J. R. Statist. Soc. A (2001)
164, *Part* 2, *pp.* 389–405

Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students

Jeremy P. Smith and Robin A. Naylor

University of Warwick, Coventry, UK

[Received December 1999. Final revision August 2000]

Summary. From individual level data for an entire cohort of undergraduate students in the 'old' universities in the UK, we use a binomial probit model to estimate the probability that an individual will 'drop out' of university before the completion of their degree course. We examine the cohort of students enrolling full time for a 3- or 4-year degree in the academic year 1989–1990. We find evidence to support both the hypothesis that the completion of courses by students is influenced by the extent of prior academic preparedness and the hypothesis that social integration at university is important. We also find an influence of unemployment in the county of prior residence, especially for poorer male students. Finally, we draw conclusions regarding the public policy of constructing university performance indicators in this area.

Keywords: Probit estimation; Student drop-out (non-completion) probabilities; University graduates; University performance

1. Introduction

The question of why student drop-out rates vary across higher educational institutions is both important and topical and has been the focus of considerable research activity, especially in the USA. For the UK, unlike the USA, most of the analysis has been based on university level data (see, for example, Johnes and Taylor (1989, 1990)) or, in the case of student level data, has been conducted on relatively small samples of students (see Johnes (1990) and Johnes and Taylor (1991)). Large scale microdata on UK university students have not been available. This has now changed with research access to the full set of individual student level information stored in the Universities Statistical Records (USR). (The USR preceded the Higher Education Statistics Agency as the depository for all the statistical returns from what are typically referred to as the 'pre-1992' universities, predating the abolition of the binary divide.) These data, together with merged data on schools, provide the information that we exploit in the current paper.

A primary motivation for studying the non-completion rates of UK university students stems from the fact that the Government is currently developing a range of performance indicators for higher education institutions. A first wave of indicators was published in December 1999 (see Higher Education Funding Council for England (1999a, b)) and included a performance measure based on the non-completion rate of students by institution. One of the primary objectives of the published performance indicators is to improve on raw rankings

Address for correspondence: Robin A. Naylor, Department of Economics, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: robin.naylor@warwick.ac.uk

 0964–1998/01/164389

or league tables of universities by comparing universities against a bench-mark that takes account of subject mix and of variations in students' entry qualifications. The published performance indicators are derived from a macrolevel analysis of university level data. Our microlevel analysis enables us to identify the influence on individual student drop-out rates of a large number of potential explanatory variables and to investigate the extent to which differences in these variables across institutions might explain the differences in drop-out rates across universities that are observed in the raw data. The estimated effects on the individual's drop-out probability of control variables such as school background also enable us to comment on possible conflicts of incentives for institutions arising from different performance criteria, such as that which might arise between minimizing the drop-out rate and maximizing 'access' for individuals from relatively disadvantaged educational backgrounds.

A second motivation for our analysis is also related to public policy matters. It has been suggested that recent changes in the nature of funding students through university are likely to have had particularly adverse effects on the probabilities both of entry into university and of withdrawal from university for students from relatively poor backgrounds. In this paper, we analyse the effect of, *inter alia*, the individual's social class background on the drop-out probability. The cohort studied is the class of students admitted to university in the autumn of 1989. (The original data are organized in the form of leaving cohorts. We have transformed the data across several leaving years to generate an entry cohort. This is described in Section 3.) The results enable us to infer the differences in the withdrawal probabilities of students from different social classes, *ceteris paribus*. This is informative in itself but will also provide future work with a bench-mark against which to measure whether or how the sensitivity of the individual withdrawal rate to background characteristics has changed since the introduction of home student tuition fees for cohorts entering university from the autumn of 1998. (It is possible that the introduction of fees may reduce the numbers of drop-outs by inducing a more efficient search by applicants.) It will be a considerable time before data on these students become available. We note that our data set contains no information on individuals who did not attend university. Consequently, our results are to be interpreted as conditional on students attending university.

A third motivation is that the results of this kind of analysis should be of interest both to potential students and to institutions themselves. For example, professional bodies with responsibilities for individual subject areas may be interested to see the extent of *ceteris paribus* differences in drop-out probabilities by discipline.

A fourth reason for studying the drop-out probability is more theoretically based. In the study of rates of return to education, there has been a long-standing debate on whether drop-outs are rewarded for their human capital acquisition in the same way as those graduating successfully from their course or whether, in contrast, there is a premium to completion *per se*, i.e. a 'sheepskin' effect. This is the famous 'human capital theory' *versus* the 'screening hypothesis' debate: see, for example, the seminal papers by Spence (1973), Layard and Psacharopoulos (1974) and Riley (1979). In this literature, it is typically supposed that the drop-out rate is exogenously determined. If, instead, the drop-out rate is influenced by the same characteristics that affect post-university earnings, then there is likely to be a sample selection bias in the estimates of any earnings premium that is associated with completing a degree course. We cannot model this bias, as we do not have reliable post-university first-destination information for those not completing their studies. However, modelling the drop-out rate is an important step in the analysis of these issues.

The rest of this paper is organized as follows. In Section 2, we provide a brief literature survey which motivates our choice of explanatory variables in the context of the major

hypotheses concerning student drop-out behaviour. In Section 3, we describe the important features of the data set and, in Section 4, we present the main results of our analysis. Section 5 focuses on the implications of the analysis for the construction and interpretation of university performance indicators against the criterion of non-completion rates for students. Section 6 closes the paper with a discussion and further remarks.

2. Literature review: models and hypotheses

That the non-completion rate for students has been the focus of much more analysis in the USA than in the UK may in part reflect the fact that the non-completion rate is much higher in the USA (at about 37%) than in the UK (currently about 18% in the expanded higher education sector). This may change if the growing participation rate in higher education in the UK continues to generate a rising drop-out rate. Surveying the US literature on college drop-out behaviour, Kalsner (1992) emphasized that decisions to withdraw are typically based on personal, social and financial considerations, with only a small minority of departures resulting from academic dismissal. 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 the educational institution is the major determinant of completion and identifies some key influences on integration. These include the student's family background, personal characteristics, previous schooling, prior academic performance and interactions between students and the faculty.

Family background is likely to influence not only the financial capacity of students to complete their studies but also their preparedness for and commitment to college and, related to this, their post-college occupational aspirations. Previous schooling can have effects that are similar to those of family background. For the UK, Johnes (1990) found significant effects of both parental social class and of school type on non-completion probabilities for university students. There is strong evidence that pre-college academic preparedness is important (see, for example, Noel and Levitz (1985)). For the UK, Fielding *et al.* (1998) identified the prior educational achievements of students as a major determinant of college non-retention for the 16–19 years age group. In our analysis, we include indicators of the individual student's family background together with explanatory variables both for previous school characteristics — including school type — and for the student's own pre-university qualifications. These include not only the student's level of overall performance but also a measure of the closeness of fit between university subject area and the subjects studied at school. We are also interested in examining whether previous school characteristics influence drop-out behaviour in the context of the research debate on the issue of whether school quality affects pupils' performance and subsequent outcomes: see, for example, Krueger (1999) and FitzGibbon (1996).

Related to Tinto's integration hypothesis is Astin's (1979) theory focusing on the student's 'intensity of involvement' in the social and academic life of the college community. To capture effects of this kind, we include control variables for whether the student lived on campus, off campus or in the parental home. We also include various measures of the social mix of the university department in which the student studied, and we interact these with the individual's own characteristics, e.g. investigating whether a female student's drop-out probability is influenced by the proportion of males in the university department. Tinto (1997) argued that the effects of classroom activity and interactions have been underexplored in the literature on non-completion. We are interested to see whether a university's teaching quality assessment TQA score — which should reflect the effectiveness of the teaching

relationship — is related to non-completion. In our analysis, then, we examine the effects both of university (including university department) characteristics and of previous school characteristics on university student drop-out behaviour.

Finally, we examine whether student drop-out behaviour is influenced by labour market conditions and, in particular, by unemployment in the county of prior residence. Various, potentially off-setting, mechanisms might be operating to generate an effect of local unemployment on the drop-out rate. First, although the graduate labour market is likely to have the characteristics of a national market, the opportunity cost of remaining at university rather than dropping out is likely to be affected by the probability of obtaining employment in the local labour market. This would suggest that there might be a negative relationship between local unemployment and the withdrawal probability. However, students may base their expectations of post-university employment probabilities on the unemployment rate observed in their pre-university local labour market. In this case, given that there are direct as well as opportunity costs of attending university, a high rate of unemployment in the local labour market would lead to a low expected rate of return on investments in human capital. This generates a positive relationship between the local unemployment rate and the withdrawal probability. We hypothesize that this effect is stronger for individuals who expect their post-university employment to be in their previous local rather than in the national labour market. Thus, we examine whether any positive relationship between the unemployment rate and the withdrawal probability is stronger for individuals with poorer post-university prospects and for students with limited access to capital markets for funding their studies.

3. Data and modelling

The data set is based on anonymized individual universities student records for the full populations of undergraduate students leaving the traditional 'pre-1992' universities in one of the years 1990, 1991, 1992 or 1993. The data contain information on approximately 400000 students: about 100000 per cohort. From information on each of these 'leaving cohorts', we have generated a data set comprising all those students who entered university at the start of the academic year 1989–1990 to study for a full-time 3- or 4-year undergraduate degree and who either completed their degree course (successfully or unsuccessfully) at the end of 3 or 4 years or left university before completion. (We note that the total student enrolments in UK universities in 1989 were not unusual: they were in line with the increasing trend in numbers of students. 3- and 4-year degrees account for 93.5% of students leaving university in 1993.) In the event of non-completion, we have an administrative leaving date.

Accordingly, in what follows we report the results based on the analysis of completion *versus* withdrawal, where withdrawal occurs at any point after entry and before an arbitrary cut-off date in the final year. We define this cut-off as occurring at the start of the second term of the three-term (final) year. (There are very few reported cases of non-completing students withdrawing after this date: less than 1% of our sample. Including these individuals as drop-outs did not change the results.) For estimating the individual's drop-out probability, the constructed entry cohort is used, instead of the leaving cohort, to standardize for time-varying influences. We drop those cases where the reason reported for the student's leaving was ill health or death. (There may be cases where ill health or death is generated endogenously and related to the tensions that are associated with a withdrawal decision. However, ill health or death is a little-cited reason for leaving university, with only 125 individuals so reported in our data set.) Our final sample consists of 33851 female and 42407 male students who entered university in the autumn of 1989.

The basic USR data set has been augmented by merging official Government department (the Department for Education and Employment) information on the school that each student last attended before entering university, and also by matching the county level rate of unemployment to the student's county of prior residence.

3.1. Description of the data

For our sample, the non-completion rate was 10.3% for male and 7.1% for female students. In Fig. 1, we show the pattern of how drop-out rates vary across universities, separately for male and female students. The universities are ordered in Fig. 1 by the drop-out rates for males. There is a considerable variation in the drop-out rate across universities and this variation is different for male and female students. The range is from about 1% to 23% for males and from 2% to 19% for females.

It is likely that part of the variation across institutions can be explained by differences across universities in the characteristics of the students and courses. Fig. 2 shows the differences in the drop-out rate by subject studied, again by gender. There are considerable differences in the drop-out rate by academic discipline. For example, the withdrawal rates for females for computer science (subject group GB) and for architecture (K) are 15% and 12% respectively. Conversely, the withdrawal rates for law and politics (M) and for literature and classics (Q) are less than 6%. For males, there is a minimum around 6% for law and politics and a maximum of about 14% for computer science.

Differences in drop-out rates by subject studied are likely to reflect, in part, differences in characteristics of students across subject area. Table 1 presents drop-out rates for males and females broken down by students' prior academic performance, by type of school and by social class of family background. Table 1 shows that the drop-out rate is lower the higher is the student's entry level of performance at A-level and varies with the type of school attended before university. (Average A-level points (out of 10) are defined as the total A-level points divided by the number of A-levels taken, and similarly for Highers scores.) In particular, the drop-out rate appears to be higher for students who had previously attended a local

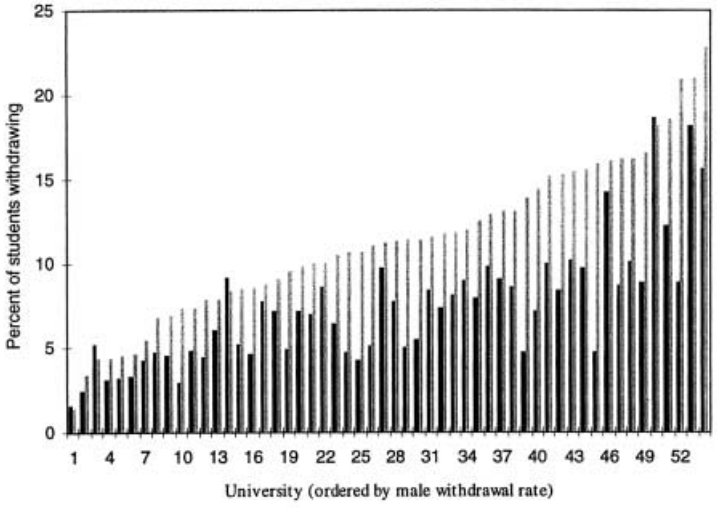

Fig. 1. Distribution of withdrawing students by university: ■, females; ■, males

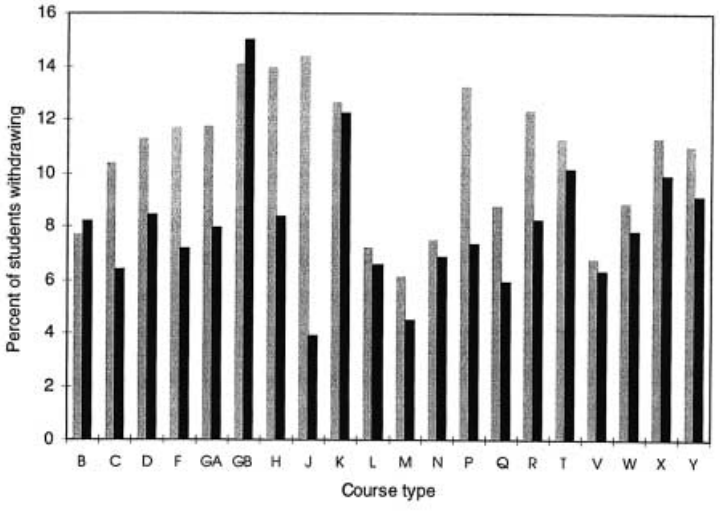

Fig. 2. Distribution of withdrawing students by course: ■, females; ■, males

Table 1. Drop-out rates by specific characteristics

	Females (%)	*Males (%)*
A-level scores		
> 8.5	3.73	4.25
7.5–8.5	4.68	6.59
5.5–7.5	5.77	9.61
0–5.5	8.09	15.31
School type		
LEA	7.27	10.82
Independent	5.96	9.08
Further education	5.65	9.35
Other	5.22	8.94
Parental occupation†		
SC I	5.57	8.22
SC II	6.37	9.53
SC IIINM	6.23	9.77
SC IIIM	7.38	10.42
SC IV	7.81	12.68
SC V	6.45	12.24
Unemployed	14.21	18.82

†SC I, professional; SC II, intermediate; SC IIINM, skilled non-manual; SC IIIM, skilled manual; SC IV, partly skilled; SC V, unskilled.

education authority (LEA) school. Table 1 also shows that the drop-out rate has a social gradient, especially for males.

3.2. Statistical modelling

We conduct a binomial probit regression analysis of the probability that an individual withdraws from university. (Our results are not sensitive to the choice between probit and logit estimation. Similarly, Dey and Astin (1993) found that results from logit and probit

models are essentially equivalent in their US sample.) An alternative competing risks framework, adopted by Booth and Satchell (1995) for British doctoral students, has not been chosen as we do not have reliable information on the reasons for leaving or on post-university destinations of withdrawing students, as we discuss below. We model the incidence of withdrawal, where withdrawal is defined as departure before a given point in the scheduled final year of the course.

The US literature suggests that first-year withdrawals may be different from others. Tinto (1987) emphasized the transitional difficulties of adjustment into college life and Porter (1990) showed that about half of all student attrition occurs in the first year. This is approximately the proportion that we find in our data: of the 7.1% of females (and 10.3% of males) who withdrew at some point before completion, 56% left during their first year, for both females and males. Tinto (1988) argued for a longitudinal approach in which the non-completion behaviour of students is time varying. DesJardins *et al.* (1999) used an event history model to analyse the temporal aspects of non-completion. We have conducted separate analyses for the probability of dropping out during the first year as well as for the incidence of dropping out in any year of study. We find that the results from both analyses are essentially the same and, in Section 4 later, we present the results based on the analysis of the incidence of withdrawal in any year of study.

As in related studies, we conduct separate analyses for male and female students. Johnes (1990), for example, found significant gender differences in the determinants of non-completion of courses by students. From the summary statistics discussed above, it appears that the drop-out behaviour for males and females is rather different and, indeed, a likelihood ratio test on the equality of the estimated coefficients from our separate models for males and females is rejected at less than the 0.01% level.

Students who drop out of university are likely to do so for diverse reasons. In particular, some may quit for 'negative' reasons related to disutility associated with their university studies whereas others may quit for 'positive' reasons related to alternative opportunities. Although all students leaving UK universities are asked to complete a first-destination survey indicating their labour market status in the year following their departure from university, very few university drop-outs respond to the survey: just 0.3% of drop-outs in our sample. Thus, the data do not provide information for inferring motives for withdrawal from university.

There is, however, direct information in the data regarding the student's reason for leaving, as reported by the university. The university records the withdrawing student as having left either because of academic failure or for 'other reasons'. None-the-less, for various reasons, we are not satisfied that this administrative coding is a reliable guide to the student's underlying motive for withdrawal. In particular, it is very likely that the first indication that a student has withdrawn will be their absence from examinations or their failure to submit other work. Whether this is — or should be — coded as academic failure or not is unclear, and the practice may vary across institutions. Similarly, students who are about to withdraw may have incentives to sit examinations so that, for example, they are not liable for the repayment of grants and bursaries made conditional on non-withdrawal. In this case, apparent academic failure does not capture the real reason for withdrawal. Conversely, students who are coded as withdrawing for other reasons may have calculated that they are about to fail and hence leave before sitting examinations. In view of this issue, we have conducted an analysis separately for students whose reasons for leaving were coded as 'other' and found that the results are essentially unchanged from those reported in the results section below.

Table 2. Probit models for withdrawing students

	Results for males		*Results for females*	
	Marginals	*p-value*	*Marginals*	*p-value*
Educational background				
Qualifications				
A-level points	−1.38	0.00	−0.51	0.00
Highers points	−1.03	0.00	−0.16	0.10
Relative A-level	−0.52	0.02	−0.79	0.00
Relative Highers	−0.39	0.00	−0.21	0.00
Subject preparedness				
Science	0.69	0.11	2.08	0.00
Other science	1.52	0.09	0.57	0.60
Social science	2.09	0.00	1.29	0.01
Languages	−0.71	0.52	0.55	0.40
Literature	0.47	0.58	−0.54	0.33
School type				
Further education college	−0.11	0.88	−0.31	0.63
Independent school	3.90	0.00	3.61	0.00
Other school	0.33	0.42	−0.77	0.03
Single-sex school	−1.03	0.02	−1.05	0.01
School points				
England and Wales	0.00	0.92	−0.09	0.03
Ireland	0.00	0.98	−0.09	0.00
Scotland	0.14	0.00	0.05	0.08
Total pupils	0.00	0.49	0.00	0.83
Personal characteristics				
Age 24–27 years	1.03	0.01	1.12	0.00
Age 28–33 years	2.72	0.00	2.89	0.00
Age $\geqslant$ 34 years	3.02	0.01	0.22	0.80
Married	−3.00	0.00	0.29	0.67
Parental occupation				
SC I	−0.95	0.01	−0.45	0.19
SC IIINM	−0.18	0.69	−0.19	0.66
SC IIIM	0.34	0.71	0.37	0.64
SC IV	0.25	0.82	0.34	0.72
SC V	−0.79	0.59	−1.20	0.41
Unemployed	0.80	0.48	1.00	0.32
Nationality				
European Union	3.73	0.00	1.18	0.00
Overseas	15.72	0.00	18.22	0.00
Self-funded overseas	−6.14	0.00	−3.75	
University residence				0.00
Parental home	2.77	0.00	1.79	0.00
Off campus	5.10	0.00	4.39	0.00

	Results for males		*Results for females*	
	Marginals	*p-value*	*Marginals*	*p-value*
Degree subject and related attributes				
Degree subject				
B, medical related	−1.13	0.20	−0.41	0.57
C, biological science	2.13	0.00	−0.84	0.12
D, agriculture	0.23	0.82	1.19	0.31
F, physical science	4.73	0.00	0.31	0.66
GA, mathematics	5.92	0.00	1.00	0.21
GB, computing	4.45	0.00	5.06	0.00
H, engineering	5.89	0.00	1.24	0.21
J, technology	3.82	0.00	−2.22	0.08
K, architecture	2.45	0.05	4.57	0.02
M, law and politics	−0.43	0.42	−0.94	0.07
N, business administration	0.57	0.37	0.57	0.37
P, communications	2.60	0.30	0.82	0.61
Q, literature and classics	2.58	0.00	1.50	0.02
R, modern European languages	4.28	0.00	3.31	0.00
T, other languages	1.92	0.26	5.18	0.00
V, humanities	−0.02	0.98	1.19	0.08
W, creative arts	0.16	0.89	0.39	0.67
X, education	0.38	0.79	0.67	0.49
Y, other subjects	5.23	0.00	2.06	0.01
Other degree information				
Department size	−0.71	0.01	−0.29	0.22
Joint degree	0.70	0.10	0.23	0.50
4-year degree	2.87	0.00	0.89	0.03
Sandwich degree	−3.10	0.00	−2.29	0.00
Integration				
Males (%)	0.00	0.90	0.02	0.20
Independent (%)	−0.05	0.04	−0.05	0.04
Fee paying (%)	−0.02	0.84	−0.05	0.48
Lower social class (%)	−0.03	0.23	0.00	0.86
Mature (%)	−0.02	0.70	−0.03	0.33
Overseas (%)	−0.02	0.77	−0.12	0.05
Economic activity				
Unemployment	0.22	0.00	0.17	0.01
Unemployment × lower social class	0.30	0.00	0.02	0.82
N	42407		33851	
% withdrawal	10.3		7.1	
Pseudo-R^2	0.098		0.090	

4. Results

Probit estimates of the probability of dropping out are presented in Table 2, separately for male and female students. The equation includes controls for educational background, personal characteristics, degree subject and related attributes, characteristics of the university department, unemployment in the county of prior residence and the university attended. Summary statistics on the means of each of the explanatory variables are presented in Table 3. We discuss the estimated marginal effects for individual universities in Section 5.

4.1. Educational background

With respect to educational background, Table 2 shows that the student's A-level points have

Table 3. Variables and mean summary statistics

Variable	*Results for males*	*Results for females*	*Variable*	*Results for males*	*Results for females*
Educational background			*Personal characteristics (continued)*		
Qualifications			Nationality		
British Technical Education Council etc.	0.046	0.024	UK	0.911	0.919
Other	0.032	0.043	European Union	0.031	0.035
No formal qualifications	0.031	0.036	Overseas	0.057	0.046
A-levels	0.821	0.821	Self-funded overseas	0.038	0.035
Highers	0.071	0.075	University residence		
A-level points	7.050	6.934	Living at home	0.096	0.112
Highers points	7.065	7.056	Living off campus	0.110	0.113
Subject preparedness					
Science	0.111	0.135	*Degree subject and related attributes*		
Other science	0.020	0.011	Degree subject		
Social science	0.076	0.141	B, medical related	0.017	0.047
Languages	0.016	0.030	C, biological science	0.063	0.112
Literature	0.042	0.042	D, agriculture	0.011	0.012
School type			F, physical science	0.132	0.068
LEA school	0.418	0.451	GA, mathematics	0.064	0.043
Further education college	0.100	0.106	GB, computing	0.054	0.011
Independent school	0.285	0.259	H, engineering	0.184	0.033
Other school	0.198	0.185	J, technology	0.015	0.008
Single-sex school	0.244	0.282	K, architecture	0.013	0.006
School points			L, social science	0.109	0.130
England and Wales	17.180	16.936	M, law and politics	0.080	0.085
Ireland	37.326	34.784	N, business administration	0.060	0.062
Scotland	48.802	49.403	P, communications	0.002	0.006
Total pupils	835.679	803.415	Q, literature and classics	0.044	0.107
			R, modern European languages	0.021	0.076
Personal characteristics			T, other languages	0.006	0.014
Age < 24 years	0.760	0.779	V, humanities	0.071	0.084
Age 24–27 years	0.177	0.141	W, creative arts	0.012	0.026
Age 28–33 years	0.040	0.035	X, education	0.006	0.026
Age $\geqslant 34$ years	0.023	0.044	Y, other subjects	0.037	0.043
Married	0.020	0.040	Other degree information		
Parental occupation			Department size	5.053	4.952
SC I	0.183	0.189	Joint degree	0.136	0.179
SC II	0.410	0.429	4-year degree	0.217	0.272
SC IIINM	0.114	0.108	Sandwich degree	0.073	0.057
SC IIIM	0.104	0.093			
SC IV	0.063	0.054	*Economic activity*		
SC V	0.011	0.008	Unemployment	5.887	6.011
Unemployed	0.069	0.079			

statistically significant effects on the drop-out probabilities for both males and females. For males, a point on the A-level average (equivalent to one extra A-level grade for a student with two A-levels) reduces the drop-out probability by about 1.4 percentage points, *ceteris paribus*. For females, the marginal effect, although statistically significant, is smaller. For males, significant effects are also associated with the average points score in Scottish or Irish Highers. The analysis also examines the effect of the students' A-level (or Highers) score relative to the average scored by all A-level (or Highers) candidates in the previous school attended. The results suggest that, for both males and females, the drop-out probability is lower for students who performed well relative to the average of their school peer group.

The analysis also attempts to take account of the effects of academic preparedness in the sense of the closeness of the match between the subject studied at university and the prior subjects studied at A-level. For specific subject areas, we define students as poorly prepared for their degree subject if they failed to have at least two relevant subjects among their prior qualifications. For the purposes of interpretation, the default, for each subject dummy variable, is a student who is well matched to the course that they are taking. The results suggest that the subject match is important only for science and social-science-based subjects. (Science includes biology, physical science, mathematics, engineering, medical related and technology. Other science is defined as agriculture, computing and architecture.) For example, students who are poorly matched to social science subjects at university are around 1.3 percentage points more likely to drop out in the case of females and 2 percentage points in the case of males. The lack of significance in the case of languages is likely to stem from the fact that significant familiarity is a prerequisite for most university language courses: in other words, there is relatively little variation in this variable. Of language students, only 1.6% of males and 3% of females are poorly matched to their degree subject. The results give support to the view that students who are well prepared for university are less likely to drop out, especially for students taking more science-based degree subjects.

Table 2 also shows the estimated effects of previous school characteristics on the individual student's drop-out probability. Of the school type variables, the independent school variable is significant in the models for both males and females. The coefficient implies that a student who had previously attended an independent school is about 4 percentage points more likely to drop out than a student who had previously attended an LEA school. Against this, students who had attended single-sex schools are 1 percentage point less likely to drop out. Control variables are included for the average performance of the last school attended against official Department for Education and Employment criteria. There is no clear pattern of effects. The size of school does not appear to have a significant effect on the drop-out probability.

Dummy variables were also included for other types of university entry qualifications, such as British Technical Education Council qualifications, other qualifications and no formal qualifications. There are some significant effects across the different types of entry qualification, though these are not reported.

4.2. Personal characteristics

Table 2 reports the effects of age, marital status, social class, fees status and term-time residence. Age is entered in the form of three age-band dummy variables. For both men and women, the results show that the drop-out probability largely increases with age. This may indicate that older students integrate less well into the social environment of UK universities. Alternatively, there may be differences in outside opportunities and responsibilities. We find that, for males, married students have a lower probability of dropping out, *ceteris paribus*.

Compared with a student from a social class II (SC II) background — i.e. where parental occupation was defined as technical or intermediate professional — there was no significant difference in the drop-out probability for students from other social class backgrounds with the exception of male students from SC I (professional occupations). These students are less likely to drop out. (Dummy variables for social class background are also interacted with unemployment, as is discussed below.) Parents of students from SC I backgrounds are most likely to have studied at university and this may increase the preparedness and motivation of such students to complete their degree studies. It will be interesting to compare our findings with data for cohorts entering university after the introduction in 1998 of 'home student' tuition fees.

Table 2 shows that non-UK European students are significantly more likely to drop out than are UK students, as are overseas students (with a very large marginal effect) — unless they are self-funding, in which case they are less likely to drop out even compared with UK students. Self-funding students may have a greater commitment to complete their studies.

Students who live at the parental address are around 2–2.5 percentage points more likely to drop out, and students who live off campus are around 5 percentage points more likely to drop out than are students who live on campus. These findings are consistent with the hypotheses of Tinto and others on the importance of social integration at university.

4.3. Degree subject and related attributes

Dummy variables were included for each of the standard degree subject groupings. It emerges that, relative to the omitted dummy variable for the study of social sciences, four and nine of the 19 subject dummy variables in the equations for females and males respectively have estimated coefficients that are significant at the 1% level, all positive, implying a higher drop-out rate than for social sciences, *ceteris paribus*. The four subject areas with significant effects for women are computer studies (with a marginal effect of 5.1% — implying that these students are 5.1 percentage points more likely to drop out relative to a social science student, *ceteris paribus*), modern European languages (3.3%), other languages (5.2%) and other subjects (2.1%). Additionally, the subject architecture and building has a positive marginal effect of 4.6%, which is significant at 5%, as do literary and classical studies (1.5%). The nine subject variables with estimated (positive) effects that are significant at the 1% level in the equation for males are biological science (with a marginal effect of 2.1%), physical science (4.7%), mathematics (5.9%), computing (4.5%), engineering (5.9%), technology (3.8%), literary and classical studies (2.6%), modern European languages (4.3%) and other or combined studies (5.2%). The estimated equation also includes a control for whether the individual was studying for a joint degree, but this proved insignificant.

Students taking a 4-year degree are more likely to drop out than are students on 3-year degrees. One could argue that if there is some fixed probability of dropping out each year — common across 3- and 4-year degree programmes — then this result could simply reflect an extra year effect, rather than a behavioural effect. (We note, however, that from our (unreported) regression of the probability of dropping out in the first year there remains a positive and significant effect associated with studying for a 4-year degree, implying different behaviour even *within* the first-year group.) The reported results also reveal a significant negative effect associated with studying on a 'sandwich' degree, i.e. a degree in which there is an intercalated year spent in a degree-related vocational activity. This effect could reflect the vocational nature of these degrees and a calculation that these degrees give better post-university employment prospects.

We have also investigated possible effects associated with the extent of the individuals' integration into the socioacademic life of the university department. We included the percentage of male students in the university department in the models for both males and females but found that this gender mix variable has no significant effects. Additionally, we interacted the dummy variable indicating whether the individual student had previously attended an independent school with a variable measuring the proportion of students in the university department who had previously attended such a school. We also included equivalent interaction terms for students from lower social class backgrounds, for mature students (i.e. students aged over 25 years at graduation), for overseas students and for fee paying students. The interaction term for the proportion of students from independent schools is significant, indicating that for a student who had come from the independent school sector the probability of dropping out of university is lower the greater the proportion of similarly educated students in the department. The only other significant interaction term is that for female overseas students: again the estimated negative effect implies a positive externality working for this subgroup. We also note that the effect of the total number of undergraduate students in the department is negative. This goes against an alienation-by-size hypothesis and may reflect a greater probability for each student to make a good social match the bigger is the pool of fellow students.

4.4. Unemployment

From Table 2, it appears that, for both male and female students, the county level unemployment rate has a well-determined positive effect on the individual drop-out probability. The estimated marginal effect implies that if the local unemployment rate is higher by 5 percentage points then the drop-out rate will be 1 percentage point higher. Given the possibility that the effect of local unemployment on the withdrawal rate might differ across different groups of students, we interact the local unemployment rate with a dummy variable for students from a low social class background (i.e. classes SC IIIM, SC IV, SC V (see Table 1 for definitions) and unemployed). For males, this interaction term is both positive and significant, indicating that students from lower social class backgrounds are more sensitive to local labour market conditions, as hypothesized in Section 2. That social class influenced the drop-out probabilities of students starting courses in 1989 underlines the importance of ensuring financial support for students from poorer backgrounds. A second corollary is that there is a potential clash of incentives for universities attempting to widen access while being constrained by performance indicators based on withdrawal rates.

Finally, we have examined the Oaxaca (1973) decomposition to investigate the extent to which male–female differences in the drop-out probability are attributable to differences in characteristics rather than in the parameters of the estimated drop-out function. The results imply that only around a sixth of the male–female difference can be explained by differences in observed characteristics. This confirms the hypothesis that student withdrawal behaviour is very different by gender.

The next section of the paper is concerned with deriving rankings of university performance from the university marginal effects estimated from the probit models for the probability of dropping out. As well as using these marginal effects to generate performance measures by institution, we have also regressed the estimated university marginal effects on the teaching quality assessment scores of each university, separately for males and females. We found a highly significant negative relationship, indicating that the probability of withdrawal diminishes as measured teaching quality rises. This gives support to the argument

by Tinto (1997) that classroom activity and interactions are likely to be an important, and underexplored, factor influencing the withdrawal of students.

5. Drop-out rates and university performance indicators

In announcing the publication of a performance indicator based on non-completion rates, the UK Government's Funding Council has stated that

> 'If all students qualified within the expected time, then the institution would be 100 per cent efficient'

(Higher Education Funding Council for England, 1999c). This view of university withdrawal is not uncontroversial. Although clearly costs are associated with university withdrawal both for institutions and for individuals (see, for example, the discussion of costs in DesJardins (1999)) it is unlikely that the optimal withdrawal rate is 0. Johnes and Taylor (1989) argued that non-completion does not mean that a student has received no benefit from their studies. Indeed, some withdrawing students may return to their studies at a later date. (Unfortunately, we cannot obtain an estimate of the proportion of withdrawing students who subsequently return. Students who maintain their registration while delaying their studies keep their personal identification code. In contrast, the records of students who withdraw are terminated. Any returning student is then allocated a new personal identification code.) Furthermore, a successful matching between degree courses and the abilities and preferences of students is likely to require adaptability such that some withdrawal is desirable on efficiency grounds. Similarly, indicators of non-completion can potentially conflict both with policies of widening access to higher education and with the maintenance of academic quality (see Cave *et al.* (1997)). These factors should be borne in mind when interpreting data on non-completion.

The university performance indicators published by the UK Government in December 1999 are intended to be an improvement on the publication of raw data which have tended to generate potentially misleading university league tables in the press and media. The published performance indicators provide context variables by which the performance of universities can be interpreted in the light of relevant circumstances, such as the school or social class backgrounds of the student intake. In the case of university drop-out rates, the data are at the aggregate university level and the method of adjustment involves only a small number of variables. Johnes and Taylor (1990) provided a method for the construction of university performance against the criterion of the non-completion rate using university level data: see also Johnes (1992). MacPherson and Paterson (1990) argued for university performance measures to be based on microdata but, at that time, large scale individual level data for whole cohorts were not available to academic researchers. Astin (1997), in an analysis of retention rates among US universities, argued that it is vital to control for the characteristics of students when assessing institutional performance. In the current paper, we have developed a model of the individual student drop-out probability based on administrative student level data. This can be used as the basis for the construction of a university performance indicator. In this section of the paper, we outline a method for constructing such a performance indicator for the non-completion rate. As importantly, we discuss some of the limitations of performance indicators in this area.

From the estimated university marginal effects, derived from the models reported in Section 4, we can derive rankings of university performance. Fig. 3 compares these adjusted university rankings with the raw or unadjusted rankings, as depicted in Fig. 1 for male students. Fig. 3 shows the sensitivity of the rankings to the adjustment for the control variables included in the underlying estimation procedure. Each point in Fig. 3 represents an

institution and the co-ordinates represent the adjusted and unadjusted rankings. Ranked number 1 on the horizontal axis is the university with the lowest unadjusted drop-out probability and, on the vertical axis, is the university with the lowest marginal effect estimated from the probit model presented in Section 4. If a university is observed on the 45° line, this indicates that its rank position does not change after controlling for student and course characteristics.

It is clear from Fig. 3 that the ranking of universities on the basis of the adjusted effects is very different from that based on the unadjusted data. The correlation between the adjusted and unadjusted rankings of universities is 0.66 for male students and the mean absolute movement is 9.4 rank places, with a substantial variation around this mean. The largest move is 39 places. Only five of the top-ranked universities on the unadjusted basis retain a top 10 ranking after adjustment. Similarly, only three of the bottom 10 maintain a bottom 10 ranking after adjustment.

An important question concerns how confident we can be in the exact ordering of universities in performance-based rankings. In the literature on the performance of schools, Goldstein and Spiegelhalter (1996) concluded that the best statistical analysis of school performance generates rankings such that most schools cannot be separated with any significant degree of confidence. This result, they argued, undermines the validity of the league table exercise for schools. To examine this issue in the context of our ranking of universities against the course non-completion criterion, we calculate the confidence intervals around the point estimates of each university's marginal effect on the withdrawal probability, relative to the median university.

Fig. 4 presents the 95% confidence intervals for the adjusted university marginal effects on the drop-out probabilities for males, ordered by their marginal effect. A line is drawn through the point estimate for the median university. This line cuts the majority of the confidence intervals, indicating that we can have little confidence in the rank position of most of the universities relative to the median. For males, only the top six and the bottom 12 universities perform significantly differently from the median.

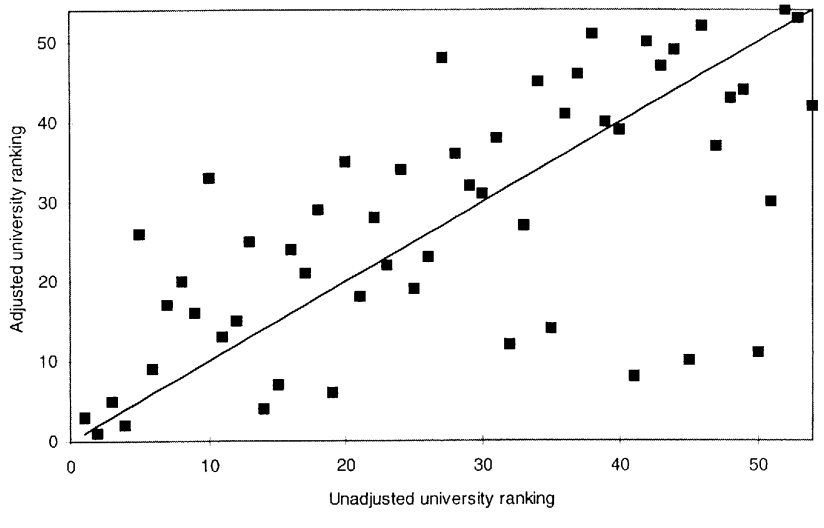

Fig. 3. Comparison of adjusted and unadjusted university rankings for males

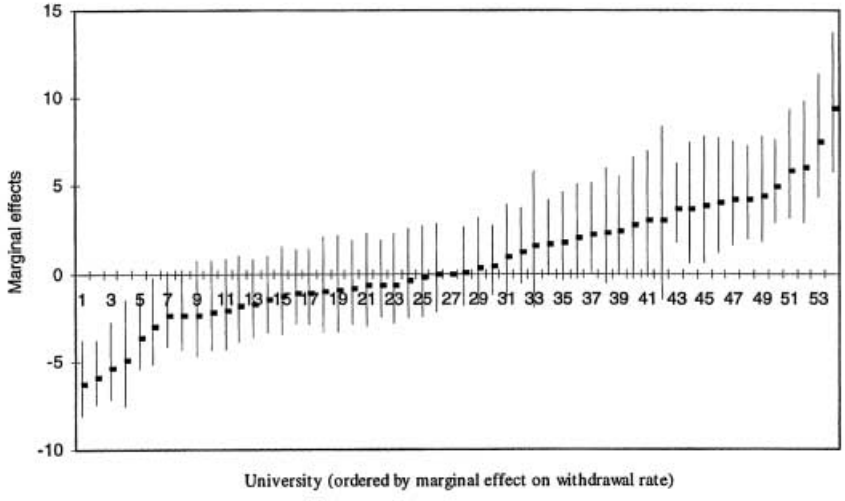

Fig. 4. 95% confidence intervals for university marginal effects on withdrawal for males

The results for females are very similar to those for males. For female students, the correlation between the adjusted and unadjusted rankings of universities is 0.77 and the mean absolute movement is 8.2 rank places, with a maximum move of 38 places. Seven of the top-ranked universities on the unadjusted basis retain a top 10 ranking after adjustment, and six of the bottom 10 maintain a bottom 10 ranking after adjustment. As for males, we can have little confidence in the rank position for females of most of the universities relative to the median: only three of the top four universities together with the bottom four universities perform significantly differently from the median.

6. Discussion

We have estimated a binomial probit model of the individual student's probability of withdrawal from university. The data set is based on the individual student level USR for 1990–1993 for the full population of undergraduate students, embarking on a 3- or 4-year degree programme in the autumn of 1989 at a pre-1992 university. We have matched these data both to school level information using data obtained from the Department for Education and Employment and from the Scottish and Northern Ireland Offices and to county level data on unemployment.

Among other results, we find that the probability of dropping out of university is influenced significantly by pre-university education, personal attributes, the degree subject and characteristics of the department and university. The county level unemployment rate also has a well-determined positive effect on the individual drop-out probability. On average, an increase in the local unemployment rate of 5 percentage points raises the drop-out probability by around 1 percentage point. For male students from lower social class backgrounds, the effect of local unemployment is more than twice as large, implying that the on-going participation of such students at university is much more sensitive to local labour market conditions. That social class influenced the drop-out probabilities of students starting courses in 1989 suggests the desirability of devising funding policies in such a way as to ensure financial support for poorer students.

The data that we have analysed provide a rich source of information on UK university students. None-the-less, there are limitations of the data and these have imposed constraints on the modelling procedure. First, the data set refers exclusively to students who entered university and, consequently, we have not been able to address the issue of selection into university. Second, the data do not capture students who substantially delay the completion of their studies and, ultimately, these students may be disproportionately more likely to fail to complete. However, we have suggested that such students are anyway relatively few in number. Third, some of the students who are observed to drop out may subsequently return to study later, perhaps elsewhere, but there are no administrative data with information on this. Fourth, although there are likely to be diverse reasons for dropping out of university, we have modelled the drop-out probability as a dichotomous variable rather than by adopting a competing risks framework, for example. This is because the data do not provide reliable information on the reason for withdrawal, either directly or indirectly through responses to the first-destination survey.

Finally, we have shown how the analysis of the individual student drop-out probability might be used to construct a university performance indicator. We have demonstrated that an adjusted ranking of universities, based on estimated university *ceteris paribus* marginal effects on the drop-out probability, is rather different from an unadjusted ranking. We have argued that a further limitation of drawing up a league table of universities based on the estimated marginal effects on the drop-out probability is that the confidence intervals are very wide around the individual university point estimates: we can have little statistical confidence in the rankings. This is so for the relatively homogeneous group of pre-1992 universities. It will be interesting to compare our results with an analysis of more recent cohorts for the expanded university sector. An issue which we do not pursue in the current paper concerns the question of what is the appropriate set of control variables to include for generating university performance indicators. For a discussion of this in the context of graduate labour market outcomes, see Smith *et al.* (2000).

Acknowledgements

We are grateful to Wiji Arulampalam, Keith Cowling, Geraint Johnes, Abigail McKnight, Mark Stewart and Jim Taylor and to the referees for helpful comments. Many people have given us invaluable help in generating the data set: in particular, we thank John McClure and Peggy Paull at the Universities Council on Admissions of Students, John McNeill at Warwick and staff at the Department for Education and Employment, the Higher Education Funding Council for England, the Higher Education Statistics Agency and the Schools Register. We acknowledge both the USR, as the original depositors, and the UK Data Archive for the use of the data set SN:3456 USR. None of these individuals or organizations bears any responsibility for any of the analysis or interpretation presented in this paper.

References

Astin, A. (1979) *Four Critical Years*. San Francisco: Jossey-Bass.

———(1997) How 'good' is your institution's retention rate? *Res. Higher Educ.*, **38**, 647–658.

Booth, A. L. and Satchell, S. E. (1995) The hazards of doing a PhD: an analysis of completion and withdrawal rates of British PhD students in the 1980s. *J. R. Statist. Soc.* A, **158**, 297–318.

Cave, M., Hanney, S., Henkel, M. and Kogan, M. (1997) *The Use of Performance Indicators in Higher Education*. London: Kingsley.

DesJardins, S., Ahlburg, D. and McCall, B. (1999) An event history model of student departure. *Econ. Educ. Rev.*, **18**, 375–390.
Dey, E. and Astin, A. (1993) Statistical alternatives for studying college student retention: a comparative analysis of logit, probit and linear regression. *Res. Higher Educ.*, **34**, 569–581.
Fielding, A., Belfield, C. and Thomas, H. (1998) The consequences of drop-outs on the cost-effectiveness of 16-19 colleges. *Oxf. Rev. Educ.*, **24**, 487–511.
Fitz-Gibbon, C. (1996) *Monitoring Education*. Trowbridge: Redwood.
Goldstein, H. and Spiegelhalter, D. J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). *J. R. Statist. Soc.* A, **159**, 385–443.
Higher Education Funding Council for England (1999a) Performance indicators in higher education. *Report 99/11*. Higher Education Funding Council for England, Bristol.
———(1999b) Performance indicators in higher education in the UK. *Report 99/66*. Higher Education Funding Council for England, Bristol.
———(1999c) *Press Release*, Dec. 3rd. Higher Education Funding Council for England, Bristol.
Johnes, G. (1992) Performance indicators in higher education: a survey of recent work. *Oxf. Rev. Econ. Poly*, **8**, 19–34.
Johnes, J. (1990) Determinants of student wastage in higher education. *Stud. Higher Educ.*, **15**, 87–99.
Johnes, J. and Taylor, J. (1989) Undergraduate non-completion rates: differences between UK universities. *Higher Educ.*, **18**, 209–225.
———(1990) *Performance Indicators in Higher Education*. Oxford: Oxford University Press.
———(1991) Non-completion of a degree course and its effect on the subsequent experience of non-completers in the labour market. *Stud. Higher Educ.*, **16**, 73–81.
Kalsner, L. (1992) Issues in College student retention. *Higher Educ. Extensn Serv. Rev.*, **3**.
Krueger, A. (1999) Experimental estimates of education production functions. *Q. J. Econ.*, **114**, 497–532.
Layard, R. and Psacharopoulos, G. (1974) The screening hypothesis and the returns to education. *J. Polit. Econ.*, **82**, 985–998.
MacPherson, A. and Paterson, L. (1990) Undergraduate non-completion rates: a comment. *Higher Educ.*, **19**, 385–390.
Noel, L. and Levitz, R. (eds) (1985) *Increasing Student Retention*. San Francisco: Jossey-Bass.
Oaxaca, R. (1973) Male-female wage differentials in urban labour markets. *Int. Econ. Rev.*, **14**, 693–709.
Porter, O. (1990) *Undergraduate Completion and Persistence in Four-year Colleges and Universities*. Washington DC: National Institute of Independent Colleges and Universities.
Riley, J. (1979) Testing the educational screening hypothesis. *J. Polit. Econ.*, **87**, 227–252.
Smith, J., McKnight, A. and Naylor, R. (2000) Graduate employability: policy and performance in higher education in the UK. *Econ. J.*, **110**, 382–411.
Spence, M. (1973) Job market signalling. *Q. J. Econ.*, **87**, 354–374.
Tinto, V. (1975) Dropout from higher education: a theoretical synthesis of recent research. *Rev. Educ. Res.*, **45**, 89–125.
———(1987) *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press.
———(1988) Stages of student departure: reflections on the longitudinal character of student leaving. *J. Higher Educ.*, **59**, 438–455.
———(1997) Classrooms as communities: exploring the educational character of student persistence. *J. Higher Educ.*, **68**, 599–623.