University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Jeremy Smith, Wiji Arulampalam and Robin A Naylor
Article Title: Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time-series analysis for UK university students
Year of publication: 2005
Link to published version:
http://dx.doi.org/10.1016/j.econedurev.2004.05.007

Effects of In-Class Variation and Student Rank on the Probability of Withdrawal: Cross-Section and Time-Series Analysis for UK University Students

Wiji Arulampalam

Robin A. Naylor

Jeremy P. Smith

Effects of In-Class Variation and Student Rank on the Probability of Withdrawal: Cross-Section and Time-Series Analysis for UK University Students

Wiji Arulampalam
University of Warwick and IZA Bonn

Robin A. Naylor
University of Warwick

Jeremy P. Smith
University of Warwick

ABSTRACT

Effects of In-Class Variation and Student Rank on the Probability of Withdrawal: Cross-Section and Time-Series Analysis for UK University Students[*]

From individual-level data for nine entire cohorts of undergraduate students in UK universities, we estimate the probability that an individual will drop out of university during their first-year. We examine the 1984-85 to 1992-93 cohorts of students enrolling full-time for a three or four-year course, and focus on the sensitivity of the probability of withdrawal to the individual's prior qualifications relative to those of the other students in their university course. We show not only that weaker students are more likely to withdraw but also that the extent of variation in prior qualifications within the student's university degree course exerts an influence on the individual's probability of withdrawal in a way that varies with the individual's own in-class rank.

Corresponding author:

Wiji Arulampalam
Department of Economics
University of Warwick
Coventry CV4 7AL
United Kingdom
Tel.: +44 (0)24 7652 3471
Fax: +44 (0)24 7652 3032
Email: wiji.arulampalam@warwick.ac.uk

1. Introduction

Since December 1999, the UK Government has published university performance indicators based on statistics such as drop-out rates from higher education institutions (see HEFCE 99/66). A focus of policy and of analysis on university withdrawal rates reflects widespread concern with evidence of a rising drop-out rate among university students. This has occurred during a period of time in which government policy has succeeded in expanding the size of the university student population in concert with strategies both to shift the financial burden of study on to students and their families (see Dearing, 1997) and yet to widen access into higher education. Both of these strategies have potential impacts on the university student dropout rate.

To date, most of the analysis of university attrition in the UK has been based on university-level data (see, for example, Johnes and Taylor (1989, 1990)). Recently, however, researchers have gained access to the full set of individual student-level information stored in the Universities Statistical Records (USR),[1] and have used these data to analyse the issue of student withdrawal. For example, Smith and Naylor (2001a) analyse the determinants of dropping out of the degree programme for students enrolling in the academic year 1989-1990, while Arulampalam, Naylor and Smith (2001) focus on medical student withdrawal. Johnes and McNabb (2001) examine the attrition of students leaving university in 1993, focussing on the influence of student-course matching and of peer group effects.

In this paper, we extend the previous analyses in particular directions. First, we concentrate our attention on the nature of the impact of prior qualifications on the

---

[1] The USR preceded the Higher Education Statistics Agency (HESA) as the depository for all the statistical returns from UK universities.

individual's probability of withdrawing from their university course. We examine the sensitivity of the student's drop-out probability to their relative position in class: that is, to their prior qualifications relative to those of fellow students on their university degree course.[2] In particular, we investigate how the extent of student in-class heterogeneity with respect to prior qualifications impacts on the probability of dropping out. Second, we analyse the extent to which differences by gender in the probability of dropping out are explained by gender differences in observed characteristics. Third, we assemble the data for nine entry cohorts between 1984-85 and 1992-93 and investigate the time-series robustness and trends exhibited by the estimated cross-section results. In particular, we decompose changes over time in the probability of dropping out into that part explained by changing characteristics and that part attributable to changes in estimated coefficients. We focus on the extent to which changes over time in prior qualifications have contributed to changes in the dropout probability.

The importance of prior qualifications of students as a determinant of their drop-out probabilities is well-established in the literature. In the extensive US literature, one of the most influential theoretical explanations of student attrition is the path analysis model of Tinto (1975, 1987). This model suggests that the student's social and academic integration into university is the major determinant of completion, and identifies a number of key influences on integration, such as the student's family background, previous schooling, prior academic performance and interactions between students and with faculty.

---

[2] Throughout, 'class' refers to fellow students on the same degree course and at the same university.

For UK university students,[3] Smith and Naylor (2001a) report that the student's prior qualifications have statistically significant effects on both the male and female drop-out probabilities. Smith and Naylor (2001a) also attempt to take account of the effects of subjects studied prior to university as a further dimension of academic preparedness.[4] Johnes and McNabb (2001) find that the probability of quitting university is higher for students whose prior performance is superior to that of fellow students. This is consistent with the idea that matching is an important element of completion: one could theorise that students who are much better qualified than their peers might quit in order to search for a better match. For the US, Light and Strayer (2000) find that the match between student ability and college quality is a significant determinant of college graduation.

Amongst other things, we are interested in examining whether the extent of variation across students within a university course with respect to prior academic attainment is also an influence on the probability of withdrawal. One of the particular motivations for our concern with this issue relates to the topical debate on policies to widen access into universities in the UK. Policies for widening access – along with policies more generally to increase the size of the graduate population – inevitably involve admitting more students with relatively weak levels of academic attainment prior to university.[5]

Broadly, one could think of two contrasting methods for widening access into the higher education sector for students who have performed relatively poorly at A-level, or

---

[3] Fielding *et al.* (1998) identify prior educational achievements of students as a major determinant of college non-retention for the 16-19 age group.

[4] They also investigate the effect of the closeness of the match between the subject studied at university and the prior subjects studied at A-level, but find relatively few significant effects.

[5] Of course, there are reasons for which prior attainment might be a biased indicator of higher educational academic potential (see, for example, Smith and Naylor (2001b)).

equivalent. One policy would be based on 'specialisation' whereby poor-performing candidates are sorted into university degree courses with low average levels of prior attainment. Such a policy runs counter to Government initiatives to widen access within institutions. A contrasting policy would be one which gave incentives to all institutions to admit a more heterogeneous population of students. One potential danger of this policy is that it might raise the overall drop-out rate if it is the case that student heterogeneity is itself a positive influence on student withdrawal. Research on this issue is therefore important. If it is the case that student academic diversity has an adverse effect, then this will need to be recognized when setting institutional targets and publishing performance indicators. It would also suggest the need for appropriate support policies to accompany potential academic diversity. On the other hand, it is not necessarily the case that diversity impacts adversely on withdrawal: this is the issue with which we are particularly concerned in the current paper.

The rest of this paper is organised as follows. In Section 2, we provide a brief discussion motivating the hypothesis that in-class variation in students' prior qualifications might impact on the individual's drop-out probability. In Section 3, we describe the important features of the data-set and in Section 4 we present the main results of our analysis for the 1993 cohort. Section 5 presents results for earlier cohorts and reports on time-series patterns in the results. Section 6 closes the paper with a discussion and further remarks.

2.   In-class variation: some hypotheses

As we have seen, in the literature on student withdrawal it is generally accepted that prior qualifications have significant effects on the probability of non-completion, reflecting the importance of students' academic preparedness. Both the absolute level of prior academic performance and the relative academic performance of students seem to have an influence on dropout probabilities. Johnes and McNabb (2001), for example, report that students are more likely to drop out the better are their prior grades relative to those of other students at their university or on their degree course. This can be interpreted as a result of the disutility associated with a poor university match. In the UK, students typically apply for a university degree course before they obtain the pre-university examination results on which any offers of a university place will be conditional. Furthermore, students respond to offers of places – being permitted to firmly accept only one conditional offer – before they know their examination results. Students who perform better than they expect or who are relatively risk-averse are likely to find themselves committed to beginning a university degree course for which they are relatively over-qualified. They may then have an incentive to drop out of the course and seek entry to a course with a higher entry standard. One incentive for this would be to improve their chances in the post-university graduate labour market in which job offer probabilities are likely to be a function of the reputation of the job applicant's university course. An important factor determining university reputation is the average score of students in their pre-university public examinations.

In our empirical analysis, we allow for the possibility that the effect of a student's prior performance relative to fellow students is not monotonic. We hypothesise that,

relative to students with average prior performance, students in the lower part of the performance distribution are more likely to drop out as a result of their relatively poor academic preparedness. Similarly, for students with relatively high prior grades we hypothesise that these students might be more likely than the average student to drop out; as a result of superior outside options and the associated incentive to search for a better university/course match. In sum, one possible hypothesis is that the relationship between the withdrawal probability and the prior academic attainment ranking of students might be U-shaped. We allow for this by including dummy variables to indicate the student's in-class ranking with respect to their prior educational attainment.

We also hypothesise that the degree of heterogeneity among students within a university course is likely to influence the individual student's drop-out probability. Furthermore, we allow for the possibility that the influence of heterogeneity on the individual's drop-out probability might itself vary according to the position of the individual in the distribution of students by prior performance. We allow for this by interacting the extent of in-class variation in prior qualifications with the binary indicator variable for the part of the distribution from which the individual is drawn. Our specific hypothesis is that students in the lower part of the prior performance distribution will feel more distant from the mean – and hence potentially be more at risk of failure – the greater is the variation in prior performance within the university course. Thus, the prediction is that such students will have a greater non-completion probability the greater is the in-class variation in A-level scores. Conversely, the hypothesis implies that the average student will feel more protected from the risk of failure the greater is the variation in prior qualifications as this will ensure a longer tail of relatively poorly

prepared students. Thus, the prediction is that students close to the centre of the distribution will be less likely to drop out the greater is the variation in prior performance. The same argument could apply to the students in the upper part of the distribution, but it is also possible that this effect is dominated by a greater propensity of such students to transfer to courses to which they are more closely matched.

We can consider the effects of in-class variation on students' drop-out probabilities more formally. Suppose, for the purposes of illustration, that within a given university course there is a uniform continuous distribution, f(x), of students by prior academic attainment, x, over some interval, r, normalised such that the mean of r is zero. Suppose that some proportion, say 10%, of these students withdraws from their university course for reasons of academic failure or of expected academic failure. If success (that is, continuation of studies into the second year) is correlated perfectly with prior academic attainment, then the withdrawing students will be the lowest decile of the prior performance distribution. If, at the other extreme, withdrawal is randomly assigned, then each individual will have a 10% probability of non-completion. Between these extremes, we could assume that the probability of withdrawal is declining with the individual's rank within the in-class prior performance distribution. Suppose that this 'conditional withdrawal probability' function is linear, as represented in Figure 1, and that the probability of withdrawal is zero for the highest ranked individual ($p_H = 0$). Given a range, r, equal to unity, say, the height of the uniform distribution is also equal to one, and the area under the linear failure-probability function is 0.1. It follows that the probability of withdrawal, p_L, for the lowest ranked individual will be 0.2 and that, for

the median-ranked individual, the probability of failure, $\bar{p}$, will be 0.1. The function cuts the horizontal axis at the upper limit of the range as we are assuming that $p_H = 0$.

Consider now what happens to the conditional probability of withdrawal if there is a mean-preserving reduction in the variation of the distribution, shown in Figure 1 by a halving of the range of the distribution. The range is now 1/2 and the height of the distribution is 2. The area under the linear function defining the conditional probability of withdrawal is still equal to 0.1. That is, $0.5*(2p_L - 2p_H)/2 = 0.1 = (p_L - p_H)/2 = 0.1$ and hence the previous results that $p_L = 0.2$ and $\bar{p} = 0.1$ remain unchanged by the reduction in variance. The result that the conditional probabilities of non-completion are unaffected by the extent of heterogeneity in prior academic attainment is a consequence of the assumption of linearity in the relationship between prior performance and the probability of failing to complete.

[Insert Figure 1 here]

Suppose instead that the conditional withdrawal probability function is convex with weaker students disproportionately more likely to fail to complete the course, as seems plausible. Then it can be shown that for the median individual the probability of failing to complete the university course is decreasing in the extent of in-class variation in prior performance. In the current paper, we illustrate this for a specific functional form.

Consider the convex function $f(x) = \alpha e^{-x}$, where $0<\alpha<1$. Within the interval, (-1/2, 1/2) the area under the function must be equal to 0.1, the student non-completion rate. The area under the function is given by

$$\text{Area} = \int_{-1/2}^{1/2} \alpha e^{-x} dx = \alpha \left[e^{1/2} - e^{-1/2} \right] = 0.1, \tag{1}$$

which implies the particular solution $\alpha = \hat{\alpha} = 0.096$. Consider now the effect of a reduction in the variance of prior performance among students within a given class, or university course. Again suppose the range is halved, implying the height of the uniform distribution is doubled. Assume that the area under the convex function $f(x) = \alpha e^{-x}$ remains equal to 0.1,

$$\text{Area} = \int_{-1/4}^{1/4} \alpha e^{-x} dx = \alpha \left[e^{1/4} - e^{-1/4} \right] = 0.1, \tag{2}$$

implying the solution $\alpha = \hat{\hat{\alpha}} = 0.198$. Evaluating the function $f(x) = \alpha e^{-x}$ at $x = 0$, yields $f(0) = \alpha$. Thus, at the centre of the range, (-1/2, 1/2), $f(0) = \hat{\alpha}$ and we infer that for the median individual the probability of non-completion is given by $\hat{\alpha}$ divided by the height of the uniform distribution, i.e., $\bar{p} = \hat{\alpha} = 0.096$. Similarly, for the range, (-1/4, 1/4), with associated height for the uniform distribution given by 2, the probability of non-completion for the median individual is given by $\bar{p}' = \hat{\hat{\alpha}} / 2 = 0.099$. It follows that

$$\hat{\hat{\alpha}} = \left\{ \frac{e^{1/2} - e^{-1/2}}{e^{1/4} - e^{-1/4}} \right. \quad \hat{\alpha} = 2.06 \hat{\alpha}$$

Thus, $\hat{\hat{\alpha}} > 2\hat{\alpha}$ and hence, under the specific assumption of the convex function, $f(x) = \alpha e^{-x}$, a reduction of in-class variation raises the non-completion probability of the median individual. This formalizes the hypothesis stated above that students close to the centre of the distribution will be less likely to drop out the greater is the variation in prior performance. Similarly, we would expect the same to be true for students in the top percentiles: that is, that the greater is the variance the lower will be their likelihood of failure. The opposite is likely to apply for students in the lowest percentiles of the

distribution by prior performance, as discussed above. Numerical simulation for the specific convex functional form considered in this section corroborates these predictions.[6]

To summarise, under the assumptions specified, the model generates the following predictions. (i) The probability of withdrawal conditional on rank performance will be greater for weaker students (Hypothesis 1), (ii) the conditional withdrawal probability of weaker students will be increasing with the extent of in-class heterogeneity (Hypothesis 2), (iii) the conditional withdrawal probability of median-ranked students will be decreasing with in-class heterogeneity (Hypothesis 3), and (iv) the conditional withdrawal probability of stronger students will also be decreasing in the extent of in-class heterogeneity (Hypothesis 4). This last prediction, however, is likely to be offset to some extent by the incentive we have identified for the relatively strong students to switch to university degree courses with higher mean scores, generating Hypothesis 5, which is the converse of Hypothesis 4 and states that the conditional withdrawal probability of stronger students will also be *increasing* in the extent of in-class heterogeneity. If Hypothesis 5 is correct, the relationship between conditional dropout probability and in-class rank will be U-shaped if there is sufficient heterogeneity in prior performance.

3.   The data and summary statistics

The data-set is based on anonymised individual Universities Student Records (USR) for the full populations of undergraduate students starting a three or four-year degree course in a UK university between 1984/5 and 1992/3. The data contain information on approximately 714,000 students: about 79,000 per cohort. We address the issue of what

---

[6] Results are available from the authors on request.

determines whether a student will drop out of their university course during their first year. We focus on first-year dropouts for two reasons. First, previous analysis of student drop-out behaviour both for the US (Tinto (1987, 1988)) and for the UK (Smith and Naylor, 2001a) shows that, typically, half of all student dropouts are first-year dropouts and also that the determinants of first-year dropout differ from those on later dropout. Hence, it follows that it is appropriate to analyse first-year dropout behaviour separately, as in the current paper. Second, the individual student-level USR data are not available beyond 1993. Smith and Naylor (2001a) restrict attention to the 1989/90 cohort, which can be followed until 1993 and hence their progression – completion, transfer, withdrawal, deferral *et cetera* – can be observed over an appropriate time frame for modelling behaviour in each year of the degree programme. In the current paper, because we choose to analyse the determinants of dropout probabilities during the first year only, we are able to exploit information for all those cohorts on whom we have first-year information prior to 1993. This includes all cohorts starting between 1984/85 and 1992/93. The last of these cohorts would have been expected to complete their degree programme in 1995 (or 1996 in the case of students taking a four-year programme) and hence the analysis is for more recent cohorts than would otherwise be possible.

Summary statistics

Table 1 reports the first-year dropout probabilities and student populations by gender for each of the nine cohorts. For the last cohort, 1992-93 entrants, the fist year non-completion rate is 5.2% for male and 3.8% for female students. The total student population for this entry cohort is about 100,000 of which 46% is female. The table

shows that the size of the student population grew by 48% from 1984-85, with the bulk of the growth occurring after 1988-89. The 65% rise in the female population is particularly striking. Despite the rapid growth in student numbers, the dropout rate for males appears to have increased only slightly, while that for females is largely unchanged, remaining consistently below the male rate. Table 2 reports summary statistics for various sets of explanatory variables. Social Class I and II refer to students from professional and managerial backgrounds. The pre-school school types show that most students went to state-sector (local education authority (LEA), grammar and further education) schools, but that almost one-quarter attended private fee-paying (Independent) schools.

Our main focus concerns the effects of in-class rank on the student's dropout probability. Students with at least three A-levels are ranked - on the basis of the score in their best three A-levels[7]- relative to other students at their university on their degree course. On the basis of this in-class ranking, students are then allocated to one of five rank categories according to their personal A-level score relative to the mean in-class score within their university degree course. The categories are defined according to whether the individual's A-level score is more or less than half a standard deviation and more or less than 1.3 standard deviations (σ) away from the mean score (μ). This choice of categories generates bands of approximately equal proportions. We also experimented with alternative classifications and found this choice to dominate in maximising the likelihood values. The reason for the use of the five rank categories defined against the criteria of the mean and the standard, rather than using a simple quantile method, arises from the clustering associated with the discrete nature of the A-level scores. Such

---

[7] We consider the best 3 A-levels as this is the basis on which places are awarded by universities.

clustering on particular points scores means that it is not always possible to rank students in such a way as to define distinct quantiles by university course.

Specifically, A-Group 1 (A-Group 5) consists of individuals whose A-level scores are 1.3 or more standard deviations above (below) the mean score and A-Group 2 (A-Group 4) comprises individuals between 0.5 and 1.3 standard deviations above (below) the mean. A-Group 3 consists of individuals within half a standard deviation either side of the mean. This method of measuring relative performance is not sensitive to the possible phenomenon of grade inflation. This would not be the case, for example, if our measure related to the absolute difference between the individual's A-level score and the university course mean.

For the construction of in-class rank, a student's 'class' is based on the university they attend and the subject they study. There are 56 universities and 19 broad subject areas, generating a total of about 1000 'classes', with an average of about 100 students per class for the 1992-93 cohort. In Table 3 we report our estimates for σ and μ for 1984-85 and 1992-93 averaged across universities for eight aggregated subject areas. The table shows substantial variation in both the mean and in the standard deviation both over time and across subject areas. Table 4 shows the proportions of students in each of the categories (A-Groups 1 through to 5), separately for males and females: these proportions are approximately constant over time. The table also shows the average coefficients of variation. The coefficient of variation is the measure we use to capture the effects of in-class heterogeneity on the dropout probability.[8]

---

[8] Use of the coefficient of variation also ensures that our measurement of heterogeneity is independent of (uniform) A-level grade inflation.

4. Results for the 1992-93 entry cohort

We conduct a binomial logit regression analysis of the probability that an individual withdraws from their university degree course during their first year of study in 1992-93. We conduct separate analyses for male and female students as, from the summary statistics discussed above, it appears that male and female drop-out behaviour is rather different. Indeed, a likelihood ratio test on the equality of the estimated coefficients from our separate models for males and females is rejected (p-value = 0.00).

In the logit regression analysis, dummy variables are included for A-Groups 1, 2, 4 and 5 so that the default case is that of an individual with three A-levels and in the middle of the distribution of A-level scores within their university degree course.[9] On the basis of the discussion in section 2 of the paper, Hypothesis 1 predicts that the dropout probability is likely to be negatively related to the student's in-class rank group, with the weakest students (that is, those in A-Group 5) most likely to dropout. The regression equation also includes interactions between the A-Group dummy variables and the coefficient of variation for the in-class distribution of A-level scores.[10] The interaction terms are included in order to test Hypotheses 2 through 5, that the individual's dropout probability varies with the extent of in-class heterogeneity in a way which itself depends on the student's in-class rank, as measured by the A-Group categories.

For students whose prior qualifications are not A-levels, dummy variables are included for particular types of alternative qualifications. For students who took Scottish Higher qualifications prior to university, dummy variables analogous to the A-Group

---

[9] Dummy variables are included in the analysis to pick up the effects of having fewer than 3 A-levels.
[10] In the construction of the interactive dummy variables, A-Groups 1 and 2 are aggregated, as are A-Groups 4 and 5.

variables are included for students with at least 5 Highers and studying at a Scottish university. The in-class coefficient of variation for these students is also included, but not the interaction between this and any measure of the student's rank position. Because of small cell sizes, we have not computed the in-class rank positions of Scottish students studying outside Scotland, but rather include a simple dummy variable for these students. Finally, in the light of evidence of a significant effect of prior Mathematics on student performance (see Smith and Naylor 2001a, 2001b) a dummy variable is included to indicate whether the student had obtained an A-level (or Higher) in Mathematics.

Logit estimates of the probability of dropping out are presented in Table 5, separately for male and female students.[11] The equation includes controls for educational background, personal characteristics, degree subject and related attributes, and university attended. For male students, the probability of dropping out of university tends to be increasing in age whereas for women the dropout probability is lowest for students in the highest age category. The effect of fees status also varies by sex, with non-UK fee paying males less likely to drop out than other male students but with no significant effects of fees status for women. The effects of accommodation type are similar for men and women. Relative to a student living on campus, the dropout probability is higher for students living either at the parental home or otherwise off-campus. This is consistent with Tinto's emphasis on the importance of social integration.

Students with part-time status do not differ from full-time students in their *ceteris paribus* dropout probability. However, the student's social class background has a significant effect - for both male and female students - with a significantly lower

---

[11] Note that the definition of the student's in-class rank is not gender-specific.

16

probability of dropping out for students from Social Class I and II parental occupation (that is, professional and managerial) backgrounds. School background has significant effects only for male students, with a higher dropout probability for students who had previously attended a private Independent school. In general, these results are in line with those of Smith and Naylor (2001a, 2001b),

Table 5 also reports results for the effects on the dropout probability associated with the individual's performance at A-level as measured by (i) the dummy variables indicating the student's in-class rank, (ii) the dummy variables interacting the rank categories with the in-class coefficient of variation, and (iii) the dummy variable for Mathematics A-level.[12]

With respect to the effects of the dummy variables indicating the student's in-class rank, we see that the majority of the estimated coefficients are significant. The same is true for the estimated coefficients on the interactive dummy variables indicating the effects of heterogeneity and how these vary across the different segments of the prior performance distribution. Because of the presence of the interaction terms in the regression equation, we cannot interpret the estimated coefficients on the dummy variables indicating in-class rank in isolation. In order to assess the importance of the estimated prior qualification effects, we have calculated the predicted dropout probabilities for each of the A-Groups for different values of the coefficient of variation.

Figures 2 and 3 present the plots of the predicted probabilities against the coefficient of variation for the rank groups A-Group 1 through 5, for men and women respectively. The vertical line at 0.2 indicates the mean value for the coefficient of

---

[12] Results for students who had taken alternative qualifications - not reported for reasons of space - are available from the authors on request.

variation (see Table 4). At this mean value, the predicted probabilities of dropping out for the five A-Groups are broadly consistent with Hypothesis 1, which predicts that the dropout probability will vary systematically according to the individual's in-class rank, as discussed above.

Specifically, we observe from Figure 2 that, for male students, the predicted dropout probabilities are monotonically related to the in-class rank groups: with the highest-ranked students (A-Group 1) the least likely to dropout, *ceteris paribus*. The predicted dropout probability for A-Group 1 is 2% and for A-Groups 2-5 is 2.5%, 3.5%, 4.2%, 5.7%, respectively. The results, then, are consistent with Hypothesis 1 and with the conventional wisdom in the literature that relative academic preparedness - as indicated by prior performance - is an important determinant of the likelihood of dropping out of college.

For female students, the picture in Figure 3 is only slightly different. At the mean value of the coefficient of variation, the lowest-ranked students (A-Group 5) have the highest predicted dropout probability of 3.5%. The next lowest ranked students (A-Group 4) have the next highest predicted probability of 2.8% and the middle group (A-Group 3) have a still lower predicted probability of 2%, as predicted under Hypothesis 1. However, the predicted probabilities for the higher-ranked students in A-Groups 1 and 2 do not conform to a monotonic pattern, as was the case for males. Instead, for female students, the predicted dropout probability for A-Group 2 is 2% - the same as that for the middle group - and the predicted probability for the highest-ranked students (A-Group 1) is a little higher at 2.4%, giving a U-shaped relationship between the dropout probability and in-class rank. A possible explanation for the behaviour of the higher-ranked female

students is that these students will have better outside options and may have incentives to leave in order to make a better university course match. We also note from Figures 2 and 3 that there is much less dispersion in the predicted probabilities across the rank groups in the case of women.

Hypotheses 2 through 5 predict that the individual's dropout probability varies with the extent of in-class heterogeneity. This is supported by the fact that interactions between the A-Group dummies and the in-class coefficient of variation are typically significant. More specifically, these hypotheses predict that the effects of heterogeneity will vary with the student's in-class rank, as measured by the A-Group categories. Hypothesis 2 suggests that a greater coefficient of variation will raise the probability that the weaker students will fail to complete their studies successfully. From Figure 2, this is consistent with the estimated effects for males, for whom the predicted probabilities increase with the coefficient of variation in the case of the relatively weak students in A-Groups 4 and 5. Similarly, the results for males ranked close to the center of the distribution (in A-Group 3) are consistent with the prediction that these students will be less likely to drop out the greater is the degree of in-class heterogeneity. This is consistent with Hypothesis 3. However, contrary to the prediction of Hypothesis 4 - that relatively well qualified students will also be less likely to drop out the greater is the coefficient of variation - the results suggest the opposite. That is, the predicted probability rises with the degree of heterogeneity for these students. Our interpretation of this result is that the risk of failure for these students is anyway so low that such considerations – which form the basis for the analysis described in section 2 – are dominated by other factors. The result is consistent with Hypothesis 5, that strong students have an incentive to switch university

degree course if they find themselves on a course for which they are relatively over-qualified.

In the case of female students, Figure 3 shows that, as for males and consistent with Hypothesis 3, the predicted probability of dropping out falls for students in A-Group 3 – those ranked close to the centre of the in-class distribution by prior performance at A-level. Contrary to Hypothesis 2 and Hypothesis 4, however, the predicted probabilities for the weaker students (in A-Groups 4 and 5) *fall* as the degree of heterogeneity increases while those for the stronger students (in A-Groups 1 and 2) rise with the extent of heterogeneity, as was the case for males. The behaviour of the stronger students may be explicable in terms of the incentive to switch to university degree courses with higher average A-level scores, as predicted by Hypothesis 5. The reasons behind the observed behaviour of the weaker female students are less obvious.

Finally, we also note from Table 5 that having Mathematics A-level significantly reduces the dropout probability for both male and female students. In preliminary analysis, we have also investigated possible effects associated with the extent of the individuals' integration into the socio-academic life of the university department but found no significant effects.

5.   Earlier cohorts and time-series evidence

The analysis reported in the previous section for 1992-93 was replicated on data for the preceding eight cohorts. Results for 1984-85 are presented in Table 5 and can be compared directly with the results for 1993-93. In terms of the key variables relating to the effects of prior qualifications, Figures 4 and 5 present the predicted dropout

probabilities against entry year for particular values of the coefficient of variation, in the range (0.1, 0.4) for a median male and a median female student, respectively. The results are very robust, with the dropout probability falling as the coefficient of variation increases.

The estimated effects associated with the other sets of explanatory variables are also reasonably robust over time. For males, the effects of attendance at an Independent school, off-campus accommodation and of part-time status are typically positive and significant across the cohorts. In addition, there is generally a higher dropout probability for male students aged 21 or above. Conversely, having a Mathematics A-level has a consistently significant and negative effect, as does non-UK fees status and having parents from a Social Class I or II background. For females, the same results obtain with respect to the social class, part-time status, Mathematics and accommodation variables.

On the basis of the results for the both the 1984-85 and the 1992-93 cohorts, we have decomposed the differences – both by gender and over time – in the predicted dropout probabilities into that part explained by differences in characteristics and that part attributable to differences in estimated coefficients. Table 6 presents the results of this Oaxaca decomposition. Consider first the decomposition of the gender differences for 1992-93 cohort. The table shows that the predicted probability of dropping out was 3.80% for females and 5.22% for males. If females are attributed male characteristics, the predicted probability is a little higher at 3.98% and if males are attributed female characteristics the dropout probability falls slightly - to 4.89%. Hence, the gender difference in the dropout rate in 1992-93 is not explained by differences in observed characteristics by gender: the difference is attributable to differences in estimated

coefficients. The same picture emerges from a gender composition based on the 1984-85 cohort.

Consider now the decomposition over time. For males, the results presented in Table 6 suggest that the reason for the rise in the predicted dropout probability from 1984-85 to 1992-93 was attributable to a deterioration in characteristics. For example, if 1984-85 males are assigned 1992-93 male characteristics, the predicted probability of dropping out increases from 4.93% to 5.52%, compared to 5.22% for 1992-93 males with their actual characteristics and estimated coefficients. Thus, changing coefficients acted to reduce the predicted male dropout probability over the period, but *not* sufficiently to fully offset the deterioration in characterisitcs. The reverse is true for females: if 1984-85 females are assigned 1992-93 female characteristics, the predicted probability of dropping out increases from 4.09% to 4.66%, compared to 3.80% for 1992-93 females with their actual characteristics and estimated coefficients – hence the effect of a deterioration in characteristics *is* more than offset by changed coefficients.

6.   Conclusions

We have examined the first-year undergraduate university dropout behaviour of UK university students from administrative data for full entry cohorts between 1984-85 and 1992-93. We have focused on the impact of prior qualifications and on differences by gender and over time. With respect to prior qualifications, our theoretical framework generated a number of testable hypotheses.

First, we tested the hypothesis that the probability of dropping out is greatest for students with relatively poor levels of prior attainment. Our method for doing this is to

rank each individual in each university and in each degree course on the basis of their performance at A-level relative to their fellow students. Each student is then assigned to one of five groups according to this rank. Our results show that, for male students, the predicted probability of dropping out is related monotonically to the in-class rank group to which the individual belongs, with the weaker students more likely to drop out, as predicted. For females, the results are the similar, with the exception that the very strongest students are *more* likely to drop out than are those with close to the mean. One possible explanation for this is that the strongest students are likely to have better outside options. In particular, they are more able to switch to a university degree course with higher average levels of prior attainment, which may enhance their labour market prospects.

Second, we have found that the dropout probability is significantly affected by the degree of in-class heterogeneity of students with respect to levels of prior attainment. Third, and in particular, we have found that – for both male and female students – the probability of dropping out of a degree course is decreasing in the extent of heterogeneity for students close to the center of the in-class distribution by prior attainment. This is consistent with the predictions arising from the theoretical framework set out in the paper. Intuitively, the greater is the extent of in-class variation by prior performance, the less at risk of failure are the students close to the mean, for any given failure rate. The same intuition lies behind the converse prediction that the dropout probability is increasing in the extent of in-class heterogeneity for relatively weak students. This prediction is consistent with the data on male students, but not supported by the evidence for female students. Finally, the evidence suggests that for the stronger students the

probability of dropping out is rising in the extent of in-class heterogeneity. This is consistent with the view that the over-riding effect of increasing heterogeneity for these students is to increase their incentives to switch to a university degree course with a higher prior-qualification average score.

We have also examined the extent to which differences in the predicted dropout probabilities by gender and over time are attributable to differences in characteristics rather than to differences in estimated coefficients. Our results imply that differences by gender are explained largely by differences in coefficients. We also show that the changes in predicted probabilities over time are attributable to deteriorating characterisitcs being only partially offset by changed coefficients in the case of males but being more than offset in the case of females.

In terms of government policy to widen participation in higher education in the UK, our results indicate that admissions policies which increase heterogeneity within university degree courses run the risk of raising the dropout rate. For male students, for example, a higher coefficient of variation raises the dropout rate of all students but those close to the mean. For female students, the effect is less clear: a rise in the coefficient of variation raises the dropout rate of the higher ranked students in the university course-specific distribution, but lowers the dropout rates of the rest. We conclude that policies aimed at widening participation not through specialisation but through encouraging increased heterogeneity *within* university courses should be complemented with appropriate strategies – educative, social, financial and pastoral – to minimise the risk that the dropout will rise as a result.

References

Arulampalam, W., Naylor, R. and Smith, J. (2001), "A hazard model of the probability of medical school dropout in the UK," Warwick Economic Research Papers no. 597.

Dearing, R., (1997). *Higher Education in the Learning Society*. National Committee of Enquiry into Higher Education, HMSO, London.

Fielding, A., Belfield, C. and Thomas, H. (1998), "The consequences of drop-outs on the cost-effectiveness of 16-19 colleges," *Oxford Review of Education*, vol. 24(4), pp. 487-511.

Gomulka, J. and Stern, N. (1990), "The employment of married women in the United Kingdom 1970-1983," *Economica*, vol. 57 (226), pp. 171-200.

HEFCE, (1999), "Performance indicators in higher education in the UK," Report 99/66, HEFCE, Bristol.

Johnes, G. and McNabb, R. (2001), "Never give up on the good times: student attrition in the UK," mimeo.

Johnes, J. and Taylor, J., (1989), "Undergraduate non-completion rates: differences between UK universities," *Higher Education*, 18, pp. 209-225.

Johnes, J. and Taylor, J., (1990), *Performance indicators in higher education*, SRHE/OUP, Oxford.

Light, A. and Strayer, W., (2000), "Determinants of college completion: school quality or student ability?" *Journal of Human Resources*, vol. 35, pp. 299-332.

Oaxaca, R. (1973), "Male-female wage differentials in urban labour markets", *International Economic Review*, 14, 693-709.

Smith, J., McKnight, A. and Naylor, R., (2000), "Graduate employability: policy and performance in higher education in the UK," *Economic Journal*, 110, pp. F382-411.

Smith, J. and Naylor, R., (2001a), "Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students," *Journal of the Royal Statistical Society Series A*, 164, 389-405.

Smith, J. and Naylor, R., (2001b), "Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort*," Oxford Bulletin of Economics and Statistics*, 63, 29-60.

Tinto, V., (1975), "Dropout from higher education: a theoretical synthesis of recent research," *Review of Educational Research*, vol. 45, pp. 89-125.

Tinto, V., (1987), *Leaving college: rethinking the causes and cures of student attrition*, Chicago, University of Chicago Press.

Tinto, V., (1988), "Stages of student departure: reflections on the longitudinal character of student leaving," *Journal of Higher Education*, vol. 59, pp. 438-455.

Tinto, V., (1997), "Classrooms as communities: exploring the educational character of student persistence," *Journal of Higher Education*, vol. 68, pp. 599-623.

Table 1 – First-Year Dropout Probabilities (%) over time by Gender

	1984/85	1985/86	1986/87	1987/88	1988/89	1989/90	1990/91	1991/92	1992/93
Dropout probability - Males	4.93	4.77	5.04	5.02	5.13	5.32	4.83	5.16	5.22
Sample Size - Males	40257	40346	40483	40986	42298	45382	47286	50641	54725
Dropout probability - Females	4.09	3.66	3.87	3.58	4.10	3.94	3.71	3.94	3.80
Sample Size - Females	28529	28796	29049	31232	31744	35634	37771	41950	47020
Proportion of Females	41.48	41.65	41.78	43.25	42.87	43.98	44.41	45.31	46.21

Table 2 - Descriptive Statistics for variables used in the analyses

Variable	1984/5		1992/3	
	Males	Females	Males	Females
Age groups				
Aged <20	0.831	0.846	0.767	0.778
Aged 20	0.064	0.050	0.067	0.057
Aged 21-28	0.079	0.059	0.119	0.092
Aged more than 28	0.026	0.045	0.047	0.073
Student fees status				
Home student	0.947	0.966	0.948	0.957
Non-UK fee student	0.053	0.034	0.052	0.043
Accommodation at university				
Living on campus	0.788	0.773	0.765	0.742
Living at home	0.108	0.128	0.133	0.151
Living in digs	0.083	0.074	0.086	0.088
Other accommodation	0.021	0.025	0.016	0.019
Student enrolment status				
Full-time	0.991	0.980	0.988	0.982
Part-time	0.009	0.020	0.012	0.018
Degree programme length				
On a 3 year programme	0.810	0.772	0.725	0.689
On a 4 year programme	0.190	0.228	0.275	0.311
Social class				
Social Class I and II	0.586	0.609	0.549	0.560
Other	0.414	0.391	0.451	0.440
Schooling				
Local Education Authority	0.478	0.497	0.441	0.445
Grammar	0.095	0.108	0.094	0.103
Independent	0.237	0.200	0.227	0.202
Further Education college	0.084	0.094	0.091	0.101
Other school	0.106	0.102	0.147	0.149
'A' Level Mathematics				
Has 'A' Level mathematics	0.502	0.272	0.401	0.217
No 'A' Level mathematics	0.498	0.728	0.599	0.783

Table 3 – Distribution of average 'A' Level Scores (std.dev) by Faculty

	1984/85	**1992/93**
Medical Related	7.01 (1.58)	6.99 (1.71)
Physical Sciences	7.39 (1.83)	7.15 (1.98)
Mathematical Sciences	7.99 (1.68)	7.49 (1.95)
Engineering Sciences	7.50 (1.79)	7.02 (2.06)
Social Studies	7.46 (1.48)	7.75 (1.49)
Language Related	7.73 (1.58)	7.93 (1.54)
Arts & Humanities	7.29 (1.65)	7.63 (1.57)
Other	7.28 (1.98)	7.58 (1.87)
Total	7.47 (1.66)	7.49 (1.71)

Notes: (i) The above averages are calculated over the best three 'A' Level scores for those students who had at least 3 'A' Levels. (ii) Both the mean and standard deviation are the within-class mean and standard deviation averaged over the broad degree areas categories above.

Table 4 – Proportion of 'A' level students in the quantiles by gender

	1984/5		**1992/3**	
	Males	Females	Males	Females
Average Score < -1.3 σ	11.7	9.0	11.7	10.5
-1.3 σ <Av. Score < -0.5 σ	24.6	22.5	24.7	22.2
-0.5 σ <Av. Score < 0.5 σ	32.6	32.6	32.8	35.0
0.5 σ < Av. Score < 1.3 σ	21.2	23.9	20.7	21.7
1.3 σ < Average Score	9.9	12.1	10.1	10.7
Average Score	7.5 (1.7)	7.5 (1.6)	7.4 (1.8)	7.6 (1.7)
Coefficient of Variation	0.2	0.2	0.2	0.2

Notes: - (i) Average Score is calculated over the best three 'A'Level scores if the number of 'A' Level subjects taken was at least 3. The points allocated were - 'A' 10 points, 'B' 8 points, 'C' 6 points, 'D' 4 points, and 'E' 2 points. The allocation was based on within, course and university. (ii) Coefficient of Variation is also based on within course and university.

Table 5 - Logit Model Coefficient Estimates (Absolute T-ratio)

Variable	1984/5		1992/3	
	Males	Females	Males	Females
Age Dummies – base <20 years				
Aged 20	-0.114 (1.12)	0.571 (4.76)	0.127 (1.66)	0.267 (2.79)
Aged 21-28	0.263 (3.04)	0.499 (4.20)	0.081 (1.24)	0.124 (1.41)
Aged more than 28	0.255 (1.95)	0.152 (1.00)	0.236 (2.64)	-0.226 (2.03)
Non-UK fee student	-0.293 (2.33)	0.062 (0.37)	-0.349 (3.38)	0.156 (1.27)
Accommodation – base University				
Living at home	0.276 (3.10)	0.281 (2.47)	0.304 (4.65)	0.388 (4.69)
Living in digs	0.580 (7.20)	0.388 (3.41)	0.524 (8.08)	0.392 (4.65)
Other accommodation	0.620 (3.92)	0.968 (4.88)	0.560 (4.64)	0.845 (6.44)
Part-time student	1.368 (6.50)	1.840 (7.65)	0.181 (0.72)	0.170 (0.71)
On a 4 year programme	0.002 (0.03)	0.139 (1.53)	0.139 (2.48)	0.118 (1.80)
Social Class – base Skilled, Semi-skilled, Unskilled (non-manual/manual)				
Social Class I and II	-0.165 (3.16)	-0.195 (2.73)	-0.075 (1.77)	-0.089 (1.68)
Type of School – base Local Education Authority (LEA)				
Grammar	-0.052 (0.53)	-0.197 (1.47)	-0.092 (1.06)	-0.093 (0.92)
Independent	0.169 (2.42)	0.091 (0.94)	0.122 (2.10)	-0.003 (0.04)
Further Education college	0.119 (1.26)	0.077 (0.62)	0.047 (0.60)	0.090 (0.99)
Other school	0.116 (1.08)	0.298 (2.15)	0.026 (0.35)	-0.057 (0.61)
'A' Level Qualifications – base Average score is within $\pm 0.5\,\sigma$				
Average Score $< -1.3\,\sigma$	-0.997 (3.29)	-0.684 (1.68)	-1.395 (5.71)	-0.962 (3.46)
$-1.3\,\sigma <$ Av. Score $< -0.5\,\sigma$	-1.030 (3.80)	-0.878 (2.30)	-1.148 (5.42)	-1.107 (4.34)
$0.5\,\sigma <$ Av. Score $< 1.3\,\sigma$	-0.530 (2.29)	-0.485 (1.46)	-0.332 (1.94)	0.021 (0.09)
$1.3\,\sigma <$ Average Score	-0.175 (0.75)	-0.404 (1.20)	0.002 (0.01)	0.312 (1.34)
Coefficient of Variation (CV) of 'A' Level scores – interactions				
CV for Average Score $< -0.5\sigma$	1.681 (1.31)	1.557 (0.86)	2.064 (2.24)	1.840 (1.55)
CV for $-0.5\sigma <$ Av. Score$<0.5\sigma$	-1.569 (2.27)	-1.998 (1.90)	-1.974 (4.43)	-3.693 (5.91)
CV for $0.5\sigma <$ Average Score	1.889 (1.85)	1.551 (1.03)	0.614 (0.88)	-2.528 (2.34)
Has an 'A' Level Maths	-0.129 (1.74)	-0.134 (1.29)	-0.419 (6.99)	-0.326 (3.88)
Maximised Log Likelihood	-7156.95	-4306.10	-10308.59	-7117.29
LR (χ^2 (104) for covariates [P-value]	1491.58 [0.00]	1133.23 [0.00]	1804.06 [0.00]	963.96 [0.00]
Number of Observations	40242	28520	54723	47017

Notes: - (i) The dependent variable take the value of unity if the student drops out of university. (ii) Average Score is calculated over the best three 'A'Level scores if the number of 'A' Level subjects taken was at least 3. The points allocated were - 'A' 10 points, 'B' 8 points, 'C' 6 points, 'D' 4 points, and 'E' 2 points. The allocation was based on within, course and university. Dummies for those taking only one or two 'A' Levels were also included. (iii) Variables similar to 'A' Levels were included for Scottish students' 'Higher' qualifications. (iv) Coefficient of Variation also based on within course and university. (v) In addition to the above set of variables, the models also included controls for courses, universities.

Table 6 – Oaxaca decomposition of predicted probabilities by gender and by entry cohort

			Characteristics			
			1985 Male	1985 Female	1993 Male	1993 Female
Coefficients	1985	Male	4.93 (0.70)	5.12 (0.73)	5.52 (0.78)	5.68 (0.80)
	1985	Female	4.25 (0.83)	4.09 (0.77)	4.86 (0.93)	4.66 (0.87)
	1993	Male	4.50 (0.47)	4.29 (0.46)	5.22 (0.53)	4.89 (0.51)
	1993	Female	3.50 (0.51)	3.39 (0.49)	3.98 (0.56)	3.80 (0.53)

Notes:
(i)   The predicted probabilities are calculated as the average of predicted probabilities over various entry cohorts (and by gender) using estimated models.

(ii)   The standard errors are calculated as the square root of $\frac{\partial \hat{P}}{\partial \hat{\beta}'}[c\hat{o}v(\hat{\beta})]\frac{\partial \hat{P}}{\partial \hat{\beta}} + \frac{1}{n^2} \quad \left(\hat{p}_i - \hat{P}\right)^2$ where

$\hat{\beta}$ are the logit model coefficient estimates, $c\hat{o}v(\hat{\beta})$ is the estimated variance covariance matrix, and $\hat{P} = (1/n) \quad \hat{p}_i$. See Gomulka and Stern (1990) for further details about the derivation of this formula.

Figure 1   The effect of halving the range under a linear function for the conditional
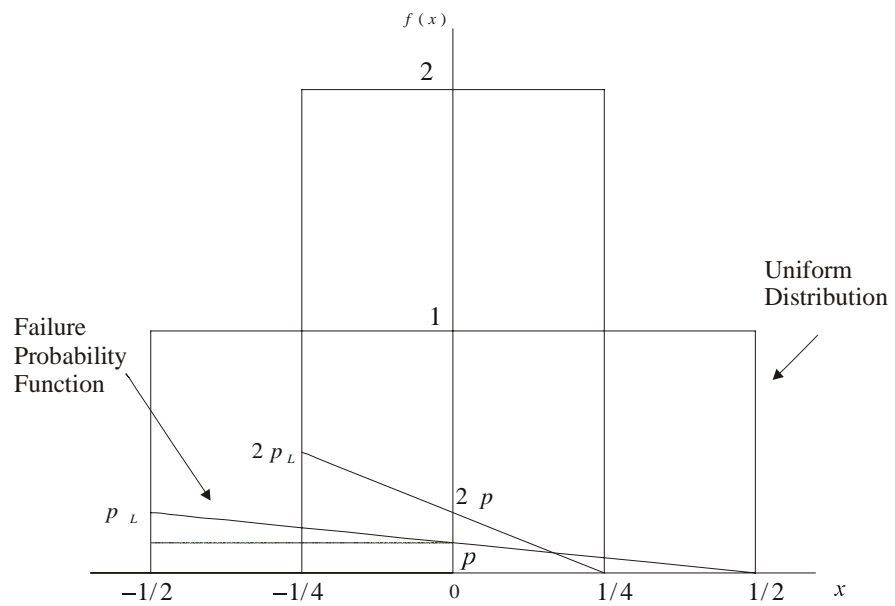
probability of non-completion.

Figure 2: Predicted dropout probabilities against the coefficient of variation for males- 1993
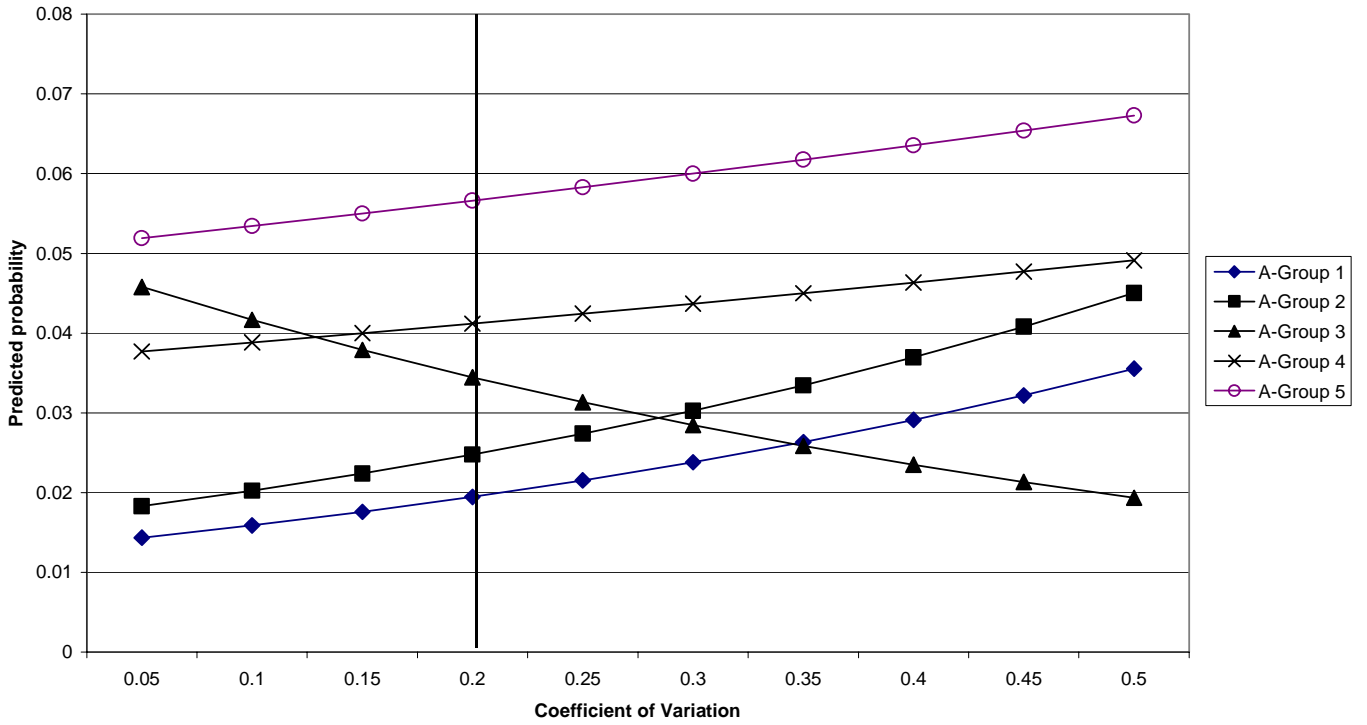

Figure 3: Predicted dropout probabilities against the coefficient of variation for females- 1993
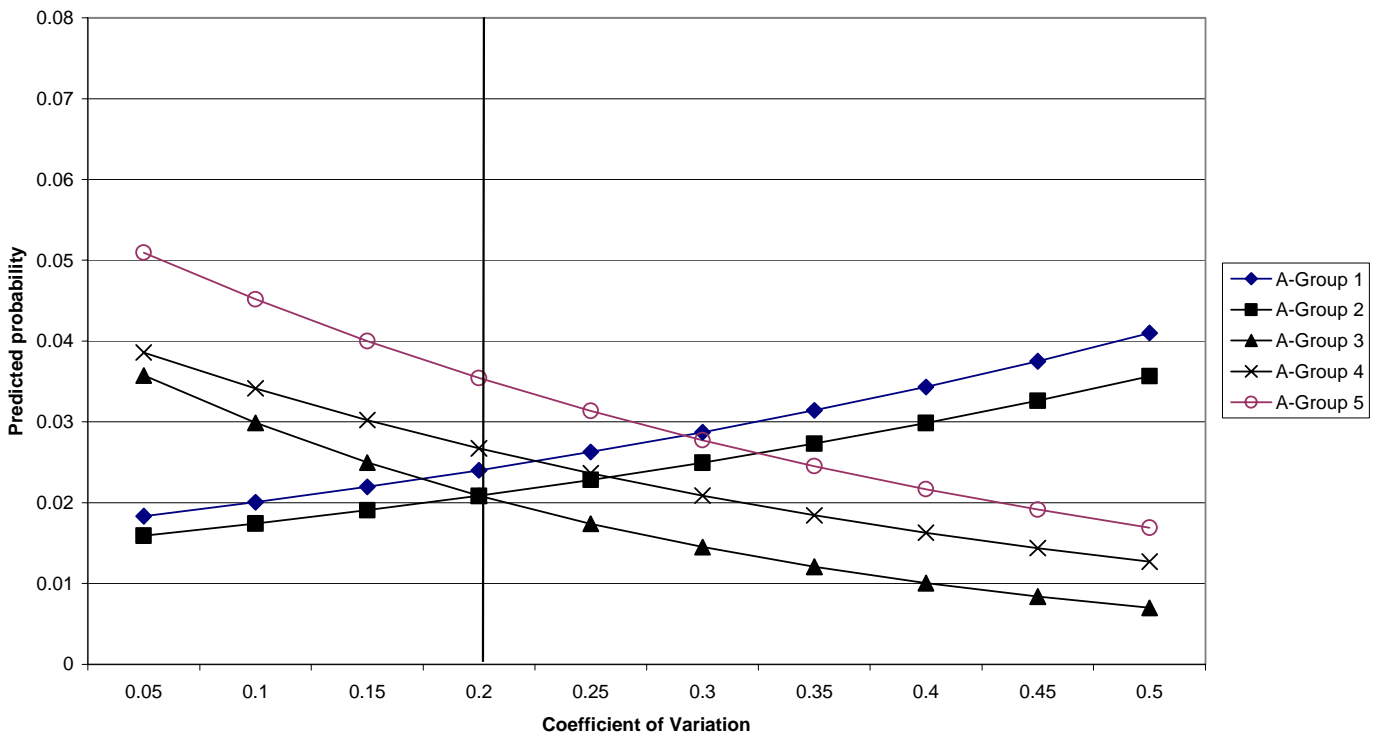
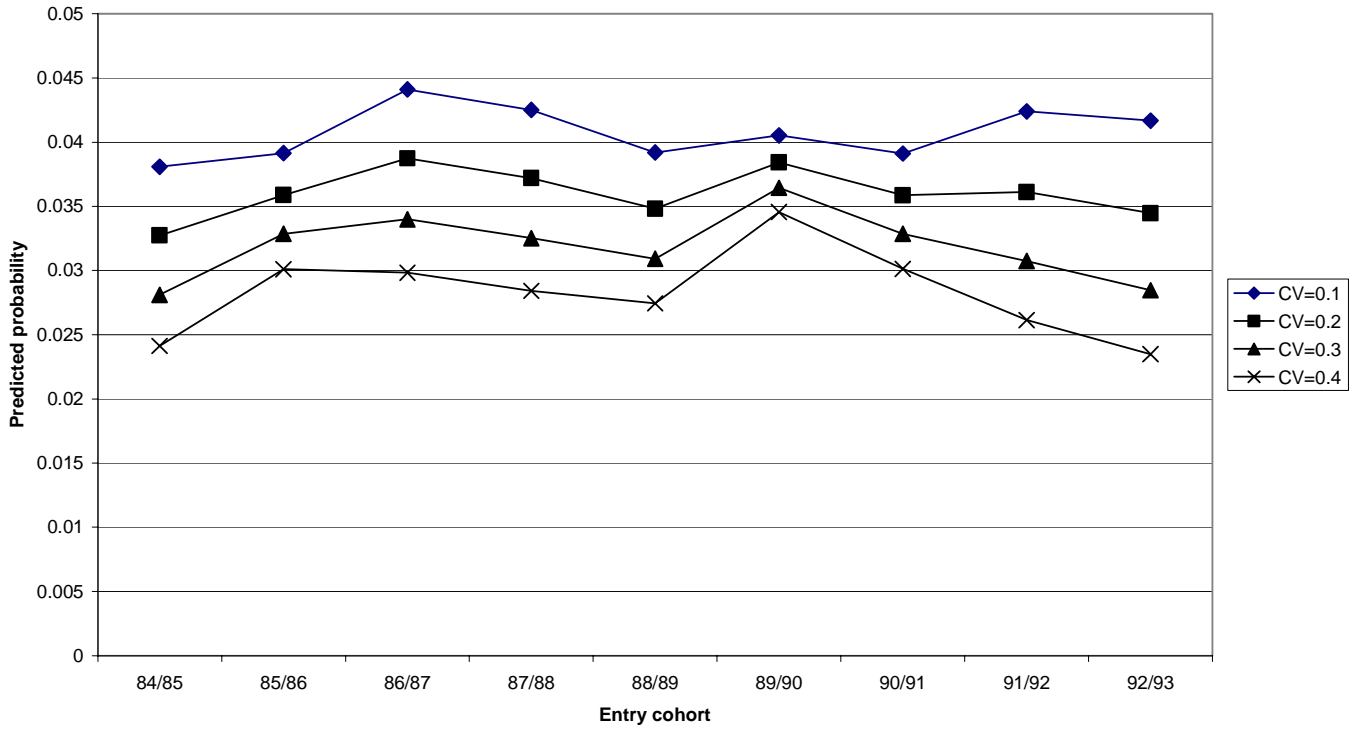**Figure 4: Predicted dropout probabilities for a median group (A-Group 3) male**
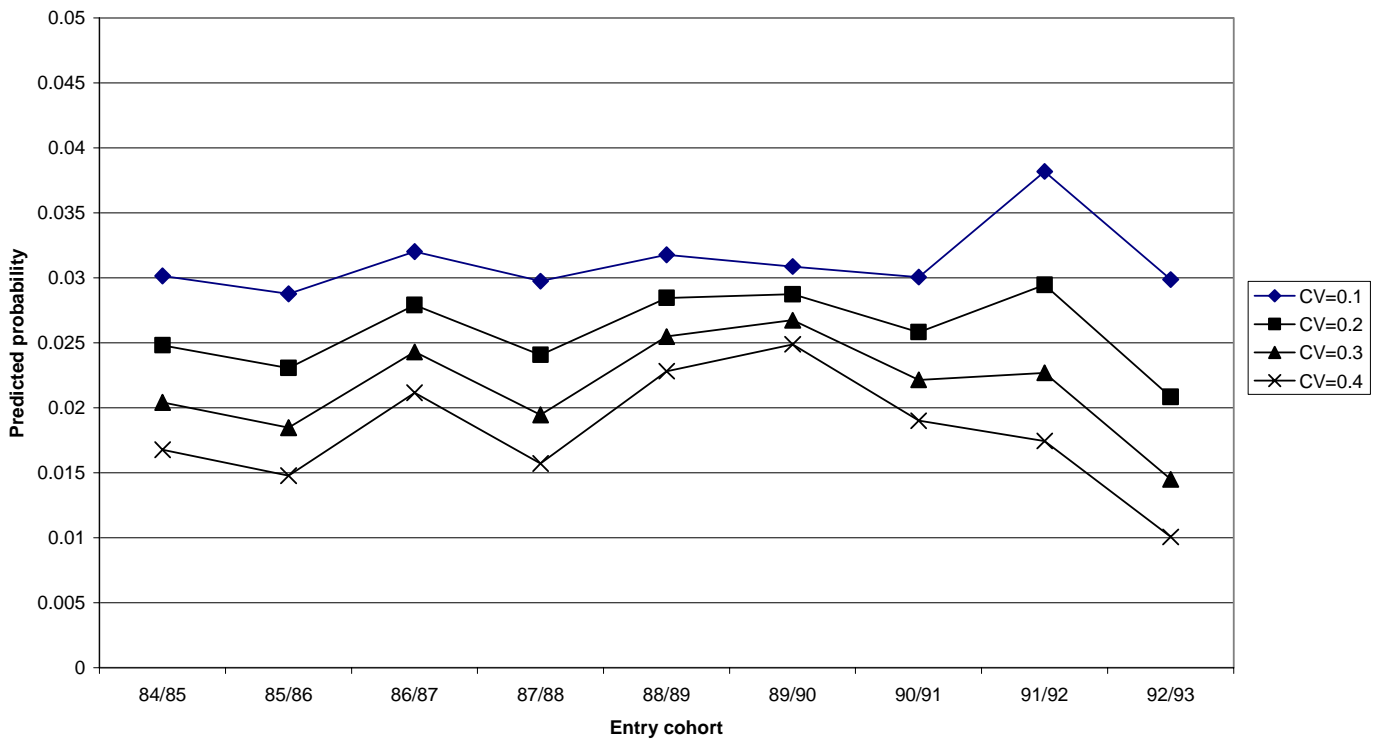

Figure 5: Predicted dropout probabilities for a median group (A-Group 3) female

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
639	G. Saint-Paul	Are Intellectual Property Rights Unfair?	3	11/02
640	J. Hartog N. Jonker H. van Ophem	Dual Track or Academic Route for Auditors: Does It Matter?	6	11/02
641	J. Hartog L. Diaz Serrano	Earning Risk and Demand for Higher Education: A Cross-Section Test for Spain	2	11/02
642	J. Hartog A. Zorlu	The Effect of Immigration on Wages in Three European Countries	1	11/02
643	A. Björklund M. Sundström	Parental Separation and Children's Educational Attainment: A Siblings Approach	6	11/02
644	J. D. Brown J. S. Earle	Job Reallocation and Productivity Growth Under Alternative Economic Systems and Policies: Evidence from the Soviet Transition	4	11/02
645	M. Karanassou H. Sala D. J. Snower	Long-Run Inflation-Unemployment Dynamics: The Spanish Phillips Curve and Economic Policy	3	11/02
646	L. Graham D. Snower	The Return of the Long-Run Phillips Curve	3	11/02
647	C. Grund D. Sliwka	Envy and Compassion in Tournaments	1	11/02
648	C. Schnabel T. Schank J. Wagner	Works Councils - Sand or Grease in the Operation of German Firms?	3	11/02
649	J. van Ours J. Veenman	From Parent to Child: Early Labor Market Experiences of Second-Generation Immigrants in the Netherlands	1	11/02
650	H. Battu P. J. Sloane	Overeducation and Ethnic Minorities in Britain	2	11/02
651	E. Schlicht	Social Evolution, Corporate Culture, and Exploitation	3	11/02
652	E. Plug	How Do Parents Raise the Educational Attainment of Future Generations?	5	11/02
653	W. Groot E. Plug H. Maassen van den Brink	Money for Health: The Compensating Variation of Cardiovascular Diseases	6	11/02
654	C. Ruhm U. G. Gerdtham	Deaths Rise in Good Economic Times: Evidence From the OECD	2	11/02
655	W. Arulampalam R. A. Naylor J. P. Smith	Effects of In-Class Variation and Student Rank on the Probability of Withdrawal: Cross-Section and Time-Series Analysis for UK University Students	2	11/02