J. R. Statist. Soc. A (2004)
167, *Part* 3, *pp.* 475–496

Higher education outcomes, graduate employment and university performance indicators

Massimiliano Bratti,

University of Warwick, Coventry, UK

Abigail McKnight

London School of Economics and Political Science, UK

and Robin Naylor and Jeremy Smith

University of Warwick, Coventry, UK

[Received January 2003. Revised December 2003]

Summary. Official employment-related performance indicators in UK higher education are based on the population of students responding to the 'First destination supplement' (FDS). This generates potentially biased performance indicators as this population of students is not necessarily representative of the full population of leavers from each institution. University leavers who do not obtain qualifications and those who do not respond to the FDS are not included within the official analysis. We compare an employment-related performance indicator based on those students who responded to the FDS with alternative approaches which address the potential non-random nature of this subgroup of university leavers.

Keywords: Employment; Higher education; Non-random response; Performance indicators

1. Introduction

There is now a substantial statistical literature on the issue of performance measures for public sector institutions. This literature has grown in recent years in the context of an increasing emphasis of public policy on institutional auditing and surveillance. In the education sector in the UK, the publication of performance-based league tables of schools, colleges and universities has become an important aspect of public policy. This is also true of other sectors, especially health, and is an increasingly common feature of both regulated and quasi-markets in many countries. In the absence of price competition or measures of profitability in such markets, performance tables are typically intended to provide information that is relevant to consumer choice and to improve accountability. For secondary schools in England and Wales, annual rankings of schools that are based on public examination results first appeared in 1992. This was extended to primary schools in 1998. Following the recommendations of the National Committee of Inquiry into Higher Education (Dearing, 1997), the UK Government began to publish performance indicators (PIs) for universities and other higher education institutions in December 1999.

Whereas league tables of schools focus primarily on performance in examinations, university performance is measured against a wide range of criteria. An assessment of the quality

Address for correspondence: Jeremy Smith, Department of Economics, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: Jeremy.Smith@warwick.ac.uk

 0964–1998/04/167475

of both research and teaching in UK universities is well established. The Higher Education Funding Council for England (HEFCE) has now also published sets of higher education PIs to cover

(a) access and participation,
(b) retention and progression,
(c) research and
(d) employability.

In the literature on the measurement of school performance it is generally accepted that school performance tables are potentially misleading if based on crude output measures with no adjustment for 'contextual' factors such as school intake. Goldstein and Spiegelhalter (1996) in particular expressed strong concern with the publication of unadjusted league tables of institutional performance in both the health and the education sectors. An awareness of the need to take account of the prior academic achievements of pupils has led to an emphasis on 'value-added' measures of school examination performance. The Government now publishes a value-added indicator of school performance that is based on performance differences between key stages 2 and 3 (see Department for Education and Employment (1995) and Department for Education and Skills (2003) for details).

Methodological issues regarding the construction of PIs for higher education institutions (HEIs) have also received increasing attention. There has been a lively debate concerning the measurement of research output (see for example, Johnes *et al.* (1993), Johnes and Taylor (1990), Goedegebuure *et al.* (1990) and Cave *et al.* (1997)). The publication of official university PIs on access, progression and employability has stimulated a renewed research interest in other aspects of university performance measurement: see, for example, Smith *et al.* (2000), Smith and Naylor (2001a,b), Johnes and McNabb (2003) and McNabb *et al.* (2002).

The current paper is concerned with a unique aspect of university PIs based on post-university outcomes. Whereas derived PIs on progression are based entirely on information that is contained in administrative records—as is typically the case with institutional PIs—indicators for employability are based on survey data, as we describe more fully below. Information on the post-university employment status of university leavers is obtained from responses to a 'First destination supplement' (FDS) survey that is sent by each institution to a target population of individuals in its leaving cohort. This target population comprises all home-domiciled students obtaining an undergraduate level qualification and who studied full time. Hence, the target population excludes students who leave with no qualifications.

The response rate to the FDS is approximately 84% on average but does vary across institutions. We note that the probability of a response from a student is unlikely to be a random event: rather, it is likely to be correlated with the individual's performance at university and with their post-university labour market outcome. This means that an institutional performance measure that is based solely on the population of respondents, in other words, conditional on responding (as is the case with the official employment PI), will be a biased estimate of the *unconditional* performance, i.e. of that not conditional on response. We are not aware of any previous work on this issue in the context of institutional performance measurement.

In the current paper, we conduct an analysis of the determinants of graduate first-destination outcomes and from this derive employment-based university PIs. We relate our analysis to that which is implicit in the official employment PI, and we pay particular attention to aspects relating both to sample selection and to non-response. Given that the average response rate across institutions is relatively high on average (for example, the Labour Force Survey has a response rate of 79% for the latest wave 1), it might be unsurprising if we find no substantial difference in

measured performance according to whether we look at conditional or unconditional measures. Thus, our data provide a reasonably strong test of the sensitivity of measured performance to conditioning on response.

The rest of the paper is set out as follows. In Section 2, we provide a brief overview of the data and methodology that we employ in our analysis. Section 3 discusses the results from the various stages of the analysis of individual graduates' labour market outcomes. Section 4 then describes how the results of the individual level analysis can be used in the construction of university level PIs. Section 5 considers an alternative approach which corrects for selection bias. Section 6 addresses some issues relating to the sensitivity of the results and Section 7 closes the paper with conclusions and further remarks.

2. Data and methodology

By far the largest survey of university leavers is the FDS, a survey of full-time undergraduate leavers from UK HEIs, conducted by the careers offices of each university and deposited with the Higher Education Statistics Agency (HESA). The response rate to the FDS survey varies markedly across institutions, but for the 1998 returns averaged around 84%. These responses can be matched to individual administrative records for each student. The survey is conducted approximately 6 months after students have successfully completed their course and identifies, *inter alia*, the main activity of the student leaver at that time. For UK students, the main activity can be grouped into four categories:

(a) entering employment or self-employment (E);
(b) undertaking further study, training or research (FS);
(c) unemployed and seeking work or further study (U);
(d) inactive—unavailable for (or not seeking) employment or further study (I).

The higher education official employment indicators are derived from the FDS. The most recently published PI for employment is that of December 2002 and is based on the 2000–2001 survey (see Higher Education Funding Council for England (2002)). In the current paper, we exploit individual HESA and FDS data for 1997–1998, the most recent cohort for which we have appropriate data.

On the basis of the information that is contained in the FDS and the matched HESA data, we present a method for the construction of a set of employment-related university performance measures. First, we distinguish between 'positive' outcomes (categories E and FS above, denoted EFS) and 'negative' outcomes (categories U and I together, denoted UI). We note that this contrasts with the work conducted by the HEFCE who choose to exclude students who report that they are 'not available for (or not seeking) employment, study or training'. It is not clear why this group should be excluded: if universities are to be assessed on their ability to prepare graduates for the labour market then arguably this group should be included. Omission generates a possible bias as this group, in part, captures 'hidden' unemployment which is unlikely to be randomly distributed across HEIs, whereas the inactive group who are travelling or ill should either be random or picked up in the modelling exercise. We note that there is considerable variation in the percentage of graduates who are 'not available for (or seeking) employment, study or training' across HEIs and hence we include this category along with the unemployed. However, the sensitivity of our results to the inclusion of this inactive group with those unemployed is addressed in Section 6 of the paper.

In our analysis, we match the FDS responses of the individual student to administrative records, which are rich in information on the student's higher education record (including

institution and course details) and prior qualifications and personal characteristics (including social class background). Therefore, when estimating the university effects in these models (for the construction of performance measures), we can control for a large amount of information on the characteristics of the students within each university.

As our aim is to assess both the effect of students' characteristics and the effect of the institution on the first-destination outcome, our concern focuses on the possible non-random nature of the sample based on those students who responded to the FDS. A conditional analysis simply focusing only on the final stage of the process from qualification through to the post-university employment outcome may give misleading inferences if the final sample of students is not a random sample of the population of students from an institution.

In this paper, we examine two alternative approaches by which to construct university performance indicators for graduates' labour market outcomes. The first is based on a series of sequential probit models for the probabilities of

(a) obtaining a qualification,
(b) responding to the FDS survey and
(c) being in employment or further study.

Under the assumption of independence across the error processes in the three models, this approach takes into account differences in universities' qualification and FDS response rates in generating an adjusted PI. This can then be compared with the HEFCE-like approach which considers only those who respond to the FDS with no adjustment for differential response rates across institutions. The results from this approach are presented in Sections 3 and 4 of the paper.

The second approach that we present for the construction of university PIs allows for non-independence in the error processes across the equations for responding to the FDS and the probability of being in employment or further study after university. Results from this approach are discussed in Section 5. Of course, this second approach is a generalization of the first. We choose to examine the approaches separately, however, as we believe it plausible that the official employment-based PI measures for universities might evolve to take account of differences in response rates under a methodology such as that based on independent probit models for response and employment outcome. We think it unlikely, however, that official indicators would be based on a methodology which corrects for endogenous selection problems. None-the-less, we think it of interest to examine this issue.

We now sketch more formally the basic methodological approach that we adopt for the development of adjusted indicators of HEI performance. We exploit the individual level data for the cohort of 1997–1998 leavers to model the probability that the student leaver obtains a qualification (OQ):

$$y_{1i} = \begin{cases} 1 & \text{obtains a qualification,} \\ 0 & \text{otherwise,} \end{cases}$$

where

$$y_{1i} = I(x'_{1i}\beta_1 + \sum_j H_{ij}\alpha_{1j} + u_{1i} > 0)$$

and

$$E(y_{1i}) = \Pr(\mathrm{OQ} = 1),$$

where i and j are subscripts for individuals and universities respectively. $I(\cdot)$ is the indicator function and assigns the value 1 to a true statement and the value 0 to a false statement. x_{1i} is a vector of individual characteristics and H_{ij} a university dummy variable which takes the

value 1 if individual i attended university j and 0 otherwise, with one university dummy variable excluded as the 'default' university, with respect to which the university effects α_{1j} are estimated. We assume that $u_{1i} \sim N(0,1)$. This simple probit essentially models the probability of both non-completion as well as the probability of failing to obtain a qualification.

For those individuals who are in the target population, we model the probability of an individual responding (R) to the FDS by the simple probit:

$$y_{2i} = \begin{cases} 1 & y_{1i} = 1 \text{ and responded to FDS,} \\ 0 & y_{1i} = 1 \text{ and did not respond to FDS,} \end{cases}$$

where

$$y_{2i} = I(x'_{2i}\beta_2 + \sum_j H_{ij}\alpha_{2j} + u_{2i} > 0),$$
$$E(y_{2i}|y_{1i} = 1) = \Pr(R = 1|\mathrm{OQ} = 1),$$
$$u_{2i} \sim N(0,1).$$

Finally, for those who responded to the FDS, we model the probability that the student is either employed or in further study (EFS), rather than unemployed or inactive (UI), again as a simple probit:

$$y_{3i} = \begin{cases} 1 & y_{2i} = 1 \text{ and in EFS,} \\ 0 & y_{2i} = 1 \text{ and not in EFS,} \end{cases}$$

where

$$y_{3i} = I(x'_{3i}\beta_3 + \sum_j H_{ij}\alpha_{3j} + u_{3i} > 0),$$
$$E(y_{3i}|y_{2i} = 1, y_{1i} = 1) = \Pr(\mathrm{EFS} = 1|R = 1, \mathrm{OQ} = 1),$$
$$u_{3i} \sim N(0,1).$$

Consequently, the *unconditional* probability of being, for example, in employment or further study is equal to the conditional probability of EFS multiplied by the conditional probability of replying, multiplied by the marginal probability of obtaining a qualification, i.e.

$$\Pr(\mathrm{EFS} = 1, R = 1, \mathrm{OQ} = 1) = \Pr(\mathrm{OQ} = 1)\Pr(R = 1|\mathrm{OQ} = 1)\Pr(\mathrm{EFS} = 1|R = 1, \mathrm{OQ} = 1). \quad (1)$$

Initially, we assume that the error processes across the three equations for OQ, R and EFS are independent, i.e. $\mathrm{corr}(u_{1i}, u_{2i}) = \mathrm{corr}(u_{1i}, u_{3i}) = \mathrm{corr}(u_{2i}, u_{3i}) = 0$. In this case the three probit models can be estimated separately. We can then calculate an individual university marginal effect on the unconditional probability of EFS, relative to the default or reference case, by calculating the difference in expression (1) for the individual university and for the default university, using sample mean values for all other determinants. Similarly, we can calculate the university marginal effect for the conditional probability of EFS by evaluating $\Pr(\mathrm{EFS} = 1|R = 1, \mathrm{OQ} = 1)$ for the individual university relative to the default university.

This model, which assumes independence across the error processes in the three equations, is similar to the so-called two-part model that is described in Cragg (1971) and Manning *et al*. (1987). The main advantage of estimating such a model, as opposed to using a full information maximum likelihood or Heckit-type estimator (see Heckman (1979)), is that the two-part model does not rely on any exclusion restriction for model identification (e.g. on variables that are specific to the different stages of the model and which are excluded from the other stages). Puhani (2000) finds that, because of the quasi-linearity of the inverse Mill's ratio and the consequent strong correlation between the variables that are included in the outcome

Table 1. Summary statistics by stages in the EFS outcome†

Variable	*Results for males*				*Results for females*			
	All	*OQ*	*R*	*EFS*	*All*	*OQ*	*R*	*EFS*
Prior qualifications								
HNC, HND	0.075	0.078	0.075	0.074	0.051	0.050	0.047	0.046
A-levels (default)	0.573	0.576	0.577	0.579	0.573	0.576	0.577	0.579
Highers	0.032	0.030	0.030	0.029	0.034	0.033	0.033	0.033
ONC, OND	0.054	0.052	0.050	0.049	0.036	0.035	0.033	0.033
Not known	0.096	0.095	0.093	0.093	0.090	0.089	0.086	0.086
Other	0.169	0.148	0.142	0.139	0.187	0.172	0.164	0.161
Point scores								
A-levels	18.6	19.2	19.3	19.5	20.0	20.2	20.3	20.5
Age groupings (years)								
<24 (default)	0.739	0.741	0.761	0.766	0.755	0.760	0.776	0.784
24–26	0.106	0.109	0.100	0.099	0.078	0.078	0.072	0.071
27–30	0.062	0.060	0.054	0.054	0.051	0.049	0.043	0.042
⩾31	0.093	0.090	0.085	0.081	0.117	0.113	0.108	0.103
Social class								
I	0.095	0.101	0.105	0.107	0.092	0.097	0.101	0.102
II (default)	0.232	0.241	0.247	0.248	0.248	0.254	0.259	0.261
IIINM	0.067	0.068	0.070	0.071	0.075	0.076	0.077	0.077
IIIM	0.097	0.096	0.098	0.098	0.097	0.096	0.098	0.098
IV	0.044	0.043	0.043	0.042	0.043	0.042	0.042	0.043
V	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009
Armed forces	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005
Not known	0.442	0.431	0.418	0.415	0.424	0.415	0.404	0.400
Retired	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Unemployed	0.005	0.004	0.003	0.003	0.005	0.005	0.004	0.004
Disability								
No disability (default)	0.879	0.875	0.874	0.874	0.879	0.876	0.874	0.875
Dyslexia	0.013	0.012	0.012	0.012	0.008	0.008	0.008	0.007
Blind	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001
Deaf	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Wheel-chair	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Mental	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001
Diabetes	0.021	0.021	0.022	0.022	0.026	0.025	0.026	0.026
Multiple	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Other	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003
Not known	0.075	0.081	0.082	0.082	0.077	0.082	0.083	0.083
Ethnicity								
White (default)	0.762	0.774	0.791	0.795	0.777	0.783	0.796	0.803
Black	0.028	0.022	0.017	0.015	0.033	0.029	0.023	0.021
Indian	0.059	0.053	0.047	0.046	0.050	0.049	0.045	0.043
Chinese	0.019	0.019	0.017	0.016	0.017	0.017	0.016	0.015
Other	0.014	0.012	0.011	0.011	0.014	0.013	0.012	0.012
Not known	0.118	0.120	0.117	0.116	0.109	0.109	0.107	0.105
Course type								
Medical related	0.026	0.027	0.028	0.030	0.080	0.081	0.083	0.088
Biological science	0.063	0.064	0.066	0.065	0.097	0.099	0.102	0.099
Agriculture and related	0.010	0.010	0.010	0.010	0.011	0.012	0.011	0.012
Physical science	0.084	0.086	0.089	0.088	0.048	0.049	0.051	0.051
Mathematical science	0.108	0.100	0.100	0.103	0.034	0.033	0.033	0.033
Engineering	0.139	0.135	0.137	0.140	0.018	0.017	0.017	0.017
Technology	0.010	0.011	0.011	0.010	0.006	0.006	0.006	0.006
Architecture and building	0.044	0.043	0.044	0.045	0.011	0.011	0.010	0.011
Social science	0.038	0.038	0.037	0.035	0.082	0.080	0.078	0.076

(*continued*)

Table 1 (*continued*)

Variable	*Results for males*				*Results for females*			
	All	*OQ*	*R*	*EFS*	*All*	*OQ*	*R*	*EFS*
Economics	0.025	0.027	0.027	0.026	0.009	0.010	0.010	0.010
Politics	0.019	0.020	0.020	0.019	0.015	0.016	0.016	0.015
Law	0.034	0.036	0.037	0.039	0.045	0.046	0.046	0.048
Business (default)	0.112	0.112	0.109	0.109	0.112	0.111	0.109	0.108
Communication	0.012	0.012	0.011	0.011	0.016	0.016	0.015	0.015
Literature and classics	0.031	0.033	0.033	0.033	0.071	0.074	0.075	0.076
Modern European languages	0.014	0.014	0.014	0.014	0.032	0.033	0.034	0.034
Humanities	0.044	0.048	0.048	0.047	0.048	0.051	0.051	0.051
Creative arts	0.055	0.056	0.054	0.050	0.071	0.071	0.068	0.066
Education	0.022	0.022	0.023	0.024	0.057	0.057	0.060	0.063
Other subjects	0.110	0.104	0.102	0.100	0.136	0.129	0.125	0.123
Study method								
Non-sandwich (default)	0.851	0.857	0.855	0.848	0.905	0.907	0.905	0.902
Sandwich	0.149	0.143	0.145	0.152	0.095	0.093	0.095	0.098
Accommodation type								
Own accommodation (default)	0.374	0.397	0.399	0.398	0.389	0.406	0.410	0.408
Institution	0.195	0.185	0.187	0.191	0.184	0.175	0.178	0.182
Parent	0.127	0.117	0.120	0.120	0.115	0.110	0.112	0.111
Other	0.160	0.170	0.169	0.170	0.164	0.172	0.171	0.173
Not known	0.144	0.130	0.124	0.121	0.148	0.137	0.129	0.126
Degree class								
First	—	0.083	0.090	0.097	—	0.072	0.077	0.080
Upper second (default)	—	0.400	0.418	0.425	—	0.502	0.516	0.518
Lower second	—	0.354	0.350	0.341	—	0.331	0.323	0.320
Third	—	0.065	0.061	0.057	—	0.033	0.030	0.028
Other	—	0.098	0.081	0.080	—	0.063	0.054	0.054
N	104148	84542	69767	60787	104437	89948	77263	68601

†All, all undergraduate students aiming at an undergraduate degree; OQ, obtained a qualification; *R*, responded to the FDS; EFS, student in employment or further study 6 months after graduation; HNC, Higher National Certificate; HND, Higher National Diploma; ONC, Ordinary National Certificate; OND, Ordinary National Diploma.

regression and the inverse Mill's ratio when no exclusion restriction is used, the Heckit and full information maximum likelihood estimators are not robust and the two-part model is a preferable alternative.

In Section 5 of the paper, we allow for the possibility that the probabilities that are associated with EFS and with response to the FDS are not independent. We use a probit model which corrects for endogenous sample selection and use this to examine the sensitivity of the derived university marginal effects to selection correction. We estimate this model by using the `heckprob` command in STATA.

3. Modelling the probabilities of qualification, response and employment outcome: independent probits

As described in the previous section, we now estimate a sequence of three probit models, under the assumption of independence. In Section 3.1, we model the probability of obtaining a qualification (OQ). In Section 3.2, for those students who obtained a qualification, we model the

probability of responding (R) to the FDS and, in Section 3.3, for those responding to the FDS, we model the probability of being in employment or further study (EFS).

3.1. Modelling graduates' probabilities of obtaining a qualification

Individuals responding to the FDS must have originally been in the target population, as defined by the HEFCE. The target population is defined over home-domiciled graduates who studied full time, qualifying with an undergraduate qualification. Table 1 presents summary statistics for the main variables broken down by gender and according to the different stages of the analysis, i.e. over

(a) all students,
(b) those obtaining a qualification (OQ),
(c) those responding to the FDS (R) and
(d) those in employment or further study (EFS).

Following the HEFCE, the sample is restricted to home-domiciled students who studied full time. We also exclude medical students from the analysis as there is essentially no variation in their reported main activity on leaving university. Additionally, we consider only the set of 'old or traditional' universities (i.e. those predating the abolition of the pre-1992 divide between universities and polytechnics) and the 'new' (former polytechnic) universities. Hence, we do not include other HEIs such as colleges of further education.

As can be seen from Table 1, the total numbers of UK-based full-time students in the cohort of 1997–1998 leavers from all universities was 104 148 and 104 437 for male and female students respectively. The majority of students had taken A-level qualifications before university. A-levels are the typical academic qualification of those who complete formal secondary education at age 18 years in England and Wales; students typically take three A-levels and each A-level pass is graded from A to E, with A being worth 10 points and E 2 points. In Scotland students take Scottish Highers. Students typically take five Highers, with each Higher graded from A to C. Smaller numbers had taken Higher National Diplomas or Ordinary National Diplomas, or Higher National Certificates or Ordinary National Certificates. About 82% and 86% of male and female university students respectively obtained an undergraduate degree level qualification. From Table 1, we note that, compared with all students, those who obtain such a qualification (OQ) had higher pre-university A-level points on average and tended to be younger and from higher social class and ethnic white backgrounds.

The probability that the student obtains a qualification, and is therefore in the target population for the FDS, is modelled by using probit regression. The analysis is conducted separately for men and women: a likelihood ratio test for the equality of coefficients for this data set easily rejects the hypothesis of no differences by gender. Full sets of regression results can be obtained for all the models that are discussed in this paper from an earlier working paper version of this paper (see Bratti *et al.* (2003)). In the current paper, we focus the discussion on selected results.

We find that, for males and females, increasing A-levels by one grade—i.e. by 2 points—increases the probability of obtaining a qualification by about 1 and by $\frac{1}{2}$ percentage point respectively. This finding is in line with the results of Smith and Naylor (2001a), who found that A-levels had a greater effect on reducing the probability of dropping out for males compared with females. Compared with students who were less than 24 years old when they left university, older students are significantly more likely to obtain a qualification. There are some significant social class effects. Relative to students from a social class II (managerial and technical occupations) background, students from social class I (professional occupations) are more

likely to obtain a qualification whereas students from social class IIIM (skilled manual occupations), social class IV (partly skilled occupations) and social class V (unskilled occupations) are typically less likely to obtain a qualification, regardless of sex. We also find some strong ethnicity effects. Relative to white students, students who are black or, in the case of males, Indian or Pakistani are less likely to obtain a qualification.

We also find that strong effects are associated with the degree subject that is studied. Our principal focus, however, concerns the large and significant institutional effects for both males and females. For males, 34 out of 101 institutions have a significant and positive marginal effect, relative to the default case, in the sense that their confidence interval does not encompass the point estimate of the default (i.e. the median) university. 28 have a significant negative effect. For females, the corresponding numbers are 35 and 26. A stricter condition for the significance of the ranking of institutions would be based on non-overlapping confidence intervals (see, for example, Goldstein and Spiegelhalter (1996)). Under this criterion, it emerges that there is evidence of considerable uncertainty in the ranking for institutions—especially away from the extremes of the ranking. For example, the confidence intervals for the 47th and 59th ranked university for males and females respectively are overlapping with all those ranked as high as 23rd and 31st and as low as 83rd and 91st respectively.

We conclude from this analysis that ignoring the selection of the target group, as does the official HEFCE PI methodology by restricting the analysis to students who successfully graduate with a first degree, risks masking some important institutional differences.

3.2. Modelling the probability of responding to the first-destination supplement

For those individuals who have obtained a qualification (OQ), an FDS questionnaire is sent out to graduates around 6 months after they have successfully completed their degrees to seek information on their main activity at that time. The careers services within universities are given the responsibility of collecting the FDS information and ensuring as high a response rate as possible to this survey. Students who do not respond to the questionnaire might receive a further questionnaire, or alternatively may receive follow-up telephone calls for the survey to be as comprehensive as possible. However, the extent to which universities attempt to track their students may differ and this is clearly a strategic decision as those likely to have a lower employment rate are also less likely to respond. Given this, it might not be in an institution's interests to invest substantial resources in actively pursuing non-respondents. In fact, there appears to be little punishment for submitting a low response rate. The HESA view 80% as an expected response rate. In the published HEFCE employment PIs, only two institutions were excluded because of a low response rate. We note that a substantial range of response rates was reported in the HEFCE employment indicators for 2001: 98% for Cranfield University, 94% for Imperial College, 93% for Warwick University, 60% for Coventry University, 63% for Middlesex University, 66% for Staffordshire University and 72% for Oxford Brookes, Anglia Polytechnic University, King's College London and Thames Valley University.

From Table 1, we note that, of those obtaining a qualification, 83% and 86% responded to the FDS, giving a sample of 69 767 and 77 263 male and female students respectively for the subsequent analysis of first-destination outcomes. Table 1 also shows that students who graduate with a good degree are more likely to respond to the FDS survey than are other students. For example, 9% of males who responded had obtained a first-class degree, compared with 8.3% of those who obtained a qualification—and hence were in the target population for the FDS survey.

The results indicate that there are few clear effects of prior qualifications on the response probability for both males and females, although there are some differences according to the type of qualification. For males and females respectively an extra 2 A-level points (one extra

Table 2. Marginal effects on the probability of employment or further study (EFS)—males and females†

Variable‡	*Marginal effects (%) for males*			*Marginal effects (%) for females*		
		95% confidence interval			*95% confidence interval*	
		Lower	*Upper*		*Lower*	*Upper*
Prior qualifications (A-levels)						
A-level points score	0.19	−3.34	3.72	−0.01	−0.06	0.04
Age groupings (<24) (years)						
24–26	−0.87	−1.75	0.02	−1.50	−2.43	−0.56
27–30	1.39	−0.48	3.26	−3.04	−5.04	−1.05
⩾31	−2.16	−4.28	−0.04	−5.93	−7.88	−3.98
Social class (II)						
I	0.79	−0.55	2.12	0.67	−0.40	1.75
IIINM	1.30	0.03	2.58	0.35	−0.73	1.43
IIIM	0.53	−0.75	1.81	0.22	−0.82	1.26
IV	0.37	−1.16	1.89	1.44	0.23	2.66
V	1.25	−1.22	3.73	0.63	−1.64	2.89
Armed forces	0.95	−2.21	4.11	2.85	0.09	5.61
Not known	0.41	−0.68	1.50	−0.21	−1.10	0.69
Retired	−2.93	−10.56	4.71	0.25	−6.25	6.74
Unemployed	−1.66	−5.96	2.64	0.55	−2.47	3.57
Disability (no disability)						
Dyslexia	−0.66	−3.48	2.16	−1.61	−4.51	1.28
Blind	−3.15	−9.88	3.58	1.87	−4.05	7.80
Deaf	−6.76	−13.06	−0.46	−2.75	−7.61	2.10
Wheel-chair	−7.76	−15.89	0.36	3.35	−1.48	8.18
Mental	−16.08	−30.45	−1.72	−4.51	−13.88	4.86
Diabetes	−0.60	−2.86	1.65	1.00	−0.66	2.65
Multiple	−7.47	−15.52	0.58	−5.86	−13.86	2.13
Other	−3.48	−7.99	1.03	−0.48	−4.18	3.22
Not known	−2.60	−4.48	−0.72	−0.77	−2.23	0.69
Ethnicity (white)						
Black	−3.75	−6.35	−1.16	−2.16	−4.07	−0.26
Indian	−2.96	−4.88	−1.04	−4.94	−6.69	−3.20
Chinese	−3.39	−6.00	−0.79	−3.85	−6.08	−1.63
Other	−1.47	−4.20	1.27	−0.54	−2.67	1.60
Not known	−0.52	−1.93	0.90	−1.49	−2.70	−0.27
Course type (business)						
Medical related	6.98	5.66	8.30	6.16	5.26	7.06
Biological science	−0.29	−1.89	1.31	−1.56	−2.94	−0.18
Agriculture and related	−0.68	−3.44	2.08	1.72	−0.41	3.85
Physical science	0.66	−0.84	2.15	−0.27	−1.79	1.24
Mathematical science	4.08	2.86	5.30	0.34	−1.26	1.95
Engineering	3.40	2.15	4.66	0.41	−1.52	2.35
Technology	1.29	−1.15	3.74	2.79	0.20	5.37
Architecture and building	4.18	2.84	5.52	2.02	−0.08	4.12
Social science	−2.60	−4.62	−0.58	−1.08	−2.50	0.33
Economics	−0.51	−2.53	1.52	0.27	−2.14	2.67
Politics	−2.38	−4.72	−0.04	−1.57	−3.71	0.57
Law	3.71	2.10	5.31	3.85	2.66	5.04
Communication	−0.10	−2.69	2.50	−2.87	−5.20	−0.54
Literature and classics	−1.45	−3.47	0.57	0.45	−0.88	1.78
Modern European languages	−0.17	−2.63	2.29	1.27	−0.25	2.80
Humanities	−2.54	−4.47	−0.60	−0.18	−1.65	1.30
Creative arts	−4.12	−6.13	−2.12	−1.70	−3.20	−0.19

(*continued*)

Table 2 (*continued*)

Variable‡	*Marginal effects (%) for males*			*Marginal effects (%) for females*		
		95% confidence interval			*95% confidence interval*	
		Lower	*Upper*		*Lower*	*Upper*
Education	6.15	4.60	7.69	5.75	4.79	6.72
Other subjects	−0.89	−2.51	0.72	−1.41	−2.76	−0.05
Study method (non-sandwich)						
Sandwich	3.50	2.75	4.24	1.69	0.87	2.51
Accommodation type (own accommodation)						
Institution	0.21	−0.59	1.02	1.31	0.60	2.02
Parent	−0.27	−1.14	0.59	0.15	−0.63	0.92
Other	−0.20	−1.12	0.72	−0.11	−0.93	0.71
Not known	−0.18	−1.22	0.86	−0.07	−0.98	0.85
Degree class (upper second)						
First	3.80	0.10	7.51	0.76	−2.58	4.11
Lower second	−1.84	−3.78	0.10	−0.72	−2.44	1.01
Third	−2.81	−5.95	0.32	0.71	−2.07	3.48
Other	0.65	−2.00	3.30	2.20	−0.12	4.51

†The model also includes controls for a type of prior qualification other than A-levels, region of residence, university attended and interactions between degree class and social class, disability, ethnicity and type of course.
‡The default category for each group of categorical variables is noted in parentheses.

grade) raise the probability of responding by 0.08 and 0.14 percentage points. Mature students (those aged 24 years or more on leaving) are less likely to respond to the FDS, regardless of sex.

There are strong degree subject effects and also a strong and monotonic association between degree class and the probability of responding to the FDS. Graduates who perform better in their degrees have a significantly higher probability of responding to the FDS: for the default student (white, studying business, with no disability and a social class II background), a first-class degree is associated with being approximately 10 percentage points more likely to respond to the FDS than a student with a third-class degree, *ceteris paribus*. From the inclusion of various interaction dummy variables, we find that the effect which is associated with a first-class degree carries over for students regardless of ethnicity, subject, disability status and social class background. As we know that the probability of EFS is positively related to degree class (see Smith *et al.* (2000)), this finding supports the argument that those who do not respond are likely to have poor labour market prospects.

University effects are also important. For males, 38 out of 101 institutions have a significant and positive marginal effect, whereas 28 have a significant negative effect, relative to the default. For females, the corresponding numbers are 28 significantly positive and 30 significantly negative. We observe, however, that there is considerable overlap in the confidence intervals for the university marginal effects and hence that there is little precision in a 'league table' of universities that is based on these estimated effects. We note, for example, that the confidence interval for the 49th and 36th ranked university for males and females respectively is overlapping with all those ranked as high as 14th and 6th and as low as 89th and 74th.

We also note that there is little correlation (−0.03 for males and −0.12 for females) between the university marginal effects on responding (R) to the FDS and the university marginal effects on obtaining a qualification (OQ). In contrast, the correlation in the university marginal effects on the response rates for males and females is very high at 0.95. Again, we would infer from

these results that ignoring the issue of non-random response is likely to lead to potential bias in the construction of institutional PIs that are based on conditional employment outcomes.

3.3. Modelling the first-destination outcomes (employment or further study)

As is shown in Table 1, of the 69 767 male graduates who responded to the FDS, 60 787 (approximately 87%) were in either employment or further study (EFS). Of the 77 263 females who responded to the FDS, 68 601 (approximately 89%) were in either employment or further study. From Table 1, we can also see that, compared with the population of those responding (R) to the survey, those in employment or further study (EFS) tended on average

(a) to have higher A-level scores,
(b) to have a higher class of degree and
(c) to have taken a sandwich course.

Table 2 shows the estimated marginal effects on the EFS probability, along with the 95% confidence intervals. With respect to the graduate's educational background, Table 2 shows that the individual's total A-level points score has no significant effect on the post-university EFS probability, for either males or females. With respect to personal characteristics, Table 2 shows the effects of age at graduation on the EFS probability, with more mature students more likely to be unemployed or inactive than students who are aged less than 24 years at graduation.

As with the analysis of the probability of responding, interactions between degree class and subject, ethnicity, disability and social class were included in the estimations. There are few significant social class effects. Graduates' disability status has some significant effect on the EFS probability of male graduates. There are strong ethnicity effects. For the default case, black, Indian and Chinese graduates have lower EFS probabilities than otherwise equivalent white graduates.

Table 2 shows that significant marginal effects are associated with the subject that is studied at university, compared with the default case of students studying for business studies degrees. The ranking of degree course effects is similar to that reported in Smith *et al.* (2000) for 1993 leavers. Table 2 also shows the estimated effects of the student's degree class on the EFS probability. For the default case of a white student of business with no disability and from a social class II background, a male graduate with a first-class degree is more likely to enter further study or employment than an otherwise equivalent student with a lower class degree.

Fig. 1 plots the institutional marginal effects on the probability of being employed or in further study, conditional on responding to the FDS (for brevity, figures are included for male students only). There are only limited statistically significant effects: for males and females, 15 and three institutions respectively have positive and significant effects and 13 and 12 institutions have negative and significant effects, relative to the default case. Furthermore, we observe that there is considerable overlap in the confidence intervals for the university marginal effects and hence that there is little precision in a league table of universities based on these estimated effects. We note for example that the confidence interval for the 12th and 43rd ranked university for males and females respectively is overlapping with all those ranked as high as third and second and as low as 78th and 101st. Thus, for females in particular, there is a remarkable degree of overlap in the confidence intervals. We note that the point estimates themselves are large. For example, the point estimate for the top ranked and lowest ranked university for males indicates respectively a probability of EFS of 10 percentage points higher and lower than that for the median university.

The correlation between the university marginal effect on the probability of being in the EFS category, conditional on response, and that on the probability of responding, conditional on

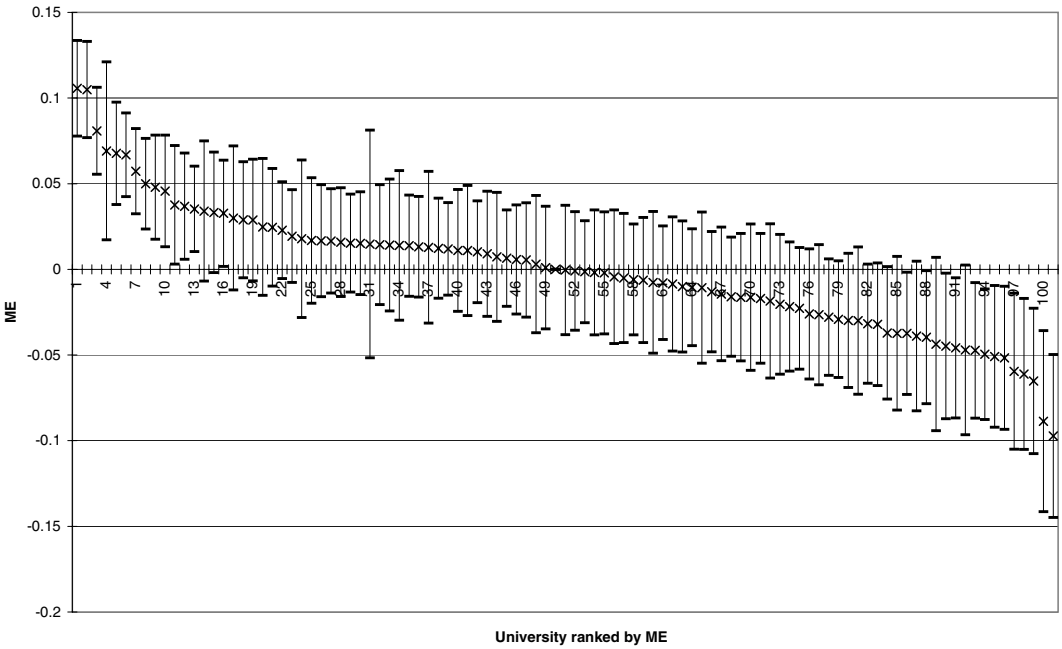

Fig. 1. University marginal effects and 95% confidence intervals for EFS conditional on *R* and OQ for males

qualification, is 0.29 for males and 0.27 for females. The correlation across males and females in the university marginal effect on the probability of being in the EFS category is 0.67. The correlation between the university marginal effect on the probability of being in the EFS category, conditional on response, and that on the probability of qualification is 0.06 for males and 0.01 for females.

4. University marginal effects: independent probits

In the preceding section, we considered how individual level characteristics influence various outcome variables, under the maintained hypothesis of independence between the processes determining the outcomes. The first two outcomes—the probability of obtaining a qualification (OQ) and the probability of responding (*R*) to the FDS—determine the prerequisites for inclusion in the official university or HEI employment PIs. We have found that a range of important individual level personal characteristics (age at graduation, social class background, entry level qualifications) and course characteristics (subject of degree) along with degree performance were associated with differences in these probabilities. In addition, even after controlling for all these factors, we find statistically significant 'unexplained' differences across HEIs. This suggests that differences across universities either in qualification completion rates or in survey response rates could contaminate the final measures of HEI employment performance.

In terms of published university PIs, we note that there are PIs for student progression rates by institution. Thus, factoring qualification rates into a PI for employment outcomes could be regarded as double-counting qualifications. In view of this, in this section we restrict ourselves to an assessment of the sensitivity of HEI rankings to whether or not we condition on the probability of responding to the FDS. Specifically, we compare HEI rank positions obtained from modelling

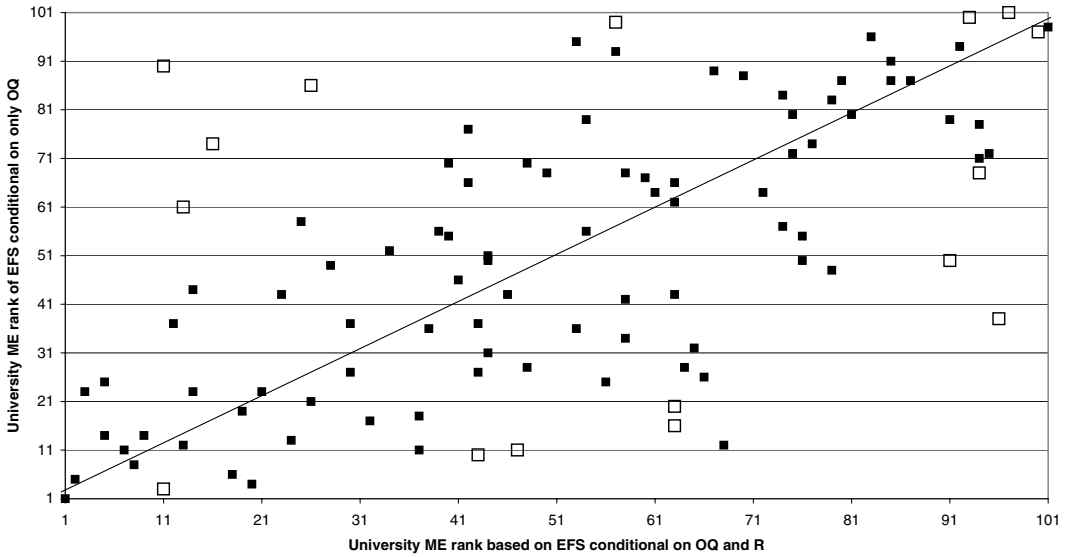

Fig. 2. Comparison of university marginal effects rankings by EFS conditional only on OQ *versus* conditional on OQ and *R* for males: □, university's 95% confidence intervals (for its two rank positions) do not overlap

(a) the probability of EFS conditional on having responded to the FDS—similar to the official HEFCE approach—and
(b) the probability of EFS conditional only on having obtained a qualification—i.e. factoring in the probability of responding, which varies by institution.

Fig. 2 shows a plot of the ranked university marginal effects on these two probabilities, for male students. Following the approach of Marshall and Spiegelhalter (1998), the rankings are derived as the median ranking (based on the university marginal effects), from 500 bootstrap replications of the two probit models, where the errors are generated as independent normal random variates. For each of the two probabilities, the institution with the highest level of performance is denoted as having the top ranked (1) marginal effect.

If the probability of responding to the FDS were a random event, then on average HEIs would be expected to have the same response rates and hence to line up along a 45° line from the origin. Instead, HEIs occupy different rank positions across the two alternative measures of performance, i.e. according to whether or not we factor in the differences in the probability of responding to the FDS across universities. Universities with low response rates do better under the measure which is constructed conditional on response—in Fig. 2, they appear above the 45° line. When we adjust for differential response, universities with low response rates are penalized as non-response is effectively treated as a bad outcome, equivalent to being unemployed or inactive. This is likely to overpenalize low response as, although non-respondents are more likely than respondents to be unemployed or inactive, some non-zero proportion is likely to be employed or in further study.

We calculate the rank correlation between the two sets of university ranks that are shown in Fig. 2 at 0.63 for males and 0.66 for females. Thus, there is substantial sensitivity in the measure of university performance ranking according to whether differential response is factored in or not. On the basis of the 500 bootstrap replications we derived 95% confidence intervals for the ranks for both males and females. We find that there are only 16 universities for which the

university's 95% confidence interval around its ranked position conditioning only on OQ fails to overlap with its corresponding 95% confidence interval derived conditional on both OQ and R. These universities are indicated by an open square in Fig. 2 (for males). The fact that there are so few stems from the generally low level of precision in the estimated university rank positions. It is this poor precision which underlies the uncertainty in the ranking of universities.

5. Models with selection

So far, we have assumed independence across the error terms in the equations for qualifying, responding and obtaining particular labour market outcomes. In this section of the paper, we relax this assumption and allow for unobservable individual factors affecting first-destination outcomes to be correlated with those driving non-response to the FDS. One possible reason for a correlation, for example, arises when students who have derived much satisfaction from their degree choose to stay on for further study. Their appreciation of their degree is also likely to make them more likely to complete the FDS.

Given that separate PIs are published for progression, as we noted in the previous section, our subsequent analysis conditions on the student's having obtained a qualification. We estimate equations for the probabilities of responding and of being EFS (employed or in further study), allowing for correlation in the error processes for the two probabilities, i.e. allowing for $\text{corr}(u_{2i}, u_{3i}) = \rho \neq 0$.

Following the approach of Van de Ven and Van Praag (1981), we model, for those individuals who are in the target population (i.e. have obtained a qualification: $y_{1i} = 1$), the probability that the student is either employed or in further study (EFS), rather than unemployed or inactive (UI), based on the selection equation for the probability that an individual responds (R) to the FDS.

The log-likelihood is

$$\begin{aligned}\ln(L) = &\sum_{\substack{y_{2i}=1\\y_{3i}=1}} \ln\left\{\Phi_2(x'_{2i}\beta_2 + \sum_j H_{ij}\alpha_{2j}, x'_{3i}\beta_2 + \sum_j H_{ij}\alpha_{3j}, \rho)\right\}\\ &+ \sum_{\substack{y_{2i}=1\\y_{3i}=0}} \ln\left[\Phi_2\{x'_{2i}\beta_2 + \sum_j H_{ij}\alpha_{2j}, -(x'_{3i}\beta_3 + \sum_j H_{ij}\alpha_{3j}\}, -\rho)\right]\\ &+ \sum_{y_{2i}=0} \ln\left\{1 - \Phi(x'_{2i}\beta_2 + \sum_j H_{ij}\alpha_{2j})\right\}\end{aligned}$$

where $\Phi_2(\cdot)$ is the cumulative bivariate normal distribution function (with mean $[0, 0]'$) and $\Phi(\cdot)$ is the cumulative standard normal distribution. The set of accommodation variables is used to identify the EFS equation as the type of university accommodation is unlikely to have a direct influence on the post-university labour market outcome but is a factor in determining the probability of response to the FDS. The results for the model with selection are very similar to those for the model without selection, especially for male students. This is not surprising as the selection term ρ is not significant in the equation for males. For females, ρ is significantly different from 0, and thus ignoring this would introduce bias into the estimated employment probabilities.

With respect to the university marginal effects on the probability of EFS, conditional on response, derived from the probit model with selection, for males and females 11 and 4 respectively out of 101 institutions have a significant and positive marginal effect relative to the default case, whereas 11 and 8 respectively have a significant negative effect. Again, there is considerable

Table 3. Marginal effects on the probability of EFS (probit model with selection)—males and females†

Variable‡	*Marginal effects for males*			*Marginal effects for females*		
		95% confidence interval			*95% confidence interval*	
		Lower	*Upper*		*Lower*	*Upper*
Prior qualifications (A-levels)						
A-level points score	0.16	−3.47	3.78	1.76	−1.08	4.60
Age groupings (<24) (years)						
24–26	−0.94	−2.33	0.44	−1.70	−2.64	−0.76
27–30	1.38	−0.69	3.44	−3.60	−5.63	−1.56
⩾31	−2.27	−4.58	0.04	−6.60	−8.52	−4.68
Social class (II)						
I	0.85	−0.53	2.24	0.72	−0.36	1.79
IIINM	1.36	0.02	2.69	0.36	−0.72	1.44
IIIM	0.53	−0.78	1.85	0.21	−0.83	1.25
IV	0.35	−1.21	1.92	1.49	0.27	2.71
V	1.26	−1.29	3.81	0.59	−1.66	2.84
Armed forces	0.94	−2.32	4.19	2.89	0.09	5.69
Not known	0.34	−0.78	1.46	−0.38	−1.27	0.51
Retired	−3.13	−10.85	4.59	−0.03	−6.49	6.42
Unemployed	−1.77	−6.14	2.59	0.38	−2.62	3.38
Disability (no disability)						
Dyslexia	−0.70	−3.58	2.19	−1.66	−4.51	1.18
Blind	−3.26	−10.06	3.54	2.11	−3.72	7.94
Deaf	−6.93	−13.32	−0.53	−2.91	−7.68	1.85
Wheel-chair	−7.86	−15.89	0.18	3.55	−1.33	8.43
Mental	−16.18	−30.14	−2.22	−4.54	−13.55	4.47
Diabetes	−0.62	−2.91	1.68	1.08	−0.58	2.73
Multiple	−7.64	−15.67	0.39	−5.92	−13.48	1.64
Other	−3.59	−8.14	0.97	−0.44	−4.06	3.17
Not known	−2.69	−4.69	−0.69	−0.80	−2.24	0.64
Ethnicity (white)						
Black	−3.79	−7.14	−0.44	−1.99	−3.88	−0.10
Indian	−3.22	−5.57	−0.86	−5.34	−7.04	−3.63
Chinese	−3.63	−6.57	−0.68	−4.19	−6.35	−2.03
Other	−1.65	−4.61	1.32	−0.73	−2.83	1.37
Not known	−0.56	−2.06	0.94	−1.59	−2.79	−0.39
Course type (business)						
Medical related	7.44	5.31	9.57	6.57	5.62	7.52
Biological science	−0.18	−1.82	1.45	−1.39	−2.74	−0.05
Agriculture and related	−0.63	−3.45	2.19	2.00	−0.11	4.11
Physical science	0.81	−0.80	2.41	0.00	−1.49	1.48
Mathematical science	4.34	2.78	5.89	0.54	−1.04	2.12
Engineering	3.64	2.03	5.25	0.57	−1.35	2.49
Technology	1.47	−1.08	4.02	2.97	0.35	5.58
Architecture and building	4.47	2.64	6.31	2.26	0.17	4.36
Social science	−2.58	−4.67	−0.49	−0.97	−2.35	0.40
Economics	−0.46	−2.53	1.62	0.41	−1.95	2.77
Politics	−2.40	−4.83	0.04	−1.52	−3.60	0.56
Law	3.99	2.03	5.94	4.22	3.01	5.43
Communication	0.02	−2.60	2.65	−2.69	−4.93	−0.46
Literature and classics	−1.46	−3.58	0.65	0.60	−0.71	1.91
Modern European languages	−0.03	−2.51	2.46	1.61	0.11	3.10
Humanities	−2.56	−4.58	−0.55	−0.09	−1.54	1.36
Creative arts	−4.15	−6.26	−2.03	−1.62	−3.08	−0.15

(*continued*)

Table 3 (*continued*)

Variable‡	*Marginal effects for males*			*Marginal effects for females*		
		95% confidence interval			*95% confidence interval*	
		Lower	*Upper*		*Lower*	*Upper*
Education	6.48	4.15	8.81	6.12	5.09	7.16
Other subjects	−0.86	−2.50	0.77	−1.31	−2.64	0.01
Study method (non-sandwich)						
Sandwich	3.79	2.77	4.82	2.04	1.23	2.85
Accommodation type (own accommodation)§						
Institution	0.10	0.05	0.15	0.13	0.07	0.19
Parent	0.27	0.20	0.34	0.31	0.25	0.37
Other	−0.01	−0.06	0.05	−0.10	−0.17	−0.04
Not known	−0.02	−0.08	0.04	−0.18	−0.25	−0.10
Degree class (upper second)						
First	4.07	0.15	7.99	1.10	−2.22	4.42
Lower second	−1.95	−4.08	0.18	−0.80	−2.51	0.90
Third	−3.05	−6.42	0.32	0.44	−2.32	3.20
Other	0.51	−2.70	3.72	2.03	−0.30	4.35
$\hat{\rho}$	−0.26	−0.67	0.28	−0.61	−0.83	−0.24

†The model also includes controls for a type of prior qualification other than A-levels, region of residence, university attended and interactions between degree class and social class, disability, ethnicity and type of course.
‡The default category for each group of categorical variables is noted in parentheses.
§Accommodation type is only included indirectly through the responding to the FDS equation.

overlap in the confidence intervals for the university marginal effects. We note for example that the confidence interval for the 11th and 36th ranked university for males and females respectively is overlapping with all those ranked as high as 2nd and 1st and as low as 78th and 100th respectively. Thus, just as in the case of the model without selection, presented in Section 3.3, there is very little precision in the ranking of universities by their marginal effects on the EFS outcomes of students who respond to the FDS.

The correlation across male and female rankings of universities is 0.66 for the probit model with selection. For male students, the correlation between the rankings of the university marginal effects according to

(a) the model with selection and
(b) the model without selection

is 0.998, reflecting the fact that ρ is not significant for males in the probit model with selection. For female students the equivalent correlation is 0.978. Thus, although for females ρ is significant, correcting for selection does not lead to substantial changes in the university ranking.

5.1. Predicted probabilities for the positive outcomes for non-respondents

We do not observe the first-destination outcomes of non-respondents, of course. However, we do observe non-respondents' characteristics from the HESA administrative database and hence we can derive predicted values for the status of non-respondents' first-destination outcomes. We do this on the basis of the probit model with selection. We then derive university marginal effects on the EFS probability, together with 95% confidence intervals, defined over both respondents

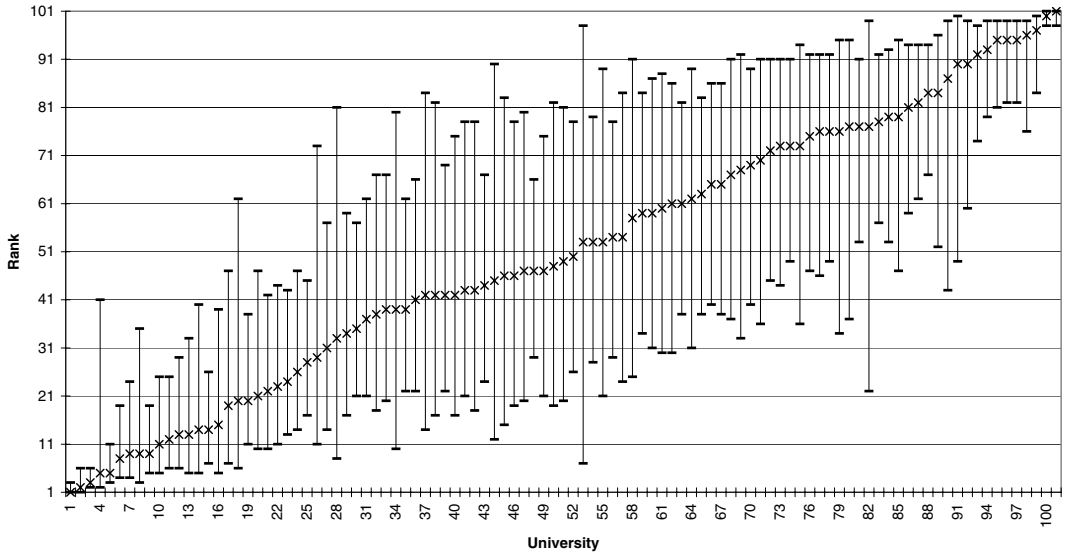

Fig. 3. University ranks and 95% confidence intervals based on the probability of EFS for both responders and non-responders for males

and non-respondents, obtained by bootstrapping the probit model with selection 500 times, drawing from a bivariate normal distribution using the ρ-estimate that is reported in Table 3 for males and females. For males and females, 13 and 11 out of 101 institutions respectively have a significant and positive marginal effect relative to the default case, whereas 7 and 10 have a significant negative effect. Once more we observe considerable overlap in the confidence intervals for the university marginal effects. We note for example that the confidence intervals for the 20th and 12th ranked university for males and females respectively are overlapping with all those which are ranked as high as sixth and third and as low as 62nd and 89th respectively.

Again following Marshall and Spiegelhalter (1998), we calculate the median and the 95% confidence interval for the rank position of each university, based on 500 bootstrap replications. These are represented in Fig. 3, which shows the considerable uncertainty in the universities' rank positions. Although a small number of universities at the extremes of the ranking have small confidence intervals, most other universities have very wide overlapping confidence intervals, indicating considerable uncertainty in the rankings derived. Fig. 4 plots the university ranking based on the university marginal effects for both respondents and non-respondents against the ranking based on the marginal effects on the probability of EFS conditional on response *for respondents only*. The correlation is 0.96 in the case of males and 0.94 for females. The extent of uncertainty in the rankings of university marginal effects that is demonstrated in Fig. 3 explains in part why in Fig. 4 there is no university for which the confidence intervals of the ranks over the two criteria do not overlap.

There is, then, further evidence of uncertainty in the rankings of universities by their marginal effects. As we would expect, the correlation in Fig. 4 is much higher than that in Fig. 2. This is because, in Fig. 4, non-respondents are attributed a predicted value for EFS, based on the probit model with selection. In Fig. 2, instead, non-response is treated as a negative outcome in the same way as is unemployment or inactivity. This is a rather extreme assumption and generates a substantial difference in university rankings according to whether or not the rankings are

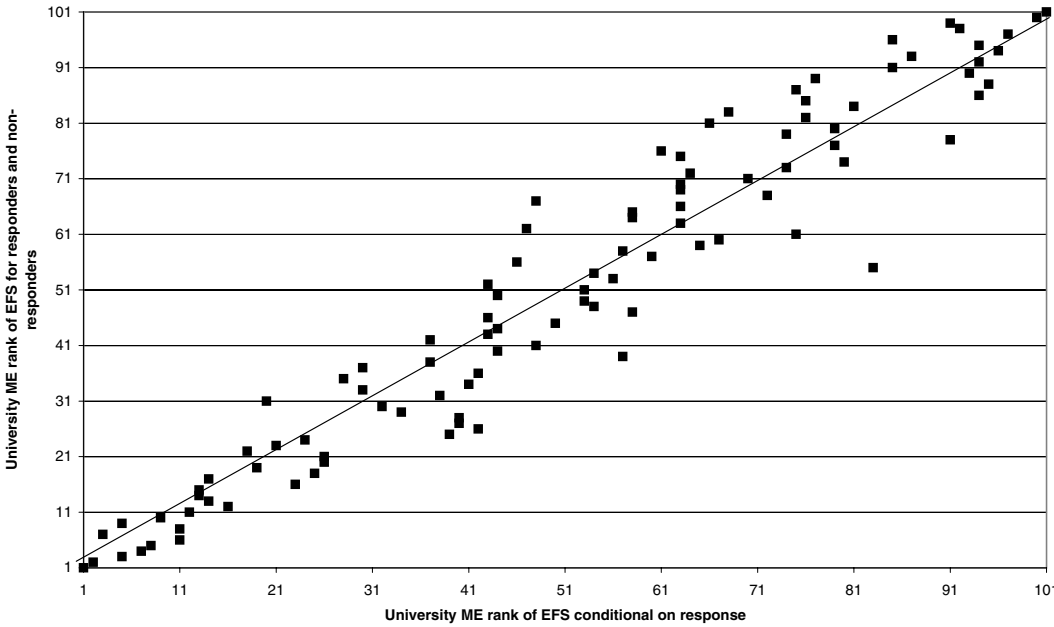

Fig. 4. Comparison of university marginal effects rankings by EFS conditional on reponse *versus* EFS for both responders and non-responders for males

calculated as conditional on response. We note that ignoring the potential problem of selection on unobservable characteristics does not introduce variations in the university marginal effects derived, which are as substantial as the variations that are associated with differences in the treatment of non-respondents. The fact that there is only limited selection on the unobservables is not inconsistent with substantial selection on observables: indeed, the results suggest that non-respondents have observed characteristics which predict a much lower EFS probability in the probit model with selection compared with respondents.

6. Sensitivity issues

In this section, we look at the sensitivity of our results to

(a) the type of university,
(b) missing information on social class background of students and
(c) the treatment of economic activity.

6.1. Type of university

In the analysis that is presented in this paper, for each of our (gender-specific) specifications we have included dummy variables for each university. Our models do not allow for interactions between universities and other co-determinants, with the exception of gender. The most likely source of differential performance across universities stems from the distinction between the 'traditional' or 'old' universities (those which predate the abolition of the binary divide between universities and polytechnics) and the 'new' universities (i.e. the former polytechnics), typically with objectives that are different from those of the old universities. Consequently, we have estimated the probit model of EFS, correcting for selection, on the population of students at old and new universities separately. The estimated marginal effects (which are not reported) are very

similar to those reported in Table 3. Additionally, we find that the rankings of universities is very robust across the two approaches. The ranking of old universities based only on the sample of old universities has a correlation of 0.99 and 0.97 with the equivalent ranking of old universities for males and females respectively. For new universities, the equivalent correlations are 0.98 for both males and females.

6.2. Missing social class information

The administrative HESA data that were exploited in our analysis provide us with a rich database with little missing information for most variables. The summary statistics in Table 1 show the proportion of cases for which information on key variables is missing. One variable with a high proportion of missing observations relates to parental social class background, with unknown values for over 40% of cases. In the analysis that has been presented so far, we have included all individuals in the analysis and used dummy variables for unknown values. We have also re-estimated the EFS probability by using a probit model correcting for selection and excluding all those individuals for whom there is missing information on social class background. The estimated effects that are associated with the various determinants of EFS are robust to this re-estimation. With respect to the estimated university marginal effects, there is a high correlation across the two methods. However, at several new universities information on social class is particularly poor such that the institutional effect either cannot be estimated or is associated with very large standard errors.

6.3. Treatment of economic inactivity

The official HEFCE method for deriving university performance indicators based on employment excludes graduates who report themselves as unavailable for work. In our analysis, we have included this group with the unemployed. It is interesting to examine the sensitivity of the derived ranking of universities according to the treatment of those who are unavailable for work, whom we have labelled inactive (I) in our analysis. We have re-estimated the Heckit-type model for EFS, excluding those who are inactive and hence distinguishing between those in employment or further study (EFS) and those unemployed (U). The correlation in the derived rankings for universities across the two methods is 0.81 for males and 0.69 for females. It is not surprising that the ranking for females is the more sensitive to the treatment of inactivity: of unemployed or inactive female students, 48.4% are inactive whereas for males 36.4% are inactive.

7. Concluding remarks

This paper is concerned chiefly with examining the robustness of rankings of university marginal effects on graduate employment status to variations in the method for deriving those marginal effects. We note that the official HEFCE methodology is based only on students who responded to the FDS survey. Our main concern is with the fact that students who did not respond to the FDS survey are unlikely to be randomly assigned across either universities or employment outcomes and hence ignoring them risks introducing bias into the estimation of university effects.

Using HESA data for all UK university leavers for 1997–1998 we model the probability that a graduate is in employment or further study (EFS) rather than unemployed or inactive (UI) after university, conditional on responding to the FDS survey. We derive university marginal effects on this probability and rank universities on the basis of the estimated marginal effects. We find that there are considerable overlaps in the confidence intervals around the point

estimates, leading to little confidence in such a ranking. We also compare a HEFCE-type ranking of university marginal effects on the probability of EFS, conditional on response, with a ranking which exploits a probit model of the probability of responding to adjust for differences in response. We find that the correlation between these two ranks is 0.63 for males and 0.66 for females, suggesting that the treatment of non-response is an important influence on how universities are ranked.

Using a probit model with selection, we also derive predicted probabilities of EFS for non-respondents and derive university marginal effects for both respondents and non-respondents. Ranking the universities by these marginal effects we find that the 20th most highly ranked university for males has a confidence interval which overlaps with the confidence intervals of all those universities that are ranked as high as sixth and as low as 62nd, out of the 101 universities. Thus, we find substantial uncertainty in the university ranking derived. We also plot the university ranking based on these marginal effects (i.e. those for both respondents and non-respondents) for EFS against the equivalent ranking based on *respondents only*: but owing to the uncertainty we find that for no university is there any significant difference in the two rankings.

We conclude from these and from related results that

(a) we can have little confidence in the ranking of HEIs by their effects on the probability of graduates entering employment or further study rather than unemployment or inactivity,
(b) the treatment of non-response could potentially have an important influence on how universities are ranked by their effects on the EFS probability of graduates and
(c) there is considerable sensitivity of university rankings to the inclusion of those who are inactive, especially in the case of females.

It follows that published PIs for universities in the area of graduate labour market outcomes should be interpreted with due caution. In particular, our evidence suggests that it would be misleading to publish league tables of universities as we can have little statistical confidence in the rankings. We have also indicated that there are potential disincentives for universities to seek a high response rate to the FDS, given that the current employment PI does not adjust for poor response. Finally, like the official employment PI, our analysis is based on employment data that were gathered in the first-destination survey of graduates. Much more valuable would be data pertaining to graduates' employment status 3 years or more after graduation.

Acknowledgements

We acknowledge the support of the HESA in providing us with the data and the Department for Education and Skills for financial support. Neither of these organizations bears any responsibility for the analysis or interpretations that are presented in this paper. We would like to thank the Joint Editor, two referees, Mark Stewart, Jim Taylor and other participants at the Royal Statistical Society meeting on 'Performance monitoring and surveillance', January 2003.

References

Bratti, M., McKnight, A., Naylor, R. and Smith, J. (2003) Higher education outcomes, graduate employment and university performance indicators. *Warwick Economics Research Paper 692*. University of Warwick, Coventry.
Cave, M., Hannay, S. and Kogan, M. (1997) *The Use of Performance Indicators in Higher Education: the Challenge of the Quality Movement*. London: Kingsley.
Cragg, J. G. (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, **39**, 829–844.
Dearing, R. (1997) *Higher Education in the Learning Society*. London: Stationery Office.

Department for Education and Employment (1995) *GCSE to GCE A/AS Value Added: Briefing for Schools and Colleges*. London: Department for Education and Employment.
Department for Education and Skills (2003) *Secondary School Performance Tables 2002: Value Added Technical Information*. London: Department for Education and Skills. (Available from `http://www.dfes.gov.uk/performancetables/schools_02/sec3b.shtml`.)
Goedegebuure, L. C. J., Maassen, P. A. M. and Westergeijden, D. F. (1990) *Peer Review and Performance Indicators: Quality Assessment in British and Dutch Higher Education*.
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.
Heckman, J. J. (1979) Sample selection bias as a specification error. *Econometrica*, **47**, 153–161.
Higher Education Funding Council for England (2002) *Performance Indicators in Higher Education 2002*. Higher Education Funding Council for England. (Available from `http://www.hefce.ac.uk/Learning/perfind/2002/report.asp`.)
Johnes, G. and McNabb, R. (2003) Never give up on the good times: student attrition in the UK. *Oxf. Bull. Econ. Statist.*, to be published.
Johnes, J. and Taylor, J. (1990) *Performance Indicators in Higher Education*. Oxford: Oxford University Press.
Johnes, J., Taylor, J. and Francis, B. (1993) The research performance of UK universities: a statistical analysis of the results of the 1989 research selectivity exercise. *J. R. Statist. Soc.* A, **156**, 271–286.
Manning, W. G., Duan, N. and Rogers, W. H. (1987) Monte Carlo evidence on the choice between sample selection and two-part models. *J. Econometr.*, **35**, 59–82.
Marshall, E. C. and Spiegelhalter, D. J. (1998) Reliability of league tables in in vitro fertilisation clinics: retrospective analysis of live birth rates. *Br. Med. J.*, **316**, 1701–1704.
McNabb, R., Pal, S. and Sloane, P. (2002) Gender differences in educational attainment: the case of university students in England and Wales. *Economica*, **69**, 481–503.
Puhani, P. A. (2000) The Heckman correction for sample selection and its critique. *J. Econ. Surv.*, **14**, 53–67.
Smith, J., McKnight, A. and Naylor, R. A. (2000) Graduate employability: policy and performance in higher education in the UK. *Econ. J.*, **110**, F382–F411.
Smith, J. P. and Naylor, R. A. (2001a) Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *J. R. Statist. Soc.* A, **164**, 389–405.
Smith, J. and Naylor, R. A. (2001b) Determinants of individual degree performance: evidence for the 1993 UK university graduate population from the USR. *Oxf. Bull. Econ. Statist.*, **63**, 29–60.
Van de Ven, W. P. M. M. and Van Praag, B. M. S. (1981) The demand for deductibles in private health insurance: a probit model with sample selection. *J. Econometr.*, **17**, 229–252.