive and significant impact

[23]Note also that the Breusch-Pagan LM test statistic strongly rejects OLS against RE.

[24]Given that each two-hour lecture is equivalent to 12.5 per cent of total attendance, using the 0.04 estimate for lecture attendance in the fixed effect model one obtains an estimated effect of each hour of lecture attendance of 0.25 (0.04*6.25), as opposed to 0.17 for self study.

on performance; third, the consistent coefficient estimate for attendance in the fixed effect model is about half the size of the OLS estimate, indicating that attending an extra one percent of lectures increases test score by 0.04 percentage points.

In order to provide a complete description of the relationship between attendance and performance, we now turn to the analysis of class attendance. In particular, we first examine whether class attendance has an impact on performance comparable to that of lectures, and then whether the respective roles of lectures and classes can be identified separately. In table 5, we report estimates obtained by replacing lecture attendance with class attendance (columns 1-3) and by including classes and lectures jointly (columns 4-6), comparing in both cases the results for the OLS, RE and FE models (all results refer to the full specification that includes the complete set of controls).

The coefficient for class attendance is positive and statistically significant in all models reported (table 5, columns 1 to 3). The point estimate is about 0.05 for both the OLS and RE estimators, and only slightly lower (0.037) for the FE estimator, remarkably close to the 0.039 FE estimate for lecture attendance reported in table 4. Interestingly, in this case the Hausman test does not reject the random effects model against the fixed effects model. This result indicates that, contrary to lecture attendance, class attendance is not significantly correlated with unobservable factors. One possible explanation for this result is that the decision to attend classes is less related to ability, given that it is commonly believed by students that class attendance has a higher return than lecture attendance for exam performance. Overall, the results suggest that the effect of class attendance on performance is significant and quantitatively similar to that of lecture attendance: an extra percentage point of class attendance increases test score by about 0.05 percentage points.

Next, we consider the estimates obtained inserting lectures and classes *jointly* in the full specification, to assess whether the respective roles of lectures and classes can be identified independently. As in the previous case, the OLS and RE estimates are quite similar (about 0.05 and 0.03 for lecture and class attendance, respectively), and the Breusch-Pagan LM test statistic strongly rejects OLS against RE. These results would seem to indicate that lecture and class attendance have independent effects on performance, and that lectures have a larger impact than classes.[25] However, the Hausman

[25]Note, however, that the difference between the 0.051 and 0.030 estimates for lectures

test rejects the consistency of the random effects model. The fixed effects model provides estimates of about 0.03 for both lectures and classes, but the coefficients are no longer statistically significant. This result indicates that, once we control for omitted variable bias, as we should, it is not possible to identify separately the effects of lecture and class attendance.

6 Discussion and conclusions

The results of the empirical analysis reported in this paper suggest two main conclusions. First, neither proxy variables nor instrumental variables provide a viable solution to the omitted variable bias in estimating the effect of attendance on academic performance. The alternative solution proposed in this paper is to exploit the panel structure of the data set to explicitly take into account the effect of unobservable factors correlated with attendance, such as student ability, effort and motivation. Second, after controlling for unobservable factors, attendance of either lectures or classes is found to have a smaller but significant impact on test scores in an Introductory Microeconomics course. On the basis of this evidence, can we conclude that teaching has a positive impact on student learning?

One possible objection could be that test scores are not a good measure of learning: attendance could affect exam performance because students learn how to do well in the exam, without any actual effect on the quality of learning (see e.g. Deere, 1994). This can be true if, for example, lectures only improve exam-taking skills, or provide information on the topics and type of questions that will be in the exam or, more generally, lectures present examinable material that is not covered in the textbook.[26] This kind of critique, however, does not apply to the data set investigated in this paper: all students had access to detailed lecture notes and past exam papers on the course web site, so that attendance did not reveal any private information. In addition, lectures and classes followed very closely the textbooks, so that all exam questions could be answered correctly by students not attending lectures or classes, who had relied exclusively on the texts to prepare for the exams. It should also be observed that the marking scheme was fully objective, so that test scores could not be used to reward students for attendance.

---

and classes in the RE model is not statistically significant.

[26]It is also possible that grades are used, either explicitly or implicitly, to reward for attendance.

A second possible argument is that, although the coefficient for attendance is significant, it is quantitatively small. Our results indicate that the estimated effect of attendance can be considered quantitatively relevant: missing one lecture is associated to about a half percentage point drop in test score. The opportunity cost of missing lectures is relevant not only in absolute terms but also in relative terms: the return to each hour of self-study is substantially lower than that to each hour spent attending lectures or classes. In assessing the size of the estimated coefficient for attendance we should also consider that measurement error, due to the self-reported nature of attendance, is likely to produce a downward bias in the estimate of its effect on performance. In addition, to the extent that regressors such as grade point average and exam speed reflect the effect of attendance in other courses (and that attendance is positively correlated across courses), the inclusion of these variables could lead to underestimate the effect of attendance on performance in Introductory Microeconomics.

Summing up, can we conclude that we, as academics, are doing something useful for student learning? According to the results of this study, yes. Alternative educational schemes, such as e-learning, would imply a positive and significant cost in terms of the quality of student learning. When considering the introduction of alternative educational models, the benefits of distance learning in terms of cost reduction for suppliers and time saving for students should be carefully weighted against the loss for student learning. Should then anything be done to increase attendance? Maybe yes. We find that the costs of absenteeism are significant and quantitatively relevant. In addition, we should consider that absenteeism implies not only a direct negative effect on learning, as reported in this study, but also significant negative externalities, such as the nuisance to the rest of the class and the high costs to the lecturer outside class.[27]

Does this mean that attendance should be made compulsory? Definitely not. A compulsory attendance policy would distort the opportunity cost of absenteeism and impose a welfare loss on students.[28] In addition, besides the fact that a captive audience is not a good learning environment, compulsory attendance would take away an important signal for lecturers on the quality

[27]See Brauer (1994).

[28]As observed by Deere (1994), a compulsory attendance policy would contradict many of the principles typically taught in introductory economics courses: "While students may not always make the wisest use of their time, it seems rather arrogant to suggest that we faculty know better the value of our subject just because we know our subject".

of their teaching.[29] The solution to the problem of high levels of academic absenteeism is not to make attendance compulsory, nor to design exams so as to make attendance necessary, but to improve the quality of our teaching, in terms of both content and format, to provide students with the right incentives and let them vote with their feet.

[29]See the comments in Brauer (1994) for a comprehensive discussion of the arguments against enforcing attendance.

7 References

Brasfield, D., McCoy, J. and Milkman, M. (1992). "The Effect of University Math on Student Performance in Principles of Economics." *Journal of Research and Development in Education*, 25(4), pp. 240-47.

Brauer, Jurgen, (1994), "Correspondence" *Journal of Economic Perspectives*, 8(3), pp. 205-215.

Bratti, M. and Staffolani, S. (2002) "Student time allocation and educational production functions", Working Paper n. 170, Economics Department, University of Ancona.

Browne, N. M., Hoag, J., Wheeler M. V., and N. Boudreau. (1991) "The Impact of Teachers in Economic Classrooms." *Journal of Economics*, 17, pp. 25-30.

Buckles, S. G. and M.E. McMahon (1971) "Further Evidence on the Value of Lecture in Elementary Economics" *Journal of Economic Education*, 2(2), pp. 138-41.

Coates, D. (2003) "Education Production Functions Using Instructional Time as an Input", *Education Economics*, vol. 11(3), pp. 273-92.

Cohn, E. and Cohn, S. (2000). Class attendance and performance in Principles of Economics. Proceedings of the Business and Economic Statistics Section. American Statistical Association, pp. 201-206.

Deere,D. (1994), in Brauer, J. "Correspondence", *Journal of Economic Perspectives*, 8(3), pp. 205-215.

Devadoss, S. and Foltz, J. (1996) "Evaluation of factors influencing student class attendance and performance", *American Journal of Agriculture Economics*, vol. 78, pp. 499-507.

Dolton, P., Marcenaro, O. D., Navarro, L. (2003) "The effective use of student time: a stochastic frontier production function case study" *Economics of Education Review*, 22 (6), pp. 547-560.

Durden, G. C. and Ellis, L. V. (1995) "The effects of attendance on student learning in principles of economics", *American Economic Review*, vol. 85, pp. 343-46.

Kirby A. and McElroy B. (2003) “The Effect of Attendance on Grade for First Year Economics Students in University College Cork”, *The Economic and Social Review*, 34(3), pp. 311-326.

Lazear, E. P. (2001) ”Educational Production”, *Quarterly Journal of Economics*, 116(3), pp. 777-803.

Lumsden, K. G. and A. Scott. (1987) ”The Economics Student Reexamined: Male-Female Differences in Comprehension.” *Journal of Economic Education*, 18(4), pp. 365-375.

Manahan, J. (1983) ”An Educational Production Function for Principles of Economics”, *Journal of Economic Education*, 14(2), pp. 11-16.

Marburger, D. R. (2001), “Absenteeism and Undergraduate Exam Performance”, *Journal of Economic Education*, p. 99-110.

McConnell, C. R. and C. Lamphear (1969) ”Teaching Principles of Economics Without Lectures.” *Journal of Economic Education*, 1(4), pp. 20-32.

Paden, D. W. and M. E. Moyer (1969) ”The Effectiveness of Teaching Methods: The Relative Effectiveness of Three Methods of Teaching Principles of Economics” *Journal of Economic Education*, 1, pp. 33-45.

Park, K. H. and P. M. Kerr (1990) “Determinants of Academic Performance: a Multinomial Logit Approach.” *Journal of Economic Education*, 21(2), pp. 101-11.

Pritchett, L. and Filmer, D. (1999) ”What Education Production Functions Really Show: A Positive Theory of Education Expenditures”, *Economics of Education Review*, 18(2), pp. 223-39.

Rodgers, J. R. (2001) “A panel-data study of the effect of student attendance on academic performance”, unpublished manuscript.

Rodgers, J. R. (2002) “Encouraging tutorial attendance at university did not improve performance”, *Australian Economic Papers*, 41(3), pp. 255-266.

Romer , D. (1993) ”Do students go to class? Should they?”, *Journal of Economic Perspectives*, 7, pp. 167-74.

Schmidt, R. 1983, "Who maximises what? A study in student time allocation", *American Economic Review* Papers and Proceedings, 73, pp.23-28.

Siegfried, John (1979) "Male-Female Differences in Economic Education: A Survey." *Journal of Economic Education*, 10(2), pp. 1-11.

Siegfried, J. J. and R. Fels (1979) "Research on Teaching College Economics: A Survey." *Journal of Economic Literature*, 17(3), pp. 923-969.

Todd, P. and Wolpin, K. (2003) "On the Specification and Estimation of the Production Function for Cognitive Achievement" *The Economic Journal*, 113, pp. F3-F33.

Williams, M. L., C. Waldauer and V. G. Duggal (1992) "Gender Differences in Economic Knowledge: An Extension of the Analysis." *Journal of Economic Education*, 23(3), pp. 219-31.

Table 1: Descriptive statistics

Variable	Mean	St. Dev.	Min	Max
Test score (%)	58.71	24.45	-41.67	100.00
Lectures attended (%)	70.82	27.78	0.00	100.00
Classes attended (%)	67.36	35.42	0.00	100.00
High school grade (%)	77.24	12.34	60.00	100.00
Grade point average (%)	76.86	8.72	60.00	100.00
Exams per annum	2.05	1.20	0.00	6.00
Hours of study (per week)	10.85	5.69	1.00	35.00
Subject evaluation	73.73	11.36	10.00	100.00
Teacher evaluation	80.89	12.20	10.00	100.00
Travel time (minutes)	46.91	26.88	1.00	150.00
Age	20.43	0.83	19.00	27.00
Year of registration	1.31	0.62	1.00	4.00
High school type	3.06	2.12	1.00	10.00
Father education	2.76	0.91	1.00	6.00
Mother education	2.63	0.89	1.00	6.00
Father occupation	4.03	2.57	1.00	11.00
Mother occupation	4.31	2.37	1.00	11.00
Province	1.53	1.23	1.00	7.00
Female	0.47	0.50	0.00	1.00
Foreign language	0.06	0.23	0.00	1.00
Calculus	0.36	0.48	0.00	1.00
Work	0.43	0.49	0.00	1.00
Web at home	0.74	0.44	0.00	1.00
Live away from home	0.10	0.30	0.00	1.00

Note: Number of observations: 3064 (N=766, T=4). See section 3 for details on the definition and construction of the variables.

Table 2: Determinants of academic performance: OLS estimates

Independent variable	(1)	(2)	(3)	(4)	(5)
Lectures attended	0.090	0.090	0.076	0.087	0.073
	(5.757)	(5.729)	(4.858)	(5.547)	(4.687)
High school grade			0.182		0.197
			(4.799)		(5.146)
Grade point average			0.231		0.223
			(4.240)		(4.100)
Exams per annum			1.217		1.185
			(2.973)		(2.875)
Calculus			3.838		3.421
			(3.804)		(3.391)
Hours of study				0.169	0.139
				(2.298)	(1.913)
Subject evaluation				0.084	0.066
				(2.371)	(1.887)
Teacher evaluation				0.144	0.155
				(3.817)	(4.149)
Year of registration		-3.243	-2.671	-3.242	-2.618
		(-4.735)	(-3.849)	(-4.790)	(-3.824)
Female		0.641	-0.869	0.161	-1.334
		(0.773)	(-1.043)	(0.194)	(-1.590)
Foreign language		-3.424	-4.144	-4.235	-5.026
		(-1.510)	(-1.826)	(-1.886)	(-2.216)
Away from home		-4.085	-3.718	-4.795	-4.450
		(-2.499)	(-2.237)	(-2.925)	(-2.667)
Adjusted R^2	0.253	0.276	0.302	0.283	0.308

Note: Dependent variable: *test score*. t-statistics reported in brackets (robust standard errors). Number of observations: 2913. All specifications include time fixed effects. Models (2) to (5) also include dummy variables for high school type, education and occupation for both father and mother, and province of residence. See section 3 for details on the definition and construction of regressors.

Table 3: Determinants of academic performance: IV estimates

Independent variable	(1)	(2)	(3)	(4)	(5)
Lectures attended	0.148	0.125	0.038	0.148	0.065
	(2.095)	(1.719)	(0.515)	(2.049)	(0.873)
High school grade			0.187		0.198
			(4.802)		(5.050)
Grade point average			0.233		0.223
			(4.279)		(4.109)
Exams per annum			1.239		1.190
			(3.015)		(2.873)
Calculus			3.973		3.449
			(3.809)		(3.321)
Hours of study				0.169	0.139
				(2.306)	(1.905)
Subject evaluation				0.077	0.067
				(2.125)	(1.878)
Teacher evaluation				0.142	0.156
				(3.710)	(4.143)
Year of registration		-3.125	-2.787	-3.039	-2.643
		(-4.264)	(-3.782)	(-4.200)	(-3.655)
Female		0.579	-0.833	0.067	-1.328
		(0.695)	(-0.998)	(0.079)	(-1.582)
Foreign language		-3.439	-4.152	-4.233	-5.033
		(-1.517)	(-1.824)	(-1.885)	(-2.218)
Away from home		-4.135	-3.654	-4.861	-4.439
		(-2.527)	(-2.186)	(-2.965)	(-2.654)
Adjusted R^2	0.249	0.275	0.300	0.279	0.308
Endogeneity test	0.724	0.239	0.270	0.763	0.014
(p-value)	(0.395)	(0.625)	(0.604)	(0.382)	(0.905)
Ov. restr. test	18.109	12.116	7.956	111.131	5.838
(p-value)	(0.000)	(0.002)	(0.019)	(0.000)	(0.054)

Note: Dependent variable: *test score*. t-statistics reported in brackets (robust standard errors). Number of observations: 2913. All specifications include time fixed effects. Models (2) to (5) also include dummy variables for high school type, education and occupation for both father and mother, and province of residence. See section 3 for details on the definition and construction of regressors. Instruments for *lectures attended*: *travel time, work, web.*

Table 4: Determinants of academic performance: panel estimates

Independent variable	F.E.	R.E. (1)	R.E. (2)	R.E. (3)
Lectures attended	0.039	0.082	0.082	0.070
	(2.158)	(5.184)	(5.160)	(4.446)
High school grade				0.210
				(4.754)
Grade point average				0.221
				(3.419)
Exams per annum				1.280
				(2.680)
Calculus				3.339
				(2.688)
Hours of study				0.166
				(1.916)
Subject evaluation				0.066
				(1.525)
Teacher evaluation				0.154
				(3.456)
Year of registration			-3.308	-2.620
			(-4.047)	(-3.291)
Female			0.474	-1.577
			(0.466)	(-1.563)
Foreign language			-2.964	-4.676
			(-1.145)	(-1.875)
Away from home			-4.087	-4.380
			(-2.109)	(-2.346)
Adjusted R^2	0.292	0.257	0.288	0.322

Note: Dependent variable: *test score*. t-statistics reported in brackets (robust standard errors). Number of observations: 2913. All specifications include time fixed effects. Models R.E.(2-3) also include dummy variables for high school type, education and occupation for both father and mother, and province of residence. Breusch-Pagan LM test statistic (OLS against RE(3)): 85.9 (p-value: 0.00). Hausman test statistic (RE(3) against FE): 29.28 (p-value: 0.01). See section 3 for details on the definition and construction of regressors.

Table 5: Determinants of academic performance: panel estimates

Equation	OLS (1)	F.E. (1)	R.E. (1)	OLS (2)	F.E. (2)	R.E. (2)
Lectures				0.053	0.026	0.051
				(2.763)	(1.198)	(2.777)
Classes	0.052	0.037	0.050	0.029	0.029	0.030
	(4.522)	(2.828)	(4.248)	(2.054)	(1.804)	(2.148)
Adjusted R^2	0.308	0.295	0.321	0.310	0.039	0.324

Note: Dependent variable: *test score*. t-statistics reported in brackets (robust standard errors). Number of observations: 2896. All specifications include time fixed effects and dummy variables for high school type, education and occupation for both father and mother, and province of residence.

Breusch-Pagan LM test statistic (OLS(1) against RE(1)): 89.23 (p-value: 0.00). (OLS(2) against RE(2)): 86.84 (p-value: 0.00) Hausman test statistic (RE(1) against FE(1): 2.16 (p-value: 0.99), (RE(2) against FE(2): 30.23 (p-value: 0.01).

See section 3 for details on the definition and construction of regressors.

Data appendix

Table 6: Summary statistics (means) by sub-group

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
Male	58.5	69.5	63.0	74.2	75.7	2.0	10.0	0.53
Female	59.0	72.3	72.4	80.7	78.3	2.1	11.9	0.47
Not-worker	60.1	75.2	73.2	78.2	77.7	2.1	10.9	0.58
Worker	57.0	64.5	58.9	76.0	75.8	2.0	10.9	0.42
No calculus	60.6	70.5	65.6	76.8	76.2	1.7	11.4	0.64
Calculus	64.5	73.3	71.9	80.1	78.5	2.7	10.8	0.36
Native lang.	59.5	70.8	67.0	77.2	77.0	2.1	11.0	0.95
Foreign lang.	51.5	70.1	69.7	82.3	74.6	2.1	11.2	0.05
Lived at home	60.0	71.0	67.3	77.3	76.8	2.1	10.9	0.91
Lived away	54.6	70.6	67.6	78.9	77.1	1.8	11.3	0.09

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.

Table 7: Summary statistics (means) by year of enrollment

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
First	60.0	72.3	68.3	78.3	77.1	2.1	10.9	0.77
Second	54.5	65.9	63.8	74.6	75.4	2.0	10.7	0.17
Third	56.6	66.7	66.0	70.6	78.8	1.7	11.9	0.05
Fourth	44.2	63.7	74.2	78.3	71.7	0.9	9.4	0.01

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.

Table 8: Summary statistics (means) by high-school type

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
Technical	59.2	72.0	69.5	81.2	77.2	2.0	10.8	0.41
Vocational	54.7	70.5	68.3	77.9	76.0	2.3	10.7	0.10
General	60.4	69.5	64.6	73.3	76.8	2.0	11.1	0.46
Other	50.4	77.2	75.8	77.9	77.7	1.5	11.3	0.03

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.

Table 9: Summary statistics (means) by parental education

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
Father								
Primary	61.9	69.6	66.7	81.1	76.8	1.7	10.1	0.09
Lower secondary	60.8	74.6	71.2	79.5	78.3	2.1	11.4	0.28
Upper secondary	58.6	70.1	67.1	76.9	76.1	2.1	10.8	0.43
University	59.0	65.3	60.0	75.3	76.6	2.1	10.8	0.18
Mother								
Primary	60.9	70.4	66.8	80.9	77.0	1.9	10.5	0.10
Lower secondary	60.6	71.2	68.4	78.7	77.9	2.1	11.2	0.34
Upper secondary	59.0	70.1	66.5	77.4	76.2	2.0	10.6	0.44
University	57.9	68.0	62.7	73.5	75.5	2.1	11.7	0.11

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.

Table 10: Summary statistics (means) by parental occupation

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
Father								
Manual worker	64.5	74.1	74.0	84.2	78.9	2.3	11.4	0.13
Clerk	57.2	71.8	67.1	77.1	76.6	2.0	11.2	0.24
Executive	61.0	67.9	63.6	73.9	75.2	2.1	10.2	0.14
Self-employed	61.5	68.4	61.8	77.3	75.6	1.9	11.7	0.12
Retired	59.8	69.4	66.4	78.2	77.9	2.1	10.9	0.21
Unemployed	60.4	74.0	74.2	78.1	77.4	2.1	10.1	0.11
Other	63.2	70.9	63.8	76.3	73.6	1.5	13.7	0.04
Mother								
Manual worker	61.4	67.4	64.9	82.6	78.5	2.1	12.4	0.08
Clerk	59.0	70.0	67.0	76.1	75.8	1.9	11.2	0.31
Executive	61.1	64.6	59.9	76.9	76.9	2.3	9.8	0.02
Self-employed	60.1	72.5	65.2	79.4	78.1	2.3	11.2	0.09
Retired	64.5	59.7	56.9	76.3	75.7	1.9	9.9	0.06
Unemployed	63.4	71.5	68.3	79.2	78.6	2.3	10.0	0.05
Other	60.5	73.1	69.8	78.6	77.3	2.1	11.2	0.39

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.

Table 11: Summary statistics (means) by province of origin

Variable	SCO	LEC	CLA	HSG	GPA	EPA	SSH	Freq.
Milan	58.5	70.9	66.4	77.0	76.5	2.0	11.0	0.82
Varese	54.5	70.5	70.9	77.9	80.6	1.6	12.8	0.02
Como	65.6	72.2	66.1	75.6	76.0	2.5	9.4	0.04
Lecco-Sondrio	62.6	71.1	74.3	80.4	78.8	2.4	10.2	0.09
Bergamo	63.9	58.8	56.1	72.4	78.6	1.8	10.9	0.01
Cremona-Lodi	52.2	50.1	65.5	71.8	86.7	1.0	17.3	0.01
Outside Lombardy	56.9	63.5	67.6	78.4	79.3	2.0	14.7	0.01

Note: Number of observations: 776. See section 3 for details on the definition and construction of variables.