DISCUSSION PAPER SERIES

IZA DP No. 3749

Am I Missing Something? The Effects of Absence from Class on Student Performance

Wiji Arulampalam
Robin A. Naylor
Jeremy Smith

October 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Am I Missing Something? The Effects of Absence from Class on Student Performance

Wiji Arulampalam
*University of Warwick
and IZA*

Robin A. Naylor
University of Warwick

Jeremy Smith
University of Warwick

Discussion Paper No. 3749
October 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3749
October 2008

ABSTRACT

Am I Missing Something? The Effects of Absence from Class on Student Performance*

We exploit a rich administrative panel data-set for cohorts of Economics students at a UK university in order to identify causal effects of class absence on student performance. We utilise the panel properties of the data to control for unobserved heterogeneity across students and hence for endogeneity between absence and academic performance of students stemming from the likely influence of unobserved effort and ability on both absence and performance. Our estimations also exploit features of the data such as the random assignment of students to classes and information on the timetable of classes, which yield potential instruments in our identification strategy. Among other results, we find that there is a causal effect of absence on performance for students: missing class leads to poorer performance. There is evidence from a quantile regression specification that this is particularly true for better-performing students, consistent with our hypothesis that effects of absence on performance are likely to vary with factors such as student ability.

JEL Classification: C41, J24, I2

Keywords: randomised experiments, quantile regression, selection correction, panel data, education, student performance, class absence

Corresponding author:

Wiji Arulampalam
Department of Economics
University of Warwick
Coventry CV4 7AL
UK
E-mail: wiji.arulampalam@warwick.ac.uk

* We are grateful for helpful comments from Pedro Martens, Mark Stewart, Ian Walker, and participants at: the annual conference of the American Economic Association, New Orleans, January 2008 and especially to the discussant, Tom Dee; the annual conference of the Royal Economic Society, Warwick, March 2008; the IFS/CEE/CEP seminar, November 2007; the Reading University economics seminar, October 2007; the Warwick economics department econometrics workshop, November 2006 and the European Summer Symposium in Labor Economics, Munich, September 2007. Wiji Arulampalam also thanks the IZA for their hospitality in July 2008 and participants at the lunchtime seminar.

1. Introduction

The analysis of the factors associated with educational attainment and performance has been a major focus of work in the last decade as economists have attempted to obtain a deeper understanding of the processes associated with the acquisition of human capital. Work has tended to focus on the importance of factors such as class size and peer effects; see, for example, Hanushek *et al*, (2003), Hoxby, (2000), Krueger (2000), Ehrenberg *et al*. (2001), and Burtless (1996), *inter alia*. Much of the work has concentrated on the educational attainment of pupils in compulsory schooling, with less attention paid to higher education. Yet the processes associated with post-compulsory human capital accumulation are internationally important given increasing participation rates and the economic significance of the higher education (HE) sector in modern economies.

A further motivation for analysing HE processes specifically is that their nature is likely to be fundamentally different to those characterising earlier-stage cognitive development, with greater student autonomy in study one obvious reason for this. One aspect of this autonomy is the relative freedom of students in HE to choose to absent themselves from class. More generally, the modes of study of HE students are less prescribed than in compulsory education: the responsibility for the efficient allocation of study time lies largely with the student – though this is not necessarily uniformly true across all university courses. Our work focuses on issues regarding the relationship between absence from class and academic performance of university students, a subject which has attracted attention since the influential paper of Romer (1993). Our analysis concentrates on variations across students in the causal impact of absence on performance.

Currently in the UK, as elsewhere, there are significant changes taking place in HE. Following several decades in which the unit of resource has fallen, the introduction of tuition fees for home students offers the potential prospect of better resourced teaching in the UK. There are a number of possible implications for the ways in which the nature of the teaching and learning environment in universities might evolve in response. Traditionally, university teaching in the UK is based on large-group lectures alongside small-group classes. Attendance at lectures has been seen as optional. Class attendance, however, has been regarded as compulsory for various reasons, which include the perception that: (i) the value added in class is greater than that in lectures and students might not appreciate this and (ii) class attendance by each student has positive externalities for other students through the contributions each can make to the learning process. With declining resources in the previous twenty years, class sizes have been increasing across the HE sector and this has undermined the strength of both of these arguments. Indeed, it is likely that in large classes attendance imposes negative externalities through congestion effects.[1] There is a view in the sector that class teaching is no longer as effective as it was and that students, perceiving this, have higher absence rates. There are various responses to this, including: abandoning small group class teaching; resourcing it better; reforming it; making attendance compulsory; and/or creating more explicit incentives for attendance.[2] Developing a better understanding of the effects of class attendance on student outcomes seems timely.

Our empirical analysis makes use of a rich administrative panel dataset for cohorts of economics students at a UK university. We exploit the panel properties of the data to control for unobserved heterogeneity across students and hence for endogeneity between attendance and performance stemming from the likely influence of ability, effort, and motivation on both. That students are randomly assigned to classes avoids the potential endogeneity

---

[1] Lazear (2001) develops a theoretical model in which classroom education has public good characteristics. In Lazear's model, a disruptive student reduces others' learning.

[2] For further discussion, see Romer's (1994) response to correspondence in the Journal of Economic Perspectives.

problems that occur when students can self select into classes. Finally, we use the idea that the time slot of the class in the weekly timetable produces exogenous variation in a student's attendance and hence acts as a potential instrument with which to identify a causal effect of attendance on performance.

In our empirical analysis, we establish a number of results. Conditional mean model parameter estimates based on simple OLS suggest that there is a significant negative association between missing class and student performance. When we control for unobserved individual effects, however, we find that the estimated effects of absence on performance are reduced and are only weakly significant. IV estimates which correct for selectivity in absence behaviour are similar to OLS estimates at the conditional mean. In a quantile regression specification, it emerges that the adverse effect of missing class is greater for better-performing students, consistent with the hypothesis that effects are likely to vary with factors such as student ability.

In the next section, we present a brief literature review. This is followed, in Section 3, by a theoretical motivation for the analysis of the relationship between absence and student performance. Section 4 presents the data, and key summary statistics, based on cohorts of economics students at the particular UK university. In Section 5, we describe the econometric model of the causal effects of tutorial absence on student performance and discuss the results. Section 6 concludes and offers further remarks.

2. Contextual literature

Romer (1993) presented quantitative evidence on absenteeism and performance in economics courses at 3 universities in the US, reporting absenteeism to be 'rampant', with an overall absence rate of about one-third. Romer also reported evidence consistent with the hypothesis that absence is associated with poor student performance, while acknowledging that no causal effect had been demonstrated given the endogenous nature of the relationship between attendance and performance. The general assumption in the literature is that more able (and motivated and hard-working) students are more likely both to attend and to score highly in their courses. Thus, in the absence of adequate proxies for such personal characteristics, part of any estimated effect of attendance will reflect a form of (upward) ability bias arising out of endogenous selection. Romer does include in his regression analysis controls for prior grade point average on the grounds that these will capture some of the otherwise unobserved heterogeneity across students. Indeed, Romer notes that as the estimated effect of prior performance depends in part on previous attendance, the inclusion of prior scores could cause a *downward* bias in the estimate of the effects of attendance on performance; part of the effect being captured in the control variable.

Following Romer (1993), Durden and Ellis (1995) analyse survey data on absences for 346 economics students on a principles of economics course at a single US university. They report that the average effect of absences on performance is modest, but that there are substantial adverse effects when absence exceeds certain threshold levels, a finding supported by Cohn and Johnson (2006) in their study of 1st year economics students from the University of South Carolina from 1997-2001. Devadoss and Foltz (1996) also report significant positive effects of attendance on student performance from a survey-based analysis of students, across 4 US universities, taking a course in agricultural economics. The analysis exploits survey responses to questions eliciting information on prior attainment, student effort and motivation. This information generates proxies for these typically unobserved characteristics. Stinebrickner and Stinebrickner (2008) also exploit survey data on how many classes students attended in an analysis of the causal effect of studying on performance. They produce IV estimates indicating a significant and substantial effect of study quantity on academic performance. Dobkin, Gil and Marion (2007) use a discontinuity design approach generated by a policy of making attendance compulsory only for students achieving below some threshold on a mid-term test in each of three large economics classes in a US university. They find that the policy had a significant effect on attendance and also generated a significant discontinuity in final examination performance.

Marburger (2001), using data on 60 1st year microeconomic students matched answers in multiple choice questions to attendance at specific classes, where the material for the specific multiple choice question was covered, and found a significant positive relationship between attendance and performance. Neri and Meoche (2007), undertaking a similar analysis for students taking a microeconomic course in Australia, report similar results.

It is important to note that the literature for the US typically measures attendance rates aggregated over all forms of meetings, lectures and smaller group meetings combined, and labels these as ‘classes’. There is, however, a potentially important distinction to be drawn between attendance at lectures – typically large group meetings – and at classes, which are typically small group meetings. In our own study, we are concerned with the causes and consequences of missing small group classes.

Rodgers (2002), using data on attendance in an introductory statistics course at an Australian university, finds a strong positive association between attendance at tutorials (small group classes) and performance but, comparing across cohorts, reports that the introduction of a scheme which raised attendance was not associated with enhanced performance. Rodgers infers that attendance alone does not improve achievement. However, Marburger (2006) updating the analysis of his earlier paper, finds evidence that an enforced mandatory attendance policy improves student performance. Stanca (2006) uses a survey-based panel data set of students taking a microeconomics course at an Italian university. The analysis exploits the panel nature of the data to take account of unobserved characteristics correlated with attendance and produces estimates indicating a significant positive causal effect of attendance on performance: both class and lecture attendance rates are observed and have similar effects on performance.

Chen and Lin (2006), using data on students taking a course in public finance in Korea, undertook a randomised experiment in which the course was taught twice in a single year to distinct sets of students. In the delivery of the lectures to one of the groups, 10% of the examinable material was randomly excluded from lecture slides (the authors denoted this as experimental lecture absence), although students were made aware of the scope of material to be covered and were informed that material not covered in lectures was still examinable. Using only those students who attended lectures, the authors found a significant and negative relationship between experimental lecture absence and performance. The result could be interpreted as indicating a positive marginal productivity associated with lecture-based teaching, though it is not clear that experimental absence for the whole group is equivalent to individual absence from lectures actually given – and actually received by at least some peers.

Kirby and McElroy (2003), using travel-to-college commuting times as an instrument, find that attendance has a positive and diminishing marginal effect on grades achieved at an Irish university: they find that tutorial attendance has quantitatively bigger effects than lecture attendance. In contrast, for the UK, Martins and Walker (2006) find no significant effects of class attendance on performance for students in the economics department at a UK university, and also find no significant effects of smaller classes on improved performance. Dolton, Marcenaro and Navarro (2003), exploiting rare information on time use for students at a Spanish university, find that both formal attendance and self-study activity are significant positive determinants of examination score: formal attendance includes both class and lectures, the two are not distinguished.

In our own work, we focus exclusively on absence from small group classes as we do not have information on lecture absences: as is commonly the case, attendance at lectures is regarded as voluntary and is not monitored. At the institution under analysis, comprehensive lecture notes are typically available on-line and are to some large extent substitutable by recommended textbooks and other supplementary material. Hence missing lectures is, *a priori*, likely to be less costly than missing small group classes, where teaching is more targeted to the individual's needs and is less easily substituted by other forms of study.

Our analysis attempts to uncover causal effects of small group class attendance (or, more specifically, absence – its converse) on the performance of students taking core courses in the second (intermediate) year[3] of an economics degree course at a UK university. We exploit administrative data for three different cohorts of students taking three common compulsory courses. Given the panel nature of the data, we are able to control for unobserved heterogeneity. We also control for previous attainment and use precise information on the students' class timetables to generate instruments which we employ in a strategy for the identification of causal effects.

3. Theoretical discussion

Consider an educational production function[4] of the following form:

$$p = p(c, q, r), \tag{1}$$

where p is a measure of a student's educational performance, c is the amount of time allocated by the student to attending class, q is the amount of time spent in alternative forms of study activity, and r captures personal characteristics such as ability, effort and

---

[3] We describe the choice and the properties of the data and of the institutional context in more detail in the Section 4.

[4] For a discussion of the empirical estimation of production functions for cognitive achievement, see Todd and Wolpin (2003).

motivation. Suppose that the objective function of the student is to maximise performance, given by equation (1). Among the constraints will be a time constraint of the form:

$$c + q \leq \bar{t}\,, \tag{2}$$

where $\bar{t}$ is the maximum amount of time available for study in a given period. In the production function, assume initially that c and q are neither complements nor substitutes but are independent.[5] The problem for the 'grade'-maximising student is to allocate their time efficiently between attending class and alternative study time uses, such as attending large-group lectures, private study, or completing assignments. Privately efficient time allocation – we are ignoring externality and public good characteristics of classroom attendance for now (see Lazear, 2001) – will require the student to have knowledge of the marginal productivity of c and q in (1).

In reality, marginal products are likely to be person-specific: one of the challenges for the student is to reflect on their own learning strategies and capacities in developing for themselves a mature appreciation of 'what works best for them' – that is, of their own marginal productivities for the factors in their educational production function. Implicitly, the importance of this is embodied in the current emphasis on 'reflective learning' and 'personal development planning' in UK higher education. One of the reasons for not selecting first year students in our data analysis is the acknowledgement that only by the second year will students have had sufficient experience of higher education to be able to make informed judgements about their optimal learning strategies. This is also recognised by most UK universities as, typically, first year performance does not contribute to final degree marks and classifications.

---

[5] Work by Douglas and Sulock (1995) suggests the two activities are substitutes.

Assume for now that the student has accurate information regarding the parameters of their own educational production function. Assume also that marginal products of study time are positive but diminishing in each study activity and are initially, at least, independent of each other and of ability, i.e. $\frac{\partial p}{\partial c} \equiv mpc > 0$, $\frac{\partial p}{\partial q} \equiv mpq > 0$, $\frac{\partial^2 p}{\partial c^2} < 0$, $\frac{\partial^2 p}{\partial q^2} < 0$, $\frac{\partial^2 p}{\partial c \partial q} = 0$, $\frac{\partial^2 p}{\partial c \partial r} = 0$, and $\frac{\partial^2 p}{\partial q \partial r} = 0$. We additionally assume that, $\frac{\partial p}{\partial r} > 0$. With these assumptions, we can represent diagrammatically the solution to the problem defined in equations (1) and (2) for the grade-maximising student: see Figure 1. Consider first Figure 1a.

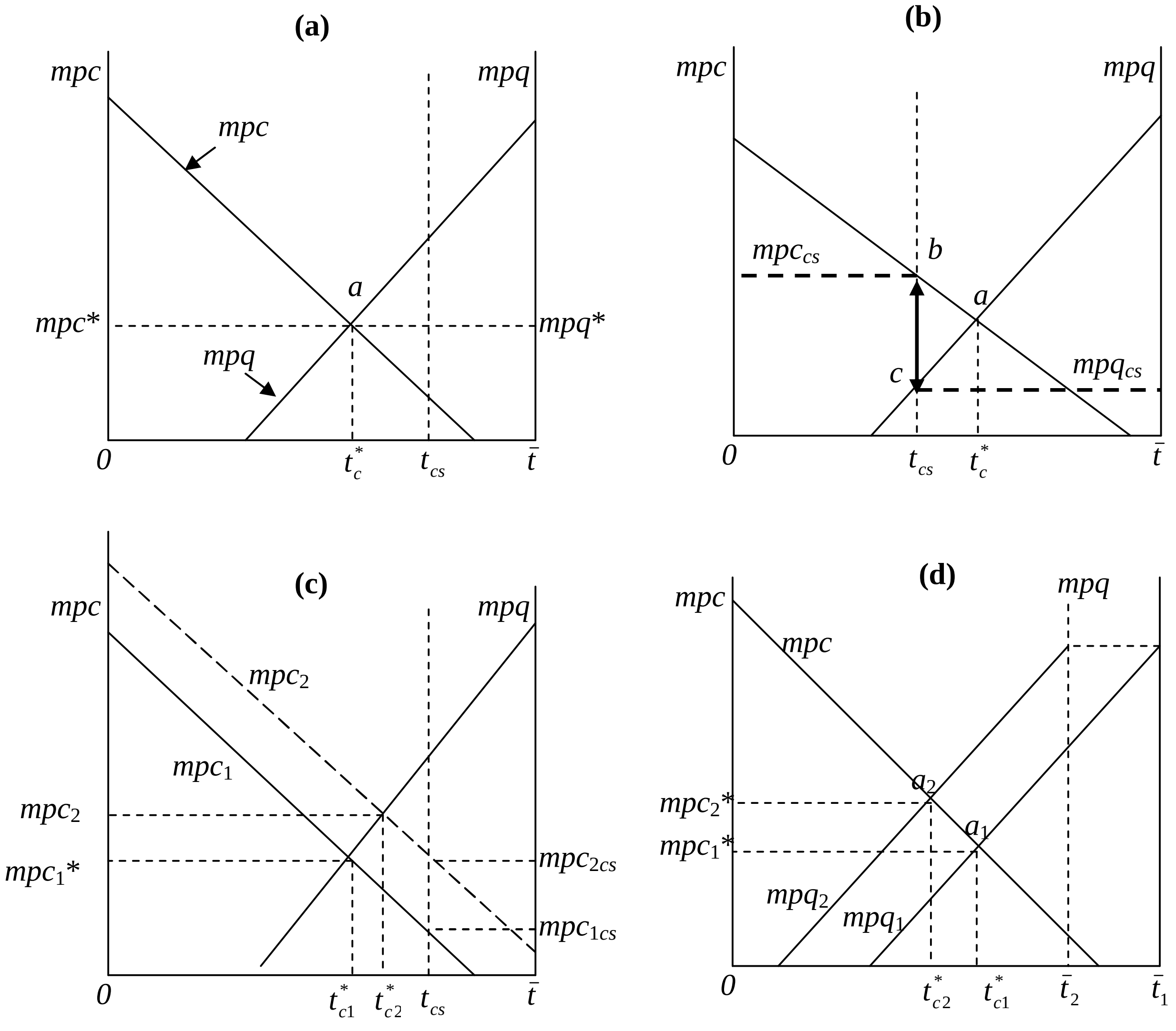

Figure 1 - Efficient study-time allocations.

From Figure 1a, we can see that the grade-maximising student will optimise at point *a*, choosing to attend t_c^* hours of class and engaging in $\bar{t} - t_c^*$ hours of additional study. Whether this involves absences from class will depend on the number of scheduled classes available to the student. If there are significant external net benefits of attending class, then – depending on institutional resources – the number of classes supplied to the student, denoted by t_{cs}, is more likely to exceed the student's optimal number, and hence $t_{cs} > t_c^*$, as in Figure 1a. If, on the other hand, $t_{cs} < t_c^*$, then the outcome will be inefficient, at least according to the student's private calculus: as shown in Figure 1b, where there is now a wedge between the marginal products: $mpc > mpq$. In this case, the optimising student will miss no classes and, indeed, would prefer more classes to be provided.

In the case described in Figure 1a, the optimising student will choose to miss $t_{cs} - t_c^*$ hours of class. At the margin, were the student required to attend all t_{cs} classes, then there would be a fall in the student's performance level as $mpq > mpc$ for the marginal classes. Suppose that class attendance is compulsory but that absence is not penalised. Then the propensity of students to miss at least some fraction of the sub-optimal $t_{cs} - t_c^*$ classes will depend on their attitudes to compliance. Suppose that this is randomly distributed across students. Then it follows that, under the assumptions of the model, in a learning environment in which class attendance is regarded as compulsory but in which, without enforcement, some individuals absent themselves nonetheless, class absences in the range $t_{cs} - t_c^*$ will be associated with improved performance. This is the opposite prediction to the standard hypothesis in the literature that predicts that absence will affect performance adversely. Our prediction arises from an optimising framework in which choices are made with perfect information: at the margin, attendance is productive, but only up to the optimising point.

So far, we have assumed that factor inputs are independent. But suppose now that, *ceteris paribus*, the marginal product of attending class is positively correlated with ability: that is, $\frac{\partial^2 p}{\partial c \partial r} > 0$. This case is represented in Figure 1c, where the mpc for more able students, mpc_2, lies above that of the less able, mpc_1. The result is that the more able students will optimally choose to miss fewer classes: $t^*_{c2} > t^*_{c1}$ in Figure 1c. In an environment in which class attendance is voluntary, performance will be greater for the more able students and, hence, will appear to have a negative association with absence from class. Of course, mpq may also be positively correlated with characteristics captured by r. In this case, the relative sign of $t^*_{c2} - t^*_{c1}$ (and hence the association between performance and absence) will be ambiguous: it will depend on comparative advantage; that is, the relative correlation of r with mpc and with mpq.

In an econometric estimation of the effects of absence on performance, correlation between r and either of the other arguments – c, q – in the education production function given by equation (1) could potentially generate endogeneity bias if r is not perfectly observed. If more able students are less likely to be absent from class – $t^*_{c2} > t^*_{c1}$, as in Figure 1c above – then the estimated adverse effect of absence on performance will be biased upwards, in absolute terms, through endogenous selection and the resulting ability bias. Hence, empirical investigation of the effects of absence from class on performance should be constructed so as to allow for heterogeneous effects of this sort. This observation lies behind the design of our later estimation strategy. In the case in which $t^*_{c2} < t^*_{c1}$, then the direction of endogeneity bias will be downward – but, again, the effects will be heterogeneous.

As we have seen, ability differences across students can affect absences from class through their influence on the educational production function, equation (1). But suppose now that there are differences across students in the time endowment for study, $\bar{t}$. In Figure 1d, we consider the effects of an exogenous reduction in the amount of time available for study activity: $\bar{t}$ falls from $\bar{t}_1$ to $\bar{t}_2$. In this case, as mpq shifts to the left from mpq_1 to mpq_2, there will be an increase in the number of classes missed (assuming $t_{cs} > t_{c1}^*$) together with an associated reduction in performance. In this model, the total time endowment for study, $\bar{t}$, is taken as exogenous. In reality, $\bar{t}$ is likely to be influenced by various arguments. For example, students from economically less advantaged backgrounds may be more likely to have to engage in part-time labour market activity, thereby reducing $\bar{t}$. The study time constraint may also be related to student ability, and hence to r in equation (1). If, for example, more able students undertake more non-curricular activities, then $\bar{t}$ will be negatively correlated with ability. In this case, more able students will be more likely to miss class. Note also from Figure 1d that the effect of missing class will be greater for more able students as $mpc_2^* > mpc_1^*$. Again, unobserved differences in ability across students will generate a bias in the estimate of the effect of absence on performance as part of the association between absence and performance is being explained by a differential propensity of the more able to be absent from class; in the case of figure 1d this is driven by differences in time constraints correlated with ability.

In summary, we have seen that, in an optimising framework, the theoretical effect of absence on performance is ambiguous. If class attendance is compulsory and students differ only in a randomly-distributed propensity toward compliance, then absence will have a positive association with performance as the less compliant will be more likely to adhere to the optimal number of classes. If, on the other hand, students are heterogeneous in ability then they will be likely to choose different optimal levels of class attendance: if ability is associated with a comparative advantage in class attendance – as in Figure 1c – then the more able will have a higher attendance rate and absence will be associated with poorer educational performance. Ability might also be correlated with the study time endowment: if more able students have a higher opportunity cost of studying, then it is likely that they will attend fewer classes. In this case, absence will be likely to have a positive association with performance. Estimation of the effects of absence on performance will be biased if ability is not observed or accurately proxied: the direction of bias will depend on the relative dominance of factors of the type we have identified. Finally, the model predicts that the magnitude of any effects of absence on performance will vary with student ability: if, for example, ability is relatively highly correlated with productivity of class attendance then the negative effect of any given level of absence on performance will be most pronounced for the more able students. These considerations inform our choice of empirical estimation strategy.

The model we have outlined so far assumes that students have sufficient information to be able to select their optimal level of class attendance. In reality, this is unlikely and students will make mistakes, attending either more or fewer classes than would be privately efficient. If students systematically under-estimate the marginal product of class attendance, then absence will tend to have an adverse effect on performance. This tendency might also be correlated with ability, so that less able students miss more classes and suffer a further reduced level of performance, thereby generating a further reason to expect heterogeneous effects of absence on performance.

Informed by this contextual optimising framework, our empirical strategy will involve: first, an analysis of the factors associated with being absent from class; second, a simple, or 'naïve', analysis of the association between student performance and student absence from class; third, an attempt to identify causal effects of absence from class on student performance; and fourth, an investigation of whether or how any effects vary systematically with student characteristics, such as those associated with ability. The following sections describe the data and the econometric strategy for investigating these and related issues.

4. Data description and summary statistics

This paper uses administrative data from the department of economics at a UK university, collected over a three year period, in order to investigate the association between absence from class and student performance. The observations are based on 444 2^{nd} year undergraduate students admitted to the economics department to commence their 3-year degree in October of 2003, 2004 or 2005; the number of students in each of the three cohorts was 134, 159 and 151, respectively.

4.1 Institutional context

As is typical in the UK, students observed in the study graduate at the end of their third year with a degree which is classified into one of four main classes. This overall degree classification of students depends on their performance in the 8 courses which they take over their 2^{nd} and 3^{rd} years of study, with the four courses taken in the 2^{nd} year having the same weight as the four taken in the 3^{rd} year. The 1^{st} year is simply a pass/fail year determining whether students progress into their 2^{nd} year of study. We focus exclusively on 2^{nd} year students in this paper for two reasons. First, we wish to analyse the behaviour of students whose motivation is likely to be to maximise their final score in each course and, as the 1^{st} year is simply a pass/fail year, this may well not be the case for first years who, arguably, are more likely to satisfice than to maximise. Second, we do not include 3^{rd} year students in the analysis as these students self-select into optional courses for three of their four modules, creating potential problems associated with endogenous module selection.

Each of the 444 2^{nd} year students observed in the study registers for three compulsory core courses (in microeconomics, macroeconomics, and econometrics) and for one optional course (selected from a potentially long list of courses from either economics or other social science disciplines). Our analysis is based on data relating only to the three compulsory courses: this gives us a balanced panel based on a total of 1332 observations.

In each course, each student is allocated to a tutorial class. These are small groups which meet with an academic tutor to complement lectures and which focus on exercise or discussion sheets given out by the lecturer of the particular course. Classes are provided for all compulsory 2nd year courses and attendance is regarded as compulsory. Students are allocated to their classes by the department and this is chiefly done on an alphabetical basis. Classes meet every week in the econometrics course (a total of 20 classes in the academic year), but are held less frequently in macroeconomics (13 classes) and in microeconomics (16 classes). In each class, the tutor takes a register for that class and this information is recorded electronically. The students are sent warnings about each of their absences with reminders about the adverse consequences associated with a poor attendance record. A 3rd absence in any one course requires the student to see their personal tutor and a 4th absence means the student must discuss their behaviour with the programme director. After the 4th absence the student is put on report and a record of their attendance across all of their courses is monitored closely. The department can impose various penalties on students with poor attendance records and, in particularly bad cases, can seek to have a student's registration withdrawn from the University for persistent non-attendance at tutorial classes, although this did not happen over the period under analysis.[6] For this system of monitoring and punishment of absence to be effective, it is important that tutors enter accurate information on each student's class attendance. This is fully understood in the department and is itself carefully monitored and policed.

---

[6] A further reason for restricting our analysis to compulsory courses is that attendance in optional courses is not necessarily recorded.

Once students have been randomly allocated to classes in each of their three compulsory courses, they are not free to re-allocate themselves to different classes – for example at more preferred times or into groups with selected class mates. One reason for this is that classes are of fixed sizes and so re-adjustment would disturb the general equilibrium allocation. There is a process by which students can appeal to be re-allocated to a different class, but permission is granted only where a special case can be made, for example on the grounds of a timetable clash involving an optional module in a different department. If in any one week a student is unable to attend their allocated class, then they can attend an alternative class on a one-off basis. This would not be recorded as an absence. The key point is that assignment to class is a random outcome: we present evidence relating to this in section 5 of the paper. We note also that the attendance database distinguishes between condoned absences – due to, for example, illness or having to attend an internship interview – and uncondoned absences. In our empirical analysis we consider both types separately but find that the results do not vary across the two types, hence we combine them.

Student performance in each of the observed courses is examined by a 3-hour end-of-year written examination worth 80% and by two pieces of assessed work, each worth 10%. In our empirical analysis of the effects of absence on performance, we have considered both overall performance and examination performance only. *A priori*, there is no reason to suppose that class attendance should affect one type of assessment more than another. We find that the empirical results are the same independent of which performance measure is used and, in the empirical section, report only results based on total performance.

Among other information on each student, we observe: prior performance of each student in each of their first year courses; both the absence record and the performance of each student in each of the three compulsory 2nd year courses; gender; home (UK/EU) or overseas student status; entry cohort (2003, 2004 or 2005); the class to which they are allocated on each course; the time of day and the day of the week of the class; and the average student rating of the tutor in each class. Students are registered either for a single degree programme in economics or a specialist variant thereof. Specialist degree students take parallel courses to those for single economics students. In our econometric analyses, we include appropriate controls for this. We have also re-estimated our econometric models on the basis of single degree students only and find that the results are robust to this. We prefer to include both sets of students as this generates a larger sample and improves the precision of estimates.

Historically, students in the department self-selected into their tutorial class groups through sign-up sheets on a first-come first-served basis in the micro and macroeconomics courses. In econometrics, students were allocated to classes on the basis of their first year performance in statistics. Neither of these allocation methods generated random assignment. Since 2003, students have been allocated to their classes – in all compulsory 2nd year courses – by the department administrators on the basis of the alphabetical ordering of family names.[7] Thus, combined with the collection of accurate data on absence, this provides the opportunity to exploit features of the data which have the characteristics of randomised experiments.

As we discuss in more detail in the description of our empirical methodology in section 5 below, the panel nature of our dataset enables us to control for unobserved heterogeneity, under the maintained assumption that unobserved individual-specific effects on performance are constant across courses. Randomised class assignment overcomes a number of possible sources of endogeneity between absence and performance – though not all sources. We are, however, able to use information on aspects of the assignment to class – such as by time of day – to construct instruments for the identification of causal effects of

---

[7] We discuss below the extent to which this generates a random assignment of students to classes.

absence on performance.

4.2 Summary Statistics

Despite being compulsory, attendance falls well short of 100%. Figure 2 plots the cumulative distribution function for the proportion of total absences across the three core courses for our 444 second-year students. From the figure, we note that the median person missed around 8% of their tutorial classes, while the 75th percentile student missed around 16% of classes. The mean level of absences, at 11%, is higher than the median, reflecting the fact that a small proportion of students missed a relatively large proportion of their classes, as is evident from the figure. Surprisingly, approximately 12% of students missed no classes. Absence is clearly less rampant than in the case reported by Romer (1993) in his US study.

While Figure 2 is based on the absence rate observed for each of the 444 students over their 3 modules, Figure 3 is based on the tutorial group as the unit of observation, the figure showing the cumulative distribution of absences across the 142 groups.[8] For simplicity of exposition, consider the case of a hypothetical group with 10 assigned students and 20 meetings over the year: a group absence rate of 10% could arise from just one student missing all their class meetings or from each student missing 2 classes over the year, or from various combinations between these extremes.[9] An absence rate of zero for any one group means that no student was absent for any of the meetings of that group during the whole year: remarkably, almost 4% of groups had a zero absence rate. The median class experienced an absence rate of 10% while 25% of classes had an absence rate of 7% or less and 75% had an absence rate no higher than 15%.

[8] With 142 groups over the 3 modules, this equates to an average of about 47 groups per module (over the three cohort years) and hence to an average of about 10 (of the 444) students per group.

[9] In the current paper, we do not exploit information on the distribution of absences by class by week.

Table 1 presents summary statistics of the variables used in the analysis, for the 444 students across their three core courses, broken down into bands according to the proportions of total class absences in each course: the band categories are defined as; (0-4]%, (4-8%], (8-15%], and over 15%.[10] Of the 2nd year students, 34% are female and 32% are overseas fee-paying. From the table, females are less likely to have high absence rates; overseas students appear to have higher rates of absence than do home students. Cohort size is not constant over time as entry is not restricted to an absolute fixed quota. Module class sizes are nonetheless constant as class provision varies with cohort size. The pattern of absences is not constant over the three cohorts: for example, the incidence of very low absences is lower among the 04/05 cohort. In subsequent analysis, we include cohort dummies and also conduct a sensitivity check in which we re-estimate our models based on just two cohorts: results are not sensitive to the dropping of any one cohort – though the standard errors do increase, causing a loss in the precision of estimates.

Table 1 shows that, in the raw data, there is a monotonic relationship between performance and absence in the second year: while the average score across all students over their three compulsory modules is 60%, the average for students with fewer than 4% absences is 65% while that for students with more than 15% absences is 55%. Of course, this does not amount to evidence of a causal effect of absence on performance. We also see from the table that there is a monotonic relationship between absence and prior performance. For example, while the average 1st year maths score is 68%, it was 73% for students subsequently missing less than 4% of their 2nd year classes and 63% for those missing more than 15%. Similar monotonic patterns are observed between 2nd year absence and (i) 1st year scores in statistics, (ii) the average score over all other 1st year modules, and (iii) having failed and re-

[10] The bin values used on the percentage of absences yield approximately equal numbers of observations in each bin.

sat 1st year modules. One might have expected students failing 1st year modules to have a better than average attendance record in the 2nd year: our discussion in section 3, however, shows that in an optimising framework, this need not follow.

Table 2 shows summary statistics for which the unit of observation is the 142 tutorial class groups. The table also shows a breakdown by absence band, where the cut-offs are selected to generate approximately equal numbers of cases per band. The mean proportion of absences across the groups is 11% - close to the median reported from Figure 3. 13% of classes were timetabled at 9am. We can see an association between the absence rate from classes and the time at which they are timetabled. Among classes with the lowest absence rates, 11% of classes began at 9am: in contrast, 17% of classes with an absence rate in excess of 15% began at 9am. Both time of day and day of week are potential instruments to use in our identification strategy. The timetable of classes is constructed around the constraints imposed by the timetable for lectures – which are not necessarily distributed uniformly across the week and hence there is likely to be clustering of tutorial classes by time and day. As teaching is typically not scheduled on a Wednesday afternoon (in order to accommodate sports activities), there are 9 morning or afternoon sessions per week: random timetabling of classes would then generate 11% of classes on a Wednesday and 22% on each of the other days. Instead, because of clustering, as many as 38% of tutorials are on a Thursday and 16% are on a Wednesday. Only 29% of classes occur in the afternoons: there is some evidence of a relationship between afternoon meetings and absence in the raw data: a relatively small proportion of classes with high absence rates are held in the afternoon. With the bunching of tutorial classes on certain days or half-days, it is likely that some students will have quite skewed timetables, with a run of consecutive classes. This is likely to have implications for absence behaviour: students do report that they absent themselves from class in order to generate a free hour in an otherwise congested day. From our data, it would be possible to construct the timetable for each of our students and investigate such congestion issues. However, in the current paper we do not address this specific issue of constraints generated by the distribution of classes across the individual students' weekly timetables.

A further tutorial characteristic concerns the students' reported evaluations of their tutors. At points during the academic year – prior to the final examinations – students record their subjective (and anonymous) evaluations of the quality of their tutors' teaching on a scale of 1-5, where 1 is excellent and 5 is poor. The average mark for the tutor across all of their students on the 2nd year courses is 1.98. There appears to be no clear pattern between absences and tutor score in the raw data.

Figure 4 plots the distribution of 2nd year performance in the core courses for those students who missed no classes and also for those who missed at least one class. We note that not only is the distribution of marks shifted towards the left for those students who missed at least one class, but the shape of the distribution is also different. These are raw data plots; we turn next to multivariate analysis to account for the effect of absenteeism on the location and shape of the students' performance distribution.

5. Econometric Model and Results

The data enable us to observe the performance and tutorial absence of students in each of their three compulsory 2nd year courses. With 444 students over three cohorts in each course, we have a panel of 1332 observations. The dependent variable p in our model is the student's end-of-year performance measured as a score out of 100 for each of the student's three core 2nd year courses. The main explanatory variable of interest is the proportion of tutorial classes missed during the second year of study, called *absence*, a, defined with respect to each of the three core courses.

As discussed in section 3, performance is likely to be affected by factors such as ability, effort and motivation. These variables are unobserved in our dataset and nor do we have reliable proxies for them: we do include a variable for prior academic performance in our analysis, but do not claim that this is a good proxy for ability, partly because it is itself an outcome measure and not purely an input. That factors such as ability are not observed creates the potential problem of an omitted variable bias in the estimation of the effect of absence on performance as ability, *inter alia*, is likely to be correlated with absence. For example, in relation to Figure 1 (c) in section 3, we hypothesized that higher ability students might have a higher marginal productivity of class attendance and hence have fewer optimal absences. In this case, OLS would generate an over-estimate of the effect of absence on performance. We also hypothesized that the effects of absence on performance would be likely to vary with ability for a variety of reasons.

In order to explore possible heterogeneous effects, we use the quantile regression (QR) framework.[11] Unlike the conventional least squares framework that looks at the effect on the conditional mean, the QR framework allows for differing effects of a on different parts of the p distribution, thus enabling us to look at the effect of a on the location, scale and shape of the p distribution. This is consistent with the theoretical discussions provided in Section 3, where the personal characteristic is allowed to interact in a non-trivial way with a and other covariates to have an effect on p.

The next econometric issue we have to address is how to estimate the causal effect of absences on performance. We have two potential problems of endogeneity bias: (i) 'Class selection' bias arising if students can choose their tutorial class, i.e. self-select into the most preferred time slot; (ii) 'Ability/motivation/effort' bias arising if, say, more able students are

---

[11] For a comprehensive introduction to the topic, see Koenker (2005).

both less likely to be absent from class and more likely to perform well. We use the fact that the students are randomly allocated to classes where some of the class-timings are unpopular, to address the bias arising from (i). We could minimise (ii) with good proxies for ability (Romer, 1993). We have information on first year performance and we do include this in the model. However, this may not necessarily be a good measure of ability. Hence – and in absence of other good measures of ability – we implement a novel two-step procedure.[12, 13] We estimate the unoberved characteristic from a reduced form equation for absences using the panel nature of our data and use this estimated variable in the QR equation in the second step. Prior to discussing the details of the procedure, we provide some information on the instruments we use to identify the causal effect of absences on performance.

***Instruments*:** The instruments we use are the days of the week and the time slots of the tutorial classes. These timings are centrally fixed and extremely difficult to change due to various constraints. Days and time slots of classes vary in popularity. For example, mornings are regarded as unattractive hours for a class, especially the 9am slot: 2nd year students live off campus and have commuting journeys to campus. There is likely to be little variation across students in the length of the commute. As there is no teaching on a Wednesday afternoon, some students elect not to attend the university on that day, thereby skipping classes: Wednesday absence rates are higher than those for other days. The combination of central timetabling and random assignment of students to class, together with the differential attractiveness of different class time slots, generates exogenous variation in class attendance.

---

12 After the seminal paper by Koenker (2004), there is now an expanding literature which deals with the estimation of quantile regression when longitudinal data are available. The models estimated allow the unobserved individual heterogeneity to have only a location shift effect. See Lamarche (2007) for an application of this technique. Abrevaya and Dahl (2006) and also Gamper-Rabindran, et.al. (2007) extend the model in different ways to allow for correlated random effects in the spirit of Chamberlain (1984).

13 This method of estimating unobserved individual specific heterogeneity is routinely performed by market analysts in issuing targeted vouchers to specific customers based on their observed behaviour (Rossi, *et al*, 1996). Also see Train (2003) Chapter 11.

We create different indicator variables to pick up these effects and use them in the reduced form equation for absence, a.

Model: The τ^{th} ($0<\tau<1$) conditional quantile of the p distribution for the j-th course, j = 1,2,3, for the i-th individual (i=1,.., 444) is specified as:

$$Q_{p_{ij}}(\tau|\, \mathbf{x}_{ij}, a_{ij}, v_i) = \beta_0(\tau) + \mathbf{x}_{ij}'\boldsymbol{\beta}(\tau) + a_{ij}\,\gamma(\tau) + v_i\,\delta(\tau) \qquad (3)$$

where v_i is an unobservable individual characteristic. Note that, unlike the standard assumption that v_i acts only as a location shift, here the effect of v_i is allowed to be different in different parts of the distribution. The full set of additional controls used is discussed in the results section.

In linear panel data models, the unobserved v_i would be eliminated via the within group or the first difference transformation prior to the application of OLS. However, these procedures are not open to us in the QR framework.[14] We exploit the panel nature of our dataset to estimate the unobservable v_i from a reduced form equation for absences and use the estimated v_{i} in place of the unobserved v_i in (3) in order to account for correlation between absences and unobservables. The details of the procedure used are summarized below. We use the timetabling characteristics of class meetings as instruments for the identification of the causal effect of absence on performance. We have described in section 4 how allocation to class is based on random assignment. The implicit assumption we are making is that class meeting times are likely to affect the absence choice of students but have no direct effect on performance once conditioned on absences. For example, we assume that conditional on absences, the time of day variables do not have any additional effect on performance. We provide tests of this hypothesis, as discussed below.

---

[14] For example, the quantile of the difference is not equal to the difference of the quantiles.

Step 1

We have absenteeism information on the three core courses for each individual i. Let a_{ij} be the proportion of tutorials missed by individual i in course j, where:

$$a_{ij} = \boldsymbol{z_{ij}}\delta + v_i + u_{ij} \tag{4}$$

v_i is an individual-specific unobservable random effect which is assumed to have the same effect on all core courses. $\boldsymbol{z_{ij}}$ is a vector of covariates which includes our instrument which is the timings of classes. As discussed previously, students are randomly allocated to classes, which are centrally timetabled to avoid timetable clashes.[15]

Given the nature of our dependent variable a, we estimate the model as a Panel Tobit model. As detailed in the Appendix, using this model, we obtain an estimate of ν for each individual ($\hat{v}_i$) which is known as the empirical Bayes prediction or shrinkage estimate (Goldstein, 2003). We use gllamm (2004) to obtain our estimates.[16]

Step 2

We use the $\hat{v}_i$ in place of v_i in equation (3). Stata 9 was used to estimate the coefficients of our QR model. The standard errors were calculated by the bootstrap method using 500 replications which accounts for clustering at the individual level.

---

[15] Note that although the assignment of students involves a role for initial of surname (and this might undermine the randomness property), this is anyway disturbed by different timetable clashes for different students (generated in a variety of ways, including through different optional module choices), requiring iterations of the assignment.

[16] Results from a model where we use a random-effects ordered probit for the first stage in order to estimate the unobserved individual-specific term produced very similar results and are therefore not reported here. We also used a generalisation of the two-step method proposed by Buchinsky (1998, 2001) and included the generalised residual and its square from the first-stage model estimates in the second-stage quantile regression model. The results are not reported here since they were very similar to the results which we do report.

5.1 Results for Step 1: model of absenteeism

Table 3 reports the results of the random effects Tobit model based on the absence variable, *a*. The results suggest that students who performed well in their 1st year Statistics course tend to have lower absenteeism in their 2nd year. There are some interesting findings in terms of the information we have on the time of the tutorial class. We find that tutorial absence is markedly higher for the 9am class and, to a lesser extent, for all morning classes (those starting before midday) compared to (pre-5pm) afternoon classes. Relative to Thursday classes, there are significantly higher absences in Wednesday classes. Monday, Tuesday and Friday classes attract higher attendance than do Thursday classes. We note that lower absence is associated with a lower tutor score: that is, with a more favourable evaluation of the tutor by the students.

On personal characteristics, we find that female students miss fewer classes than do male students. Overseas students are found to miss more classes compared to home (EU) students, holding all else constant. Attendance at classes differs significantly only for one of the subjects, with Econometrics experiencing markedly greater absenteeism.

The variance of the individual-specific random effects is found to be significantly different from zero. A plot of the Bayesian estimate of the random effects $\hat{v}_i$ is given in Figure 5a. As expected, it is unimodal and centred around 0. Figure 5b plots the estimated random effects but now separately for students who had no absences in a particular course and for those who had at least one absence in a particular course. Interestingly, the density plot is shifted towards the negative part for those students who had no absences. This implies that these students have a characteristic which makes them less prone to absenteeism.

We consider two empirical models: (1) the benchmark models treating absenteeism as exogenous, and (2) the models treating absenteeism as endogenous. Prior to discussing our results from the QR model, we discuss the results from the conditional mean models.

5.2 Results for models of conditional mean performance

Conditional mean model parameter estimates are given in Table 4. Column [1] provides the simple OLS estimates. If there is neither unobserved heterogeneity nor self-selection with regard to class attendance and if there is only a location shift effect of absences on performance, then the OLS would provide consistent estimators. From the simple OLS estimates, we find that a 10% class absence rate is associated with a statistically significant 1.4 percentage point reduction in measured performance.[17] Column [2] has the within-group (WG) estimates. These are the OLS estimates of the model that allows for an unobserved individual-specific effect (more commonly known as FE estimation). If students choose optimally to absent themselves from classes and if this can be captured fully by unobserved individual-specific characteristic, then a 10% rate of class absences is found to be associated with only a barely significant 0.5 percentage point reduction in performance – a very modest effect. Thus, as expected, OLS appears to over-estimate the effect of absence on performance. This is consistent with the hypothesis, discussed in section 3, that more able students have a higher marginal productivity of class attendance. The simple OLS estimates over-state the effect of absence on performance as absence is lower and performance is greater for more able students. In the case of the WG estimates – which control for unobserved characteristics – ability bias is corrected for, at least to some extent, and the estimated coefficients are therefore reduced in absolute size.

---

[17] We have investigated the possibility of non-linearities in the effect of absence on performance but, unlike Durden and Ellis (1995), for example, who find threshold effects, we do not find such discontinuities. Krohn and O'Connor (2005) also examine possible nonlinear effects of ability on performance.

We turn now to consider the various instrumental variable estimations. In Table 4 column [3] we use the days and the times of classes as instruments for absences. As a benchmark, although the more appropriate first stage regression is a Tobit, given the nature of the absence variable, we provide the results from a simple first stage OLS reduced form here. The F-test for the joint significance of the instruments in the first-stage of the regression indicates a rejection of the null generally at 5% significance level. That is, instruments are important in the first stage regressions. The Hansen J test for over-identification is also not rejected. However, we note that absences are now estimated to have an insignificant effect on performance. Although the estimated coefficients are large and positive, the standard errors are now also quite large.

In column [4] of Table 4, we report the results from an IV regression where we use the predicted absences from the Tobit regression as an instrument for absences. The estimated penalty is now slightly higher at 1.6 percentage points for 10% of class absence. The other coefficient estimates are very similar to the OLS estimates. Since we have an over-identified equation, we are able to estimate the model using either the times of classes or the days of classes as instruments, separately. We find that a Wald test of zero coefficients on the time of classes as additional regressors in [4] has a chi-squared p-value of 0.203 (see Table 4 column [4]). Similarly, the p-value for the test of zero coefficients on the day of classes as additional regressors in [4] is 0.108 (see Table 4 column [4]). Before turning to the two stage least squares estimates, we note the fact that since the times or days of classes are insignificant once we control for absences, the effects of class timings on performance are stemming solely from their effects on absences.

In Column [5] of Table 4 we report the results from a similar estimation where we now replace the absence variable by the predicted values. The equation is just identified. We find a slightly higher penalty attached to absences than found from Column [1]. A 10% rate of class absences reduces the expected mark by about 1-2 percentage points.

Summarising, we find, as expected, that allowance for unobserved heterogeneity is important in the model of conditional mean performance and that the estimated penalty associated with absence is over-estimated in the simple OLS specification when no allowance is made for unobserved heterogeneity.

5.3 Results from Quantile Regressions

We now turn to the results based on quantile regressions. These models allow for a heterogeneous effect of the covariates on the various parts of the performance distribution. Again, for the purposes of comparison, we present results from models that treat absence as exogenous as well as from those in which absence is endogenous. As described above, we account for the endogeneity of absences by including an estimated unobserved characteristic from a first stage Tobit regression for each individual. The individual characteristics are allowed to have different effects on different parts of the conditional quantile functions. The results are presented in Table 5. The estimated effects of absences along with the 95% confidence intervals are presented in Figures 6a and 6b.

We first discuss the results from the model that treat absenteeism as exogenous, as presented in Table 5 panel 1 and illustrated in Figure 6a. The figure suggests the presence of some heterogeneity in the effects of absences on performance, with the penalty attached to missing classes greatest for students in the lowest decile. It is estimated that for students in the top 10% of the performance distribution being absent from 10% of classes is associated with around a 1 percentage point loss in the subject score, *ceteris paribus*, compared to about 2 percentage points for those students in the bottom 10% of the performance distribution. Interestingly, the penalty for missing classes is very much smaller on the conditional mean performance from the within group estimation (column [2] Table 4). Also, we note that the OLS estimate (column [1] Table 4) is very similar to the estimate from the QR model at the median. We conclude from this that there is some effect of unobserved individual characteristics that needs to be taken into account. Comparison of WG model results with those derived from the QR model indicates not only that controls for unobservables might be relevant, but also that allowance for heterogeneous effects of covariates on different parts of the performance distribution is important.

Results for other variables included in the model are also informative. We find that females perform worse than males and that this negative effect is increasing across the quantile index. While this result might be viewed as surprising, McNabb *et. al*. (2002) find that while, on average, females perform better than males at University, they are significantly less likely to obtain a first class degree than are males. Given that the criteria for admission into the Department of Economics is AAB at A-level (or equivalent),[18] we are by definition observing already high performing individuals.

The coefficients on variables reflecting 1st year performance are as expected, although, in general, these do not show much variation across the quantiles. More interestingly, we find that the effect of the tutor score variable is negative, with students performing better in classes where the tutor has a lower score (better evaluation), with this negative effect becoming smaller across the quantile index.

We now turn to QR models that allow for the possibility of endogenous selection in the number of classes the student chooses to attend, presented in Panel 2 of Table 5. For the models presented in this panel, an estimated 'individual-specific characteristic' variable is included, although this variable is significant only in the bottom half of the distribution.

---

[18] A-levels are graded A through to E, with A being the top grade.

While the coefficient estimates on most of the covariates are again very similar in terms of magnitude and significance to those reported in Panel 1, the coefficient estimates on the absence variable are markedly reduced. Interestingly, we now find that the effect of missing classes is estimated to be significant only for high 'ability' students. Missing 10% of classes is estimated to be associated with around 1-2 marks for this group of students. The estimated effect of missing class for the low 'ability' students is insignificantly different from zero once selection is accounted for in the estimation.

What do we conclude from this set of results? The results indicate that when we take into account that individuals might be self-selecting to miss class, there is a penalty attaching to absence - but only for more able students. The fact that the estimated effects of absence on performance are weaker (that is, less negative) in the endogenous than in the exogenous model is consistent with our theoretical discussion of Figure 1c, where we predicted just such a negative ability bias – and hence that ability disproportionately raises the marginal product of class attendance, relative to its effect on the marginal product of other forms of study. That the average effect of absence on performance is still negative in the endogenous model suggests that there is no particular excess supply of classes: this is evidence against the case presented in Figure 1a, in which the causal effect of absence on performance would be predicted to be positive. Similarly, the case presented in Figure 1d – in which more able students allocate less time to total study effort – is not supported: we have shown that absence rates are lower for higher performing students.

To the extent that we are correcting fully for ability bias, the results indicate that absences have a negative causal effect on performance: this is consistent with the case presented in Figure 1b, with the implication that the number of classes offered is lower than the student's optimising number. We note, however, that the quantile results for the endogenous model show that the significant negative effects of absence on performance hold only for upper quantiles. This is consistent with a combination of the cases depicted in Figures 1b and 1c, with more able students attending more classes than the less able (Figure 1c) and with this optimal number – for the more able only – being greater than the number supplied (as in Figure 1b).

5.4 Models without controls for first year performance

In all the models reported above, first year performance has been included as a control variable. One of the motivations for this is that it acts as a proxy for ability, thereby offering the prospect of offsetting the missing variable bias in the estimate of the effect of absence on performance arising from the correlation between ability and absence. On the other hand, however, as observed by Romer (1993), inclusion of prior performance risks under-estimating the absolute size of the true effect as past performance is itself likely to be the outcome of the same kind of interaction between unobserved ability and past absence. In the light of this, we have re-estimated some of our models – those involving QR methods – with the exclusion of controls for first year performance. The results with regard to the effects of absence on performance across the quantiles are illustrated in Figure 7: 7a presents the results for the model in which absence is treated as exogenous and 7b the results for the endogenous case.[19]

Comparing the results described in Figure 7a with those from Figure 6a, we see that the general pattern is similar, but with a sharp difference in the intercept: whereas the average effect is around -15 in 6a, it is now around -30 in 7a. This is as one would expect: to the extent that the inclusion of past performance represents a control for ability, the estimated magnitude of the effect of absence falls from that shown in 7a to that shown in 6a with the inclusion of the measure of first year performance. Romer’s argument is that the true effect is

---

[19] Full results are available from authors on request.

likely to be somewhere between the two. From the QR models, this effect seems to be broadly constant across the quantiles, in line with a simple intercept difference.

In contrast, there is no obvious intercept difference arising from the exclusion of controls for past performance in the case of the QR model which controls for the endogeneity of absence. Nor are there clear differences in the pattern across the quantiles. Why should the inclusion of controls for past performance matter in the exogenous case but not in the endogenous? The difference which arises in the exogenous case does so because ability is not controlled for when first year performance measures are excluded in the analysis: hence there is an ability bias which is corrected for (over-corrected for, in Romer's terms) when past performance is included. In the endogenous case, our argument is that the effects of ability on absence, and hence indirectly on performance, are being captured by the $\hat{v}_i$ term, yielding quantitatively similar results whether or not first year controls are incorporated.

6. Concluding remarks

There is now a significant body of work which attempts to delve inside the educational 'black box' in order to deepen the understanding of the processes by which human capital is acquired in learning environments. Much of this work has focused on the importance of factors such as class size and peer effects and has concentrated on educational attainment of pupils in compulsory schooling, with less attention paid to higher education. This is surprising given that there has been a growing policy focus on higher education, with governments in many countries viewing the university sector as an important driver of research, development and growth. The nature of higher education is likely to be fundamentally different from compulsory primary and secondary education, with greater student autonomy in study in higher education. Our attention in the current paper has focused on the causal impact of class absence on student performance and on variations in the estimated effects according to particular student characteristics. The analysis has also incorporated a study of the determinants of class absence.

Our empirical analysis makes use of rich administrative panel for economics students at a UK university. We have exploited a number of key features of the data-set: (i) the panel nature of the data enables us to control for unobserved heterogeneity across students and hence for endogeneity between attendance and performance stemming from the likely influence of effort and ability on both; (ii) random assignment of students to classes avoids the potential endogeneity problems that occur when students can self select into classes; (iii) information on the time slots of the classes in the student's weekly timetable serves as a source of exogenous variation in a student's absence and hence yields potential instruments for the identification of a causal effect of absence on performance.

Our theoretical framework has outlined various possible links between absence and performance and how these might vary with student characteristics such as ability. The approach predicted that absence rates will be lower for more able students in the case in which ability is relatively highly correlated with the marginal productivity of class attendance rather than with other factors in the educational production function. We find empirical support for this prediction as, on average, absence is lower for students with better prior performance – a proxy for ability. In this same case, the model also predicted a negative relationship between absence and performance: both because of selectivity bias and because the marginal product of attendance is greater for more able students. The implications are that: one should find a negative association between absence and performance; that this negative effect should be moderated when we correct for endogenous selection; and that the causal negative effect should be stronger for more able (and hence better-performing) students.

We have reported results consistent with each of these predictions, from which we conclude that a major driver of patterns in class attendance are differences in marginal productivity of class attendance across students. In the absence of controls for unobserved heterogeneity, we find that there is a significant effect of class absence on the student's performance. This effect is weakened – though remains significant – when controlling for unobserved individual effects. We interpret this as consistent with the presence of ability bias in the naïve regression which fails to model the endogenous nature of absence and performance. In a quantile regression specification, it emerges that the adverse effect of missing class is greater for better-performing students, consistent with our hypothesis that effects are likely to vary with factors such as student ability.

What are the policy implications of our findings? We think that there are several. First, the evidence is consistent with the view that class attendance is a productive activity – the estimated causal effect of missing class is negative. Second, theory suggests that grade-maximising students might optimally choose to miss classes and hence that making classes compulsory could be inefficient. However, there is no evidence that missing class is associated with better performance, as would be implied in the model with compulsion and excess classes: compulsion does not seem to be creating problems – from the data we have examined, the administrative regime seems sufficiently flexible as to permit privately optimal choices by students. Third, the evidence suggests that class attendance is particularly productive for better-performing students: one response would be to offer additional, voluntary classes with these students in mind.

Appendix

Bayesian estimate of the random effects

v_i is the unobserved individual specific random effects in the model (see equation (2)). Then,

$f(v_i|data) = f(v_i|a_{i1}, a_{i2},\ldots,a_{im}) = f(a_{i1}, a_{i2},\ldots,a_{im}| v_i)f(v_i)/ f(a_{i1}, a_{i2},\ldots,a_{im})$

Thus,

$$E\left(v_i \mid a_{i1},...,a_{im}\right) = \int v_i \; f(v_i \mid a_{i1},...,a_{im})dv_i$$

$$= \frac{\int v_i \; f(a_{i1},...,a_{im} \mid v_i) f(v_i)dv_i}{f(a_{i1},...,a_{im})}$$

$f(a_{i1},...,a_{im} \mid v_i)$ is the conditional likelihood and $f(a_{i1},...,a_{im})$ is the marginal likelihood which are obtained during the maximising of the likelihood function. The estimated $E\left(v_i \mid a_{i1},...,a_{im}\right)$ is known as the Bayesian shrinkage estimator.

References

Abrevaya, J. and Dahl, C. (2006). The effects of smoking and prenatal care on birth outcomes: evidence from quantile estimation on panel data. Forthcoming *Journal of Business and Economic Statistics.*

Buchinsky, M. (1998). Recent advances in quantile regression models. *Journal of Human Resources*, 33, 88-126.

Buchinsky, M (2001). Quantile regression with sample selection: estimating women's return to education in the U.S. In Fitzenberger, B., Koenker, R., and Machado, J.A.F. (Eds.) (2001) *Economic Applications of Quantile Regression*, Physica-Verlag.

Burtless, G. (ed.) (1996). *Does money matter? The effect of school resources on student achievement and adult success.* Brookings Institution Press. Washington D.C.

Chamberlain, G. C. (1984). *Panel Data*. Handbook of Econometrics, Volume 2, eds. Z. Griliches and M. Intriligator. Amsterdam: North-Holland, 1984, 1247-1318.

Chen, J. and Lin, T-F. (2006). "Class attendance and exam performance: A randomized experiment". Available at SSRN: http://ssrn.com/abstract=908923

Cohn, E. and Johnson, E. (2006). "Class attendance and performance in Principles of Economics". *Education Economics*, 14, 211-233.

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

Dobkin, C., Gil, R., and Marion, J. (2007). Causes and consequences of skipping class in college. Mimeo, UC Santa Cruz.

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

Douglas, S. and Sulock, J. (1995). “Estimating educational production functions with correction for drops”, *Journal of Economic Education*, 26, 101-112.

Durden, G. C. and Ellis, L. V. (1995). The effect of attendance on student learning in Principles of Economics. *American Economic Review*, 85, 343-346.

Ehrenberg, R. G., Brewer, D., Gamoran, A. and Willms, J. D. (2001). Class Size and Student Achievement. *Psychological Science in the Public Interest*, 2 1-30.

Gamper-Rabindran, S., Khan, S. and Timmins, C. (2007). The Impact of Piped Water Provision on Infant Mortality in Brazil: A Quantile Panel Data Approach. September. Available at SSRN: http://ssrn.com/abstract=1129044. Last accessed on 8th July 2008.

Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd Edition, Kendall’s Library of Statistics 3, Arnold, London.

Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G. (2003). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, 18, 527-544.

Hoxby, C. M. (2000). The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation. *Quarterly Journal of Economics*, 115, 239-285.

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*, vol. 34(3), pp. 311-326.

Koenker, R (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91, pp.74-89.

Koenker, R (2005). *Quantile Regression*, Cambridge University Press.

Krohn, G. A. and O’Connor, C. M. (2005). “Student effort and performance over the semester”, *Journal of Economic Education*, 36, pp. 3-28.

Krueger, A. B. (2000). An Economist's View of Class Size Research. In *How Small Classes Help Teachers Do Their Best,* ed. Margaret C. Wang and Jeremy D. Finn. Philadelphia: Temple University Center for Research in Human Development in Education.

Lamarche, C. (2007). Private school vouchers and student achievement: a fixed effects quantile regerssion evaluation of the Milwaukee parental choice program. Mimeo, University of Oklahoma.

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

McNabb, R., Pal, S. and Sloane, P. J. (2002). Gender Differences in Educational Attainment: The Case of University Students in England and Wales. *Economica*, 69, 481-503.

Marburger, D. R. (2001). “Absenteeism and undergraduate exam performance”, *Journal of Economic Education*, 32, 99-109.

Marburger, D. R. (2006). “Does mandatory attendance improve student performance”, *Journal of Economic Education*, 37, 148-154.

Martins, P. and Walker, I. (2006). Student achievement and education production: a case study of the effect of class attendance. Mimeo, University of Warwick.

Neri, F. and Meloche, Y. (2007). “The impact of lecture attendance on academic performance in a large first year Economics course”. Available at SSRN: http://ssrn.com/abstract=975573

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

Romer, D. (1993). Do students go to class? Should they? *Journal of Economic Perspectives*, 7, 167-174.

Romer, D. (1994). Response in ‘Correspondence’, *Journal of Economic Perspectives*, 8, no.3, 214-215.

Rossi, P., McCulloch. R. and Allenby, G. (1996). The value of household information in target marketing, *Marketing Science*, 15, 321-340.

Stanca, L. (2006). The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics. *Journal of Economic Education.* 37, 251-266.

Stinebrickner, R. and Stinebrickner, T.R. (2008). The Causal Effect of Studying on Academic Performance. *The B.E. Journal of Economic Analysis & Policy*, Vol. 8: Iss. 1 (Frontiers), Article 14.

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

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK.

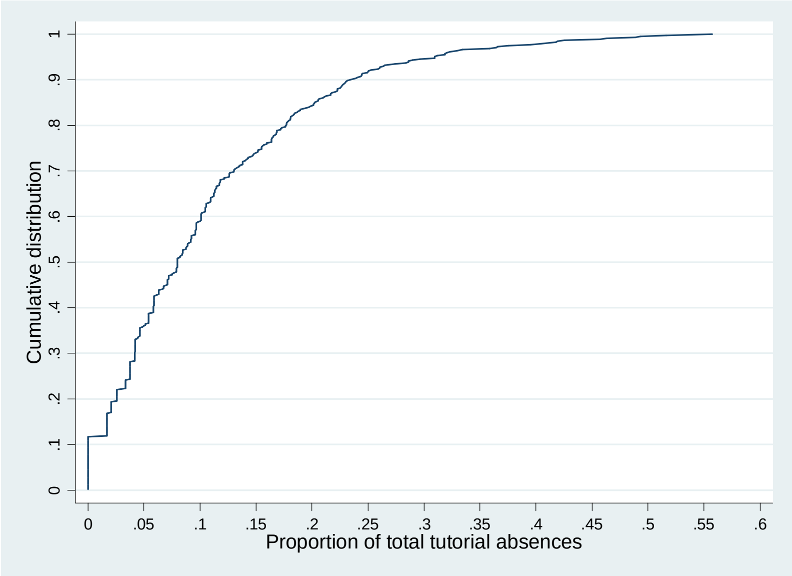

Figure 2 – CDF of absences across 3 core courses
(Unit of analysis is student)

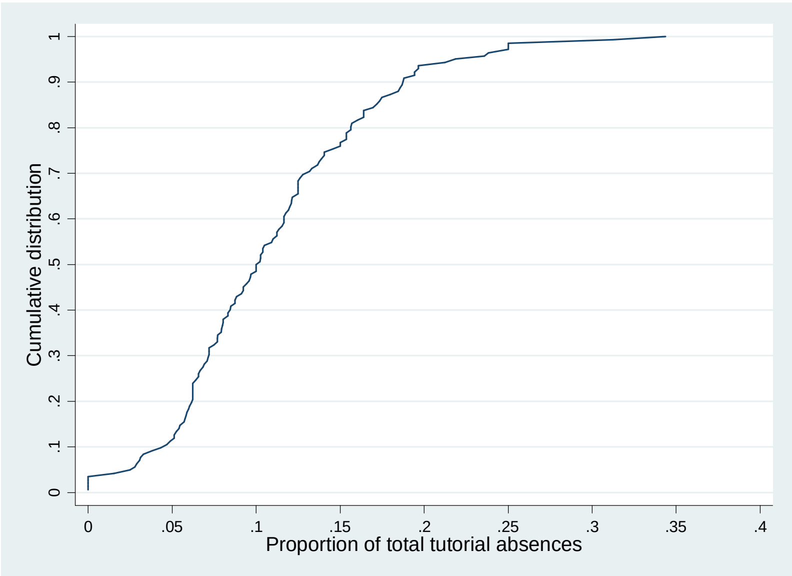

Figure 3 – CDF of absences in each tutorial class
(Unit of analysis is tutorial group)

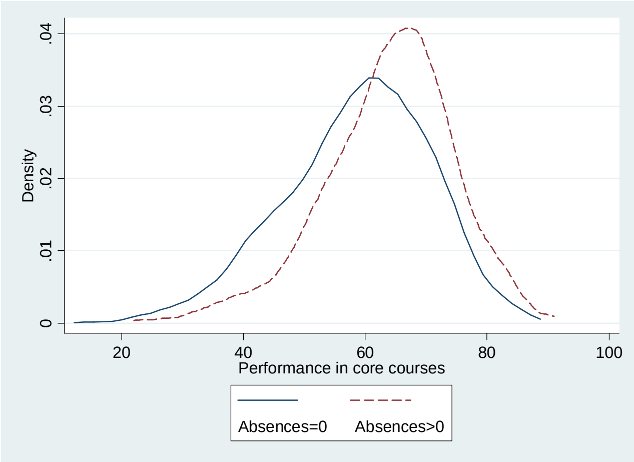

Figure 4 - Performance in core courses by absence category

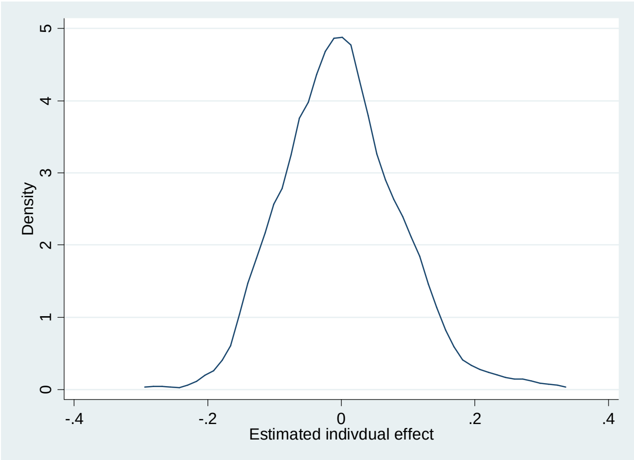

Figure 5a – Bayesian estimate of the individual unobserved effect

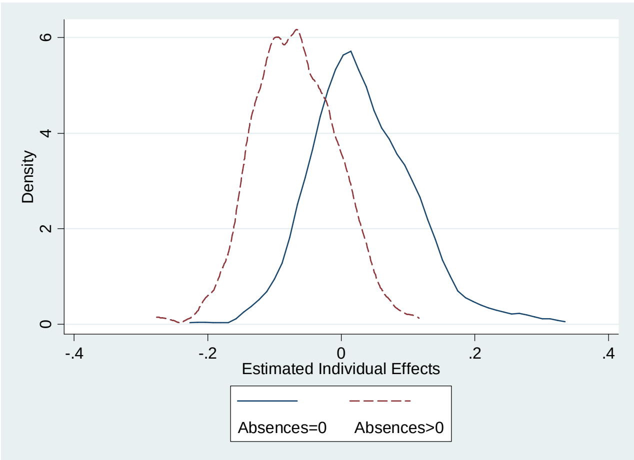

Figure 5b - Bayesian estimate of the individual unobserved effect by absence category

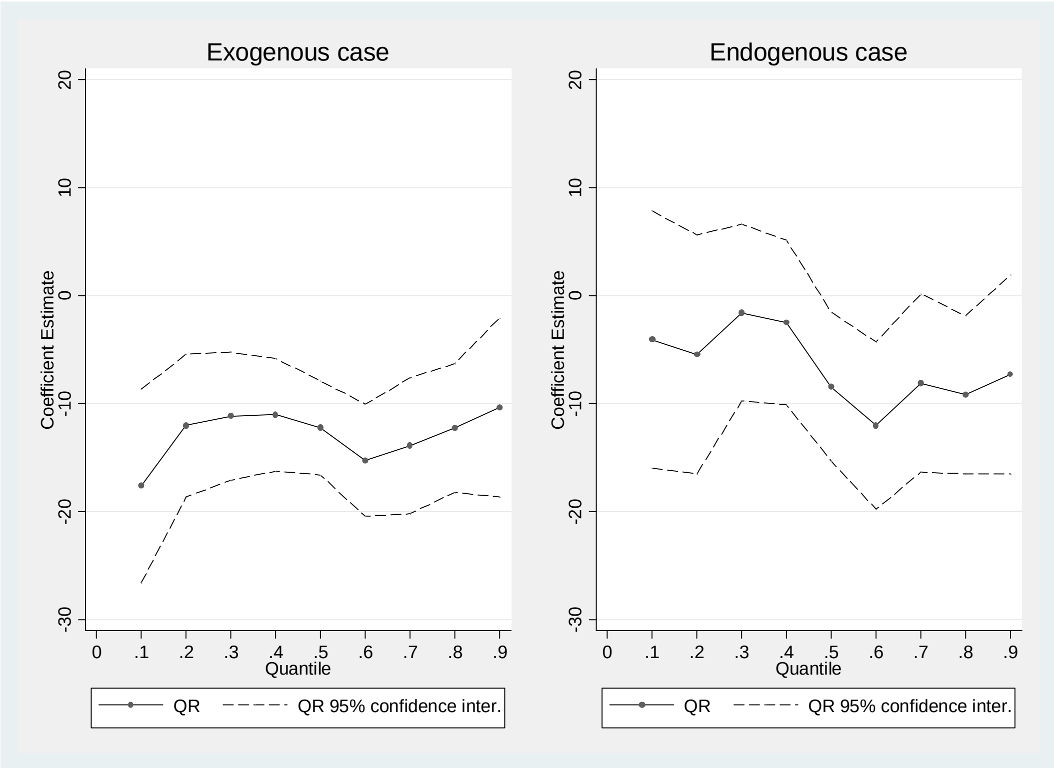

Figure 6 Quantile results for the effect of absence on performance

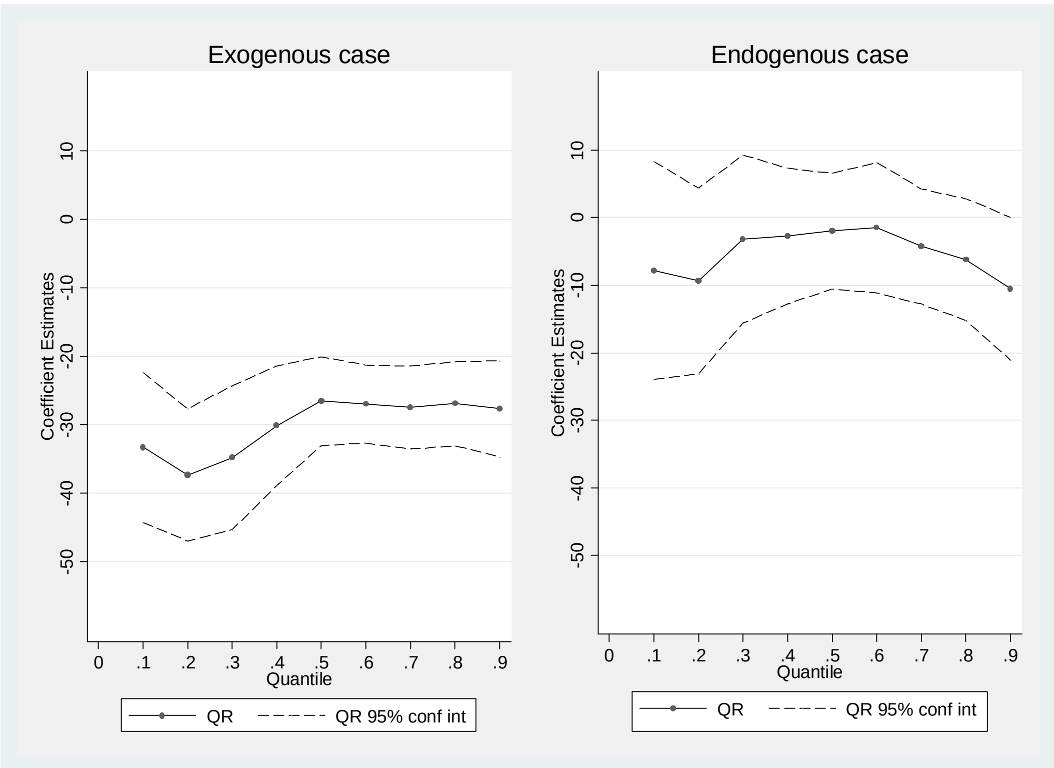

Figure 7: Quantile results for the effect of absence on performance
(without control for first year performance)

Table 1: Student level summary statistics

	All		Absences <4%		Absences (4-8]%		Absences (8-15]%		Absences >15%	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Female	0.34		0.39		0.35		0.35		0.28	
Non-UK fee student	0.32		0.25		0.31		0.30		0.43	
Cohort 04/05	0.30		0.22		0.29		0.30		0.41	
Cohort 05/06	0.36		0.38		0.41		0.39		0.26	
Cohort 06/07	0.34		0.40		0.30		0.31		0.33	
Performance: average score	60.33	10.14	65.28	8.43	62.43	9.65	58.66	9.14	54.63	9.96
Performance: 1st year score										
Maths	68.01	15.46	73.21	12.96	69.70	16.30	65.34	13.02	63.29	17.26
Stats	64.05	15.48	71.18	14.05	66.57	15.46	61.71	12.87	56.24	15.13
Other Total Average	62.46	8.38	65.69	7.68	62.42	8.41	61.69	7.88	59.69	8.46
Resit dummy	0.12		0.06		0.09		0.11		0.21	
Number of students	**444**		**125**		**101**		**102**		**116**	

Table 2: Tutorial group level summary statistics

	All		Absences <6.5%		Absences (6.5-10]%		Absences (10-15]%		Absences >15%	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Tutorial Information										
Proportion of absences	0.11	0.06	0.04		0.08		0.12		0.19	
9:00am tutorials	0.13		0.11		0.06		0.15		0.17	
Other morning tutorials	0.58		0.54		0.52		0.57		0.69	
Afternoon classes	0.29		0.34		0.42		0.28		0.13	
Monday, Tuesday, Friday tutorials	0.46		0.49		0.52		0.46		0.37	
Wednesday tutorials	0.16		0.17		0.09		0.21		0.17	
Thursday tutorials	0.38		0.34		0.39		0.33		0.46	
Tutor score [1=highest,…5=lowest]	1.98	0.57	2.02	0.65	1.94	0.49	1.93	0.54	2.04	0.59
Number of tutorial groups	**142**		**35**		**33**		**39**		**35**	

Table 3:

Random Effects Tobit Model for Tutorial Absences

Variable	**Coefficient Estimate (std errors)**
1st Year Marks	
Maths	-0.000 (0.001)
Statistics	-0.003*** (0.001)
Other	-0.001 (0.001)
Resit dummy	-0.009 (0.020)
Tutorial info (2nd year)	
9am class	0.076*** (0.013)
Other morning tutorials [excl 9:00 tutorials] + 17:00 hours	0.037*** (0.008)
Monday, Tuesday, Friday class	-0.013* (0.007)
Wednesday class	0.025** (0.012)
Tutor Score	0.031*** (0.008)
Personal characteristics	
Cohort 04/05 [base 06/07]	0.048*** (0.015)
Cohort 05/06	-0.027* (0.014)
Female	-0.026** (0.013)
Overseas	0.053*** (0.013)
2nd Year Courses	
Macro Economics	0.003 (0.009)
Econometrics	0.032*** (0.009)
Intercept	0.238*** (0.055)
σ_α (std. Error) [unobserved heterogeneity std dev]	0.103*** (0.005)
σ_u	0.100*** (0.003)
Proportion of error variance attributed to unobs heterog	0.513
Log Likelihood	252.06
Number of Individuals	444
Number of Observations	1332
Number of left censored Observations	429

Notes: (i) The model also includes appropriate controls for another associated economics degree and the core courses in that degree. (ii) ***, **, * significant at 1%, 5% and 10% respectively.

Table 4 – Models of Conditional Mean Absences – Coefficient Estimates (Standard Errors)

	OLS	WG	IV1	IV2	TSLS
	[1]	[2]	[3]	[4]	[5]
Absences	-13.39***	-4.92*	18.74	-16.14***	-15.52***
	(2.50)	(2.98)	(16.76)	(3.62)	(3.54)
Cohort (04/05)	-4.26***		-5.32***	-4.17***	-4.02***
	(0.77)		(1.03)	(0.76)	(0.77)
Cohort (05/06)	-2.69***		-2.13**	-2.74***	-2.83***
	(0.80)		(0.85)	(0.79)	(0.81)
Female	-2.14***		-1.32	-2.21***	-2.22***
	(0.71)		(0.86)	(0.70)	(0.71)
Oversees Fee	-1.60**		-2.92***	-1.49*	-1.34*
	(0.79)		(1.06)	(0.79)	(0.80)
1st Year Marks - Maths	0.07*		0.07*	0.07*	0.06*
	(0.04)		(0.04)	(0.04)	(0.04)
Stats	0.25***		0.32***	0.24***	0.23***
	(0.04)		(0.05)	(0.04)	(0.04)
Other	0.37***		0.37***	0.36***	0.36***
	(0.07)		(0.07)	(0.06)	(0.07)
Resit dummy	1.00		0.96	1.01	0.79
	(1.41)		(1.55)	(1.39)	(1.42)
2nd Year Courses - Macro	-0.82	-1.15**	-0.98	-0.81	-0.84
	(0.64)	(0.57)	(0.66)	(0.64)	(0.64)
Econometrics	4.12***	3.83***	3.80***	4.15***	4.32***
	(0.54)	(0.54)	(0.57)	(0.54)	(0.55)
Tutor Score	-0.95**	-1.54***	-1.40**	-0.91*	-0.77
	(0.47)	(0.48)	(0.56)	(0.47)	(0.49)
Intercept	22.34***	62.53***	14.81***	22.99***	23.63***
	(3.74)	(1.17)	(5.44)	(3.79)	(3.87)
Test of excluded instruments F [p-value]			7.53 [0.00]		
J Statistic for over-identification [p-value]			2.38 [0.50]		
Partial R^2 of excluded instruments			0.021	0.618	0.617
Joint test of including time of day as additional regressors in [4] χ^2 pvalue				0.203	
Joint test of including day of week as additional regressors in [4] χ^2 pvalue				0.108	

Notes: (i) Absences is the proportion of missed tutorials. (ii) The model also includes appropriate controls for another associated economics degree and the core courses in that degree. (iii) ***, **, * significant at 1%, 5% and 10% respectively. (iv) Standard errors reported in columns [1] to [4] are clustered at the individual level and in column [5] are block bootstrapped to account for clustering at the individual level (because the predicted values from the first stage Tobit regression is used as a regressor). (v) The set of instruments used are: Both timing and the day of tutorials in column [3]; predicted proportion of absences from the first stage Tobit in column [4], column [5] includes predicted absence. (vi)Test of excluded instrument is the F test that tests for the joint significance of the instruments that are used in the identification.

Table 5 – Parameter Estimates

PANEL 1 Exog Absences	**Quan 0.1**	**Quan 0.2**	**Quan 0.3**	**Quan 0.4**	**Quan 0.5**	**Quan 0.6**	**Quan 0.7**	**Quan 0.8**	**Quan 0.9**
	[4]	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**
Absences[a]	-17.61***	-12.03***	-11.15***	-11.03***	-12.23***	-15.26***	-13.89***	-12.24***	-10.35**
Cohort (04/05)	-6.23***	-4.98***	-4.10***	-3.18***	-3.82***	-3.47***	-3.91***	-3.78***	-3.97***
Cohort (05/06)	-4.38***	-2.81***	-2.48***	-2.27***	-2.17**	-1.69**	-1.50*	-1.16	-0.11
Female	-1.71	-0.99	-1.60*	-1.16	-1.37*	-1.73**	-1.96***	-2.24***	-2.35***
Oversees Fee	-2.08	-3.30***	-2.60**	-2.26***	-2.17**	-1.13	-1.18	-0.96	-0.48
1st Year Marks									
Maths	0.09	0.12***	0.12***	0.09***	0.10***	0.09**	0.05	0.03	0.01
Stats	0.29***	0.24***	0.23***	0.25***	0.25***	0.24***	0.24***	0.24***	0.19***
Other	0.30***	0.40***	0.40***	0.36***	0.34***	0.32***	0.33***	0.33***	0.35***
Resit dummy	0.37	0.95	0.49	-0.01	0.81	0.79	0.40	1.55	2.34
2nd Year Courses									
Macro	3.69***	1.98*	0.28	-0.43	-2.16**	-2.81***	-3.62***	-4.48***	-4.95***
Econometrics	9.64***	9.28***	6.81***	5.33***	3.03***	2.18***	0.91	-0.43	-0.86
Tutor Score	-1.29	-2.05***	-1.92***	-1.54**	-1.04*	-0.92	-0.52	-0.19	-0.13
Intercept	10.55*	9.92**	13.59***	18.46***	22.74***	27.22***	31.44***	35.39***	40.84***
PANEL 2 Endog Absences	**Quan 0.1**	**Quan 0.2**	**Quan 0.3**	**Quan 0.4**	**Quan 0.5**	**Quan 0.6**	**Quan 0.7**	**Quan 0.8**	**Quan 0.9**
Absences (endog)	-4.34	-5.51	-2.03	-2.7	-8.25**	-11.84***	-7.95*	-8.96**	-7.12
Cohort (04/05)	-6.92***	-5.37***	-4.01***	-3.59***	-3.60***	-3.70***	-4.17***	-3.87***	-4.21***
Cohort (05/06)	-4.39***	-3.22***	-2.50***	-1.93**	-2.13**	-1.66*	-1.24	-1.1	0.13
Female	-1.73	-1.17	-1.32	-1.14	-1.18	-1.64**	-1.84***	-2.24***	-2.38***
Oversees Fee	-3.19**	-3.56***	-3.17***	-2.63***	-2.08**	-1.3	-1.53*	-1.01	-0.79
1st Year Marks									
Maths	0.11**	0.12***	0.12***	0.10***	0.09**	0.09**	0.05	0.04	0.02
Stats	0.30***	0.26***	0.25***	0.27***	0.25***	0.25***	0.26***	0.25***	0.20***
Other	0.29***	0.40***	0.41***	0.36***	0.34***	0.31***	0.33***	0.32***	0.35***
Resit dummy	0.75	0.93	0.37	0.03	0.63	0.51	0.42	1.31	2.42
2nd Year Course									
Macro	4.09***	1.59	0.77	-0.63	-2.50**	-2.93***	-3.24***	-4.52***	-4.98***
Econometrics	10.42***	8.79***	6.98***	5.00***	2.90***	2.17***	1.07	-0.64	-0.91
Tutor Score	-1.45*	-2.00***	-1.64**	-1.97***	-1.05*	-1.03*	-0.48	-0.31	-0.09
Est'd Indiv Effect	-20.81**	-10.89	-13.52**	-13.34**	-8.58	-4.78	-7.08	-3.86	-6.89
Intercept	7.81	8.26**	10.32**	17.73***	22.69***	26.83***	29.53***	34.76***	39.66***

Notes: (i) ***, **, * significant at 1%, 5% and 10% respectively. (ii) All standard errors are block bootstrapped to account for clustering at the individual level. (iii) The model also includes appropriate controls for another associated economics degree and the core courses in that degree.