PSI Research Discussion Paper 5

The Impact of Careers Guidance for Employed Adults in Continuing Education

PSI RESEARCH DISCUSSION PAPER 5

The Impact of Careers Guidance for Employed Adults in Continuing Education
Assessing the Importance of Attitudinal Information

Michael White and John Killeen

psi

Policy Studies Institute

ISBN 0 85374 783 0

PSI Report No 875

Policy Studies Institute
For further information contact
Publications Dept, PSI, 100 Park Village East, London NW1 3SR
Tel (020) 7468 0468 Fax (020) 7468 2211 Email pubs@psi.org.uk

PSI is a wholly owned subsidiary of the University of Westminster

Abstract

The validity of the matching estimator in programme evaluation depends on the completeness of the set of variables used for matching. When an attitudinal variable is relevant for the participation decision, but is either unmeasured or measured only after entry to the programme, estimates of effects may be biased or hard to interpret. This issue is investigated with data from an evaluation study of careers guidance for employed adults, which utilised the method of propensity score matching. Job satisfaction, measured shortly after entry to the programme, was found to be strongly associated with participation, but may itself have been influenced by the early experience of careers guidance. Estimates of the impacts of guidance on several post-programme education and training outcomes are considered, both including and excluding the job satisfaction measure from the participation model. Data experiments with adjusted values of job satisfaction are also performed. It is found that estimates of treatment effects are highly sensitive to these variants, and respond in a non-monotonic fashion. The implications for evaluation methodology are discussed.

Introduction

There is currently considerable interest in ‘matching on observables’ as an econometric evaluation method[1] in non-experimental studies. The assumptions required for causal inference using non-experimental matching were defined by Rubin (1974; 1979). The conditional independence assumption (CIA) states that the effect of a non-experimental treatment can be recovered given that the outcome is independent of individuals’ assignment to the programme, conditional on values of a set of variables X. Rosenbaum and Rubin (1983) further showed that, with an additional assumption[2] about the availability of observations on participants and non-participants across values of X, an equivalent result could be obtained for a function b(X), the balancing score, such that the conditional distribution of X given b(X) is the same for treated and untreated samples. The most useful interpretation of the balancing score is, in many cases, the propensity to receive the treatment, P(X). The practical difficulty of matching on covariates when X has many members is resolved by using X to estimate the propensity score P(X) (typically by means of a logit or probit model). Once propensity scores have been obtained, a range of methods is available for matching non-participant to participant samples. Matching methods have been reviewed by Heckman, LaLonde and Smith (1999), who point out that they can be regarded as forms of sample weighting.

A distinguishing feature of the matching estimator is its requirement that X be observable and that it contain all relevant variables. An obvious diffi-

---

1 See for example: Dehejia and Wahba (1998); Frohlich, Heshmati and Lechner (2000); Lechner (1999; 2000); Brodaty, Crepon and Fougere (2000).

2 Rosenbaum and Rubin (1983) state as their central definition that ‘assignment is strongly ignorable given a vector of covariates v if (r_1,r_0) ^ z | v, and 0 < pr(z=1|v) < 1 for all v’, where r_1, r_0 are the outcomes under participation and non-participation respectively, ^ means ‘is independent of’, z indicates assignment to participation and takes values 1 and 0, and v is the population counterpart of the observed covariates. See also Heckman, Ichimura and Todd (1998) for weaker assumptions on which estimates of particular treatment parameters can be based.

culty is that some kinds of variable, notably those of an attitudinal or motivational type, may be relevant to both participation and outcomes yet hard to measure. This arises, in essence, because attitudes prior to participation cannot be retrospectively calculated nor can they be recalled. While many well-validated survey questions to measure attitudes have been developed by psychologists, it is usually only possible to administer these after entry to the programme, when the attitudes in question may already have been influenced by assignment or participation. As noted by Heckman, LaLonde and Smith (1999), conditioning on X variables which is caused by assignment or participation undermines the interpretability of the estimated effects.

This paper examines a situation in which an attitudinal variable is plausibly the most important single influence on entry to a programme, but can only be measured post-entry. One approach in this situation is to investigate the sensitivity of estimates when the variable is included in or omitted from the set of variables X on which matching is conditioned (via P(X)). Another possibility is to conduct data experiments in which the attitudinal variable is adjusted across a range of values, and the response of the estimates to these variants is observed.

Data

The study, a more complete account of which is given in Killeen and White (2000),[3] concerns the effects of careers guidance services for adult employed people. The fieldwork took place in 1997 and 1998. Eight localities[4] with public services of this type were studied. About two months after registering for careers guidance, participants were recruited[5] to the study either by postal questionnaire or by a short telephone interview. A parallel sample of employed people was recruited by telephone interview, using a random digit dialling method within the same localities where participants were found. The initial samples were followed up about 12 months later, and a face-to-face interview conducted. There was substantial sample attrition, with 65 per cent of the participants and 59 per cent of the controls being interviewed at follow-up. This partly resulted from a youthful sample with a high rate of

3 A follow-up took place in 1999 but is not considered here. Fieldwork was conducted by Public Attitude Surveys Ltd.

4 Each locality was the administrative area of a Training and Enterprise Council, typically with an employed population of 300,000. It is believed that public careers guidance services were offered to adult employed people in 10 such localities at this time.

5 A recruited person was one who completed the questionnaire or interview, and also agreed to subsequent follow-up.

mobility; the study was unable to follow up those who had left the locality. The participation models to be described later were weighted by sex, age group, full-time employment status, and locality group, to correct for bias introduced by attrition. Sources of bias at the stage of initial recruitment to the study are unknown and we perforce assume that they were of the same type as corrected by the attrition weights. After elimination of some cases with missing data, there were 669 participants in guidance, and 1015 non-participants, available for the modeling analyses to be described here.

The nature of careers guidance services was broadly similar across the eight localities covered, with one or more counselling interviews supplemented by careers library information, self-assessment exercises (usually computer-aided), assistance in CV writing, and referral to other services (such as job placement services or colleges) in some cases. Use of the services generally extended over several months.

The outcome variables which are the focus of this paper concern continuing education and training (CET). The majority of participants in guidance gave interest in CET as a reason for seeking guidance (54 per cent sought help with education and training opportunities, and 56 per cent wanted to improve their skills or qualifications), while a somewhat larger proportion stated that CET had been suggested to them in their counselling interviews. Guidance may be conceptualised as (in part) a means of reducing the costs of search for CET opportunities and of increasing the quality of CET decisions. It is plausible therefore that participation in guidance will increase CET outcomes. In this paper, we focus on three outcomes which were measured at the follow-up interview: participation during 1997 or 1998 in a CET course not paid for or provided by an employer (SELFCET); whether or not the individual had a qualification aim in such a course (SELFQAIM); and whether or not a qualification had been obtained from such a course (SELFQUAL).[6] Other CET outcomes, including those achieved by the time of a later follow-up, are examined in Killeen and White (2000).[7] Descriptive statistics for the selected outcomes are shown in Table 1. The proportions with each outcome were around two to three times greater in the participant group than in the non-participant group.

The information collected at recruitment of the sample members was designed to discriminate between guidance participants and non-participants.

6 SELFQAIM is nested within SELFCET, and SELFQUAL is nested within SELFQAIM. In this paper we ignore the additional information potentially available in these conditional relationships.

7 The general result from the wider analyses was that CET activities and qualifications continued to increase over time for the guidance participants relative to the non-participants.

Table 1: *Summary statistics for continuing education and training (CET) outcome variables*

Label	Description	Participants			Non-participants		
		N	Mean	SD	N	Mean	SD
SELFCET	Took part in CET	688	0.382	0.486	1019	0.187	0.390
SELFQAIM	Course aimed at qualification	688	0.295	0.456	1019	0.110	0.313
SELFQUAL	Obtained qualification from course	688	0.141	0.348	1019	0.047	0.212

It included sex, age of youngest child, age of respondent, job tenure, any history of broken employment, ethnicity, educational qualifications, receipt of training from the employer in the preceding year, receipt of training arranged by the respondent in the previous year, self-employed or employee status, full-time or part-time hours, size of workplace (less than 25 or 25-plus), occupational group (one-digit classification), area dummies, and an attitudinal measure of 'overall job satisfaction' obtained with a seven-point rating scale which ranged from 'completely dissatisfied' to 'completely satisfied'.

Methods of matching and estimation

The methodology of the evaluation was 'matching on observables'. Probit equations were used to estimate the probability of participation in guidance, with the variables noted in the preceding paragraph as the regressors X. Non-participants were then matched to participants on the propensity score P(X), by the 'nearest neighbour' method.[8] Under nearest neighbour matching, the match to each guidance case was that case in the non-participant group which had the smallest absolute difference of P(X). The matches were made with replacement, that is, many-to-one matches were permitted. Of the available matching methods, nearest neighbour matching produces estimates with the least bias, but at the cost of the highest variance (because only part of the comparison sample is utilised). The nearest neighbour method is appropriate here since we are chiefly interested in the sensitivity of estimates of effects.

Prior to matching, some observations in the tails of P(X) were removed, so as to impose a 'common support' condition (Rosenbaum and Rubin, 1983; Heckman et al, 1998). Tabulation of the distributions of P(X) indicated that

8 For some results using the 'calliper' or 'radius' method of matching with these data, see Killeen and White (2000).

there was common support for the two samples except in the extreme tails. The method adopted was to remove cases which were greater than the smallest maximum of the two samples (smaller than the greater minimum of the two samples), with the addition (subtraction) of a constant t=0.015. Depending on the specification adopted, trimming removed 1–3 per cent of cases overall.

On the assumption that all relevant variables have been included in the matching, the distributions of the set X should be the same in the participant and non-participant groups (Rosenbaum and Rubin, 1983). A partial specification test is to compare the equality of vectors of means, Mx1 and Mx0, where the subscripts indicate participation or non-participation in guidance. This test was performed for all the reported matching analyses, using Hotelling's T2 measure, separately for the subset of dummy variables and the subset of continuous variables. For all analyses reported here, these tests failed to reject the null hypothesis Mx1 = Mx0 at the $p<0.1$ level.

Estimates of treatment effects under matching were obtained by weighted logit models consisting of an intercept and single regressor, the dummy for receipt of guidance. Guidance participants had a weight of 1 in all analyses. Non-participants had integer weights equal to the number of participants to which they were matched. For comparative purposes, we also report some results from a simple unweighted[9] logit, and from 'regression adjustment' models in which the full set of X-variables was used as regressors along with the treatment dummy, instead of the matching method. All models used robust standard errors (the Huber-White 'sandwich' estimator) appropriate for weighted estimation, which should also provide some protection against heteroskedasticity.

The main variant analyses reported have been derived from estimating P(X) with and without job satisfaction in X, or with an adjusted value of job satisfaction for the guidance sample only, with resulting differences in matching. The rationale for these variants is discussed below.

The role of the 'job satisfaction' variable

The chief focus of this paper is on the role of 'overall job satisfaction' as a potential influence on both participation and on CET outcomes. At the initial stage of the study, the participants were on average very much more dissatisfied with their jobs than were the non-participants (Table 2). For instance, 49

9 'Unweighted' here indicates no use of matching weights; weights to adjust for sample attrition were however retained.

Table 2: *Job satisfaction of participants and non-participants*

Column percentages

	Participants	*Non-participants*
Dissatisfied to some extent	48.6	24.0
Neither satisfied nor dissatisfied	17.2	11.3
Satisfied to some extent	34.2	64.7
N	1015	669

Note: 'dissatisfied' and 'satisfied' results are collapsed from three categories each.

per cent of the guidance sample expressed some dissatisfaction with their jobs, compared with 24 per cent of non-participants.

This difference in job satisfaction between participants and non-participants appears intuitively reasonable: workers are likely to seek careers guidance precisely because they face problematic or frustrating circumstances in their jobs or careers. Without information on job satisfaction, a model of participation in guidance would appear to be seriously deficient and a key assumption underlying the matching estimator would be violated. However, it also seems plausible that job satisfaction might be affected by careers guidance if measured post-entry. This illustrates the general dilemma concerning attitudinal variables in the construction of matching estimators.

One response would be to instrument the potentially endogenous variable. As often, however, the scope for this approach is limited here by the lack of usable instruments in the dataset, and this is a general difficulty with attitudinal measures since they tend to be poorly predicted by behavioural variables (Dulany, 1967; Ajzen, 1988). The practical decision remains limited to inclusion or omission of the variable.

To the extent that the attitudinal variable can be shown not to be influenced by programme participation, it becomes more reasonable to include it in the matching process. It is not the timing of measurement as such, but the causal effect of participation on the attitudinal variable which is problematic. The theory of career guidance (Super, 1957) predicts that career counselling should increase job satisfaction, but this results chiefly from improved matching into appropriate jobs or occupations, a medium-term effect. There is no prediction about the short-term impact, which is of concern here, but it might be supposed that if there were no impact other than from job matching, the short-term impact would be close to zero. Additionally, however, theories of job satisfaction (eg, Locke, 1976) emphasise its relativity to expectations (eg, high expectations lower satisfaction, other things being equal). Again, there is no clear prediction concerning the short-term impact of guidance specifically, which could be to reduce satisfaction (if it raised job expectations), to increase satisfaction (if it led to more 'realistic' job expectations), or to leave satisfaction unaffected (if it did not address job expectations).

In the present study, job satisfaction was measured not only at the initial stage (JS1) but also at the follow-up interview (JS2). It may be possible to make some limited inferences about short-term impacts if these are connected with the medium-term picture. Omitting JS1 from estimation of P(X), we obtained a significantly negative estimate of guidance on JS2 (weighted ordered logit model: $b = -0.60$, $z = -4.13$), which is highly implausible and suggests that the impact of guidance on medium-term job satisfaction cannot be appropriately estimated without taking account of initial job satisfaction in matching. On the assumption that there is no short-term impact of guidance on job satisfaction, JS1 may be regarded as a proxy for pre-programme job satisfaction and be included in an estimate of P(X). Matching on this basis, we found that there was no impact of guidance on JS2 (ordered logit model: $b = -0.08$, $z = -0.012$), which is inherently a more plausible result, consistent with guidance having no impact on job satisfaction either in the short-term or the medium-term. However, experiments with adjusted values of JS1 for the guidance group (see section (c) below for the adjustment method) indicated that the relationship between guidance and JS2 remained insignificantly different from zero over a wide range of adjustments, both negative and positive (results[10] available from the authors). On the one hand this is consistent with the earlier inference that guidance did not impact on medium-term job satisfaction, but on the other hand it shows that such a result would be consistent with the presence of substantial bias in JS1 as a proxy for the pre-treatment level of job satisfaction. In other words, the likely absence of a medium-term impact of guidance on job satisfaction does not rule out the possibility of a short-term impact.

Results from the analysis of CET

(a) Models of participation, and matching

Since the preliminary investigations did not point to a clear decision either to include or to exclude JS1 from the participation model, it is desirable to develop sensitivity analyses concerning the consequences of inclusion or exclusion. Annex Tables A1 and A2 show the results of two probit models for participation in guidance, the first including JS1 in X, the second exclud-

10 For example, adjustment by –0.5 (implying that JS1 was upwardly biased by 0.5) resulted in $b = -0.045$, $z = -0.28$, which was very close to the result with JS1 unadjusted.

Table 3: *Results of nearest neighbour matching for basic matching models*

	Job satisfaction (JS1) included in model of participation	*JS1 not included in model of participation*
Common support condition:		
No of cases excluded from upper tail of P(X)	28	13
No of cases excluded from lower tail of P(X)	1	3
No of guidance cases included	632	647
No of comparison cases matched	350	385
Maximum no of matches to a control case	13	9
Min (P(X)) after exclusions	0.013	0.040
Max (P(X)) after exclusions	0.996	0.983

ing JS1. Across these two versions, the model chi-square fell from 309.4 to 158.3 with a difference of 6 df, confirming the power of job satisfaction to discriminate between guidance participants and non-participants. Other important variables in predicting participation were educational qualifications, full-time employment status, whether the respondent had organised training for herself/himself in the year before the study, the interaction between gender and age of youngest child, and ethnicity.

Table 3 describes some features of the nearest neighbour matches obtained in the two basic analyses. When JS1 was included in the participation model, this extended the tails of the distribution of the predicted probabilities, and resulted in more cases being trimmed to satisfy the common support requirement. Even so, only 2 per cent of the sample (before matching) were excluded when JS1 was present and only 1 per cent when JS1 was absent. As Table 4 shows, under both matches the equality of mean vectors between the guidance and non-participant matched samples was not rejected.

(b) Effects of guidance on CET outcomes: basic models

The two basic matched samples were used to generate tests of the effects of guidance on CET outcomes, in the manner described above. The basic results of these analyses are summarised in Table 5. Also included in the table are the estimates of the treatment effect in a simple model (without matching and without additional regressors), and from 'regression adjustment' (RA) models

Table 4: *Tests of equality of mean vectors of predictor variables*

Hotelling T2 tests	*T2*	*F(df)*	*P**
Variables with (0,1) values			
Nearest neighbour match, JS1 included in X	36.60	1.23 (29,952)	0.20
Nearest neighbour match, JS1 not in X	29.27	1.25 (23,1008)	0.20
Variables with continuous values			
Nearest neighbour match, JS1 included in X	1.95	0.65 (3,978)	0.58
Nearest neighbour match, JS1 not in X	2.33	0.78 (3,1028)	0.51

(without matching but with the same set of 'control' regressors as used in the participation model).

If the results from the two matching models (see rows (4) and (5) of Table 5) were closely similar, then one might infer that the omission of JS1 from the participation model would not be problematic. With SELFQAIM, the absolute difference in point estimates was around 5 per cent and this might support such a conclusion if it were the sole outcome measure of interest. However, for the other two CET outcomes, the omission of JS1 led to point estimates which differed from the results when included by 24 per cent (SELFCET) and 20 per cent (SELFQUAL). These differences are too substantial to be ignored. It is also notable that while the models including JS1 led to lower estimates of impact in the case of SELFCET and SELFQUAL, they led to a somewhat *higher* estimate of impact in the case of SELFQAIM. On this evidence, therefore, one could not generally argue for inclusion of JS1 on the grounds of producing more 'conservative' estimates.

Comparison of these results with their counterparts under the RA method (rows (2) and (3) of Table 5) is instructive. RA produced point estimates when JS1 was omitted which were consistently higher than when JS1 was included. Moreover, the differences were fairly uniform, with inclusion/exclusion of JS1 affecting the estimates by 7–10 per cent. Thus, under RA, the issue of attitudinal variables may appear more tractable. However, all the RA estimates, with or without JS1, lie considerably above the corresponding estimates from the matching models, with or without JS1. Even omitting JS1 from the participation model, nearest neighbour matching produced substantially smaller estimates of the treatment effects than did RA with JS1 included in regressors. In general we would expect the matching estimator to produce the least-biased estimates of treatment effects, and on this basis all the RA estimates appear to be upward biased.

Table 5: *Estimates of effects of guidance on CET outcomes: basic models*

Model (N of obs)		*SELFCET*		*SELFQAIM*		*SELFQUAL*	
		b	*z*	*b*	*z*	*b*	*z*
(1) Simple unmatched	(1674)	1.039	9.07	1.264	9.50	1.270	6.80
(2) Regression adjustment with JS1 in regressors	(1674)	1.015	7.55	1.225	8.01	1.213	5.61
(3) Regression adjustment with JS1 omitted	(1674)	1.111	8.64	1.309	8.88	1.323	6.33
(4) Nearest neighbour match, JS1 in X	(982)	0.754	3.99	1.072	4.94	0.860	2.84
(5) Nearest neighbour match, JS1 not in X	(1032)	0.938	5.63	1.014	5.24	1.030	3.51

Notes: Weighted logit models, b is the estimated effect on log-odds. See text for further details of models. The z-statistic is computed using the Huber-White robust variance estimator.

(c) Results from experiments in adjusting the values of JS1

If JS1 is included in the X variables to estimate P(X), it may be regarded as a proxy for the ideal measure which would have been obtained, in the case of the guidance sample, shortly before the decision to enrol. (In the case of the comparison sample, JS1 provides an appropriate measure of job satisfaction, on the assumption that satisfaction is stable, since there has been no exposure to guidance.) Suppose that for the guidance sample only, the attitude has been shifted by a constant amount, c: then the true or ideal value can be restored for the guidance sample as (JS1-c). As we do not know the value of c, we can experiment with a range of values, and examine the sensitivity of the estimates of the treatment effects to these adjustments.

While this type of adjustment may appear crude, it is consistent with one of the most widely used specifications of the evaluation estimator, termed by Heckman LaLonde and Smith (1999) the 'common treatment effect model'. In their notation, the difference in potential outcomes in the treated (subscript 1) and non-treated (subscript 0) states is Y1-Y0 = _ , leading to the widely used regression model Y = X_ + D_ + U, where D is the assignment or participation dummy, and U represents unobservables. Underlying this model are the assumptions that E(U) = 0 and U1 = U0 (or less restrictively U1|X = U0|X). The adjustment term c is the sample counterpart of the common treatment effect _ . While this is a highly simplistic model (and much of the motivation for nonparametric matching is to provide scope for investigating heterogeneous treatment effects), it provides a convenient framework for an exploratory analysis since individual heterogeneity in the effect can be ignored.

A further simplification which is necessary in order to introduce the adjustment method is to treat the variable JS1 as if measured on an interval scale. Then c can be given the same scaling. The experimental values chosen for c were –0.5, –0.25, –0.125, 0, 0.125, 0.25 and 0.5. The value 0 was included for comparison with the earlier results, since the assumption of an interval scale may in itself considerably affect the estimates. Under the interval scale assumption, the standard deviation of JS1 is 1.82 in the full analysis sample, varying within |0.05| for various subsets selected by matching. Setting c to |0.125| with a sample size of 1000 represents an assumed effect, or bias, of a little more than two standard errors of JS1. Larger values are also included to test the sensitivity of estimates, though a value of c as great as |0.50| may seem implausible for a short-term effect.

What happens to the matching process when these adjustments are introduced? Recall that JS1 was the variable most strongly associated with entry to guidance and that this reflected a lower level of dissatisfaction in the guidance group relative to the non-participant group. Upwardly adjusting the values of JS1 for the guidance group reduces the difference with the non-participant group, lowers the coefficient on JS1 in the participation model, and (since other regressor variables are not highly correlated with JS1) weakens the predictive power of the model. The converse applies when JS1 is adjusted downwards for the guidance participants. The distribution of P(X) is also affected, and this has consequences for the region of common support: the number of observations retained in matching shrinks progressively as c goes from its highest to its lowest value (see second column of Table 6).

Estimates from the variant analyses of CET outcomes are shown in Table 6. The first point to note (from the results for c=0, to be compared with row (4) of Table 5) is that treating JS1 as an interval scale measure considerably inflated the estimates for all three outcome variables. Relative to this new baseline estimate, the analyses with c>0 or c<0 had somewhat different consequences by CET outcome. For SELFCET, all adjustments whether positive or negative had the consequence of substantially reducing the estimated effect of guidance, the range of the reductions being 20–35 per cent. In the case of SELFQAIM, however, for c<0 (ie, when the short-term impact of guidance was assumed positive) the estimate effect increased somewhat (in the range 5–10 per cent), and there was also an increase with c=0.125; but with c=0.25 and c=0.5, the estimates were substantially reduced relative to the baseline case. Finally, the results for SELFQUAL were somewhat similar to those for SELFCET, except that with c=0.125 (ie, when a small negative impact of guidance on short-term satisfaction was assumed), the estimated effect was considerably increased. Across all three outcome variables, indeed, the estimates varied with c in a markedly non-monotonic fashion and the smaller

Table 6 : *Variant estimates of effects of guidance on CET outcomes*

*c**	*N of obs. matched*	*SELFCET*		*SELFQAIM*		*SELFQUAL*	
		b	*z*	*b*	*z*	*b*	*z*
0.5	1021	0.673	3.65	0.865	4.29	0.687	2.46
0.25	1003	0.653	3.56	0.944	4.50	0.896	3.18
0.125	997	0.753	3.93	1.050	4.46	1.326	3.52
0	977	1.000	5.06	1.001	4.22	1.021	3.09
–0.125	971	0.722	3.86	1.046	4.44	0.861	2.49
–0.25	958	0.806	3.80	1.082	4.97	0.966	2.99
–0.50	936	0.682	2.89	1.097	4.90	0.887	2.94

* c = adjustment to JS1 (assumed measured on an interval scale)
Notes: Weighted logit models, b is the estimated effect on log-odds. All models consist of an intercept term and a dummy for participation in guidance. The z-statistic is computed using the Huber-White robust variance estimator.

absolute values of c appeared as likely to produce large shifts in the estimates as did the larger values.

Finally, it is noteworthy that, despite the sensitivity of the point estimates, the hypothesis of no treatment effect was always rejected at least at the $p<0.01$ level, in these as in the foregoing analyses. Thus the inference that careers guidance had a positive impact on these CET outcomes was robust.

Discussion

This research has illustrated the large role that can be played by an attitudinal variable, in this case job satisfaction, within a programme evaluation study. It also illustrates the dilemma for the analyst when a measure of the attitudinal variable can only be obtained for the treated group after they have entered the treated state. The estimates in this study have differed substantially between analyses including, and omitting, the measure of initial job satisfaction. Since estimates shifted in different directions for different outcome variables, no general rule for inclusion or exclusion could be inferred. Data experiments were also carried out, in which the observed value of job satisfaction was adjusted up or down to correct for various assumed impacts of guidance on satisfaction in the short term. These data experiments again suggested a high degree of sensitivity of estimates, but did not permit any general inferences about the direction of bias in the outcome estimates. The results as a whole demonstrated the non-monotonic response of the matching estimator to inclusion, exclusion or adjustment of the attitudinal variable in the model of participation.

These results can be explained in a general sense as reflecting the nonparametric nature of the matching estimator. When a parametric 'regression adjustment' method was used, inclusion or exclusion of the job satisfaction variable in the regressors led to relatively small, and orderly, shifts in the estimates. The price paid for this, however, was a general inflation of the estimates of effects of guidance, relative to the matching estimator.

On this basis, it seems that the decision to include or exclude the post-entry attitudinal variable has to be made on theoretical or practical grounds, or in the light of evidence from other aspects of the same study or from other studies. Here, we would argue that inclusion of the job satisfaction variable in the model of participation is the least-bad decision. Relevant considerations include: the intuitive importance of job dissatisfaction as a general motive for seeking careers guidance; the measurement of the job satisfaction variable, in the present study, soon after entry to guidance thus avoiding medium-term impacts from guidance; the absence of theory-based prediction of a short-term effect of guidance on satisfaction; the absence, in the present study, of any evidence of a medium-term effect of guidance on satisfaction; and, consequent on the previous points, the low probability that any short-term effect of guidance on satisfaction would be as large as the effect of satisfaction on participation in guidance. In other studies, different considerations would apply and the opposite decision might be reached.

Another general conclusion which this research points to is the desirability of conservatism in the interpretation of estimates of treatment effects generated by the matching estimator, when attitudinal influences on participation are likely. The validity of the matching estimator depends on the assumption that all relevant variables are included in the matching set (Lechner, 2000). If an important attitudinal variable is omitted, or if the attitudinal variable is measured with bias, the estimates can be considerably affected, as the present results illustrate. In this study, the effects of guidance on CET outcomes were sufficiently large to remain significant in all the variants and data experiments conducted. However, the observed variation in the magnitudes of the estimates would be sufficient to affect policy conclusions drawn from many other programme evaluation studies reporting smaller treatment effects.

In view of the potential importance of attitudes in participation decisions, there would appear to be value in further research on this topic. One issue is how far the present type of results would be affected by the use of other matching methods, for example kernel density matching, which reduce variance and hence may also reduce the sensitivity of estimates. This is an approach which we hope to pursue in our future research with these data. More substantively, there is a need to build up empirical knowledge about programme impacts on attitudes as well as on the more usual economic outcomes. This requires special efforts to collect attitudinal data before entry to programmes as well as at various points after entry, with parallel measures on non-participants, so that effects on attitudes can be directly estimated.

References

Ajzen, I (1988) *Attitudes, Personality and Behavior*. Milton Keynes: Open University Press

Brodaty, T, B Crepon and D Fougere (2000) 'Using Matching Estimators to Evaluate Alternative Youth Employment Programs: Evidence from France, 1986–1988'. *EALE/IEEA Conference Proceedings 2000*, European Association of Labour Economists

Dehejia, R H and S Wahba (1998) *Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs*. NBER Working Paper 6586, Cambridge MA: National Bureau of Economic Research

Dulany, D E (1967) 'Awareness, Rules and Propositional Control: A Confrontation with S-R Behaviour Theory' in Horton, D and T Dixon (eds) *Verbal Behaviour and S-R Behaviour Theory*. Englewood Cliffs NJ: Prentice-Hall

Frohlich, M, A Heshmati and M Lechner (2000) 'A Microeconometric Evaluation of Occupational Rehabilitation Programmes in Sweden'. *EALE/IEEA Conference Proceedings 2000*, European Association of Labour Economists

Heckman, J, H Ichimura and P Todd (1998) 'Matching As An Econometric Evaluation Estimator'. *Review of Economic Studies*, 65, 261–294

Heckman, J J, R J LaLonde and J A Smith (1999) 'The Economics and Econometrics of Active Labor Market Programs' in Ashenfelter, O and Card, D (eds) *The Handbook of Labour Economics*, Vol III. Amsterdam: North Holland

Heckman, J, H Ichimura, J Smith and P Todd, (1998) 'Characterizing Selection Bias Using Experimental Data'. *Econometrica*, 66(5), 1017–1098

Killeen, J and M White (2000) *The Impact of Careers Guidance on Adult Employed People*. DfEE Research Report, Sheffield: Department for Education and Employment (forthcoming)

Lechner, M (1999) *Propensity Score Matching and Bias when Treatment Heterogeneity is Ignored*. Working Paper: University of St Gallen

Lechner, M (2000) 'Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods'. Paper presented at a Conference on *The Evaluation of Economic and Social Policies*, London: Royal Statistical Society

Locke, E A (1976) 'The Nature and Causes of Job Satisfaction', in Dunnette, M D (ed) *The Handbook of Industrial and Organizational*. Chicago: Rand McNally

Rosenbaum, P R and D B Rubin (1983) 'The central role of the propensity score in observational studies for causal effects'. *Biometrika*, 70(1), 41–55

Rubin, D B (1974) 'Estimating causal effects of treatments in randomized and nonrandomized studies'. *Journal of Educational Psychology*, 66, 688–701

Rubin, D B (1979) 'Using multivariate matched sampling and regression adjustment to control bias in observational studies'. *Journal of the American Statistical Association*, 74, 318–328

Super, D E (1957) *The Psychology of Careers*. New York: Harper and Row

Annex: Probit models of participation in careers guidance

Table A1: *Model with job satisfaction included in regressors*

N=1674; Wald chi-square (43 df)=309.41, p<0.001; log likelihood = –951.8

	Coef.	*S.E.*	*z*	*P>|z|*
FEMALE	–0.0838	0.0893	–0.938	0.348
Youngest child:				
3–4	–0.4803	0.2237	–2.147	0.032
5–11	0.2083	0.1529	1.362	0.173
12–16	0.2772	0.2529	1.096	0.273
FEM*3–4	0.3023	0.2959	1.022	0.307
FEM*5–11	–0.3697	0.1954	–1.892	0.059
FEM*12–16	–0.3273	0.2866	–1.142	0.253
NONWHITE	0.2683	0.1420	1.890	0.059
AGE*10	0.0072	0.0481	0.150	0.881
TENURE	–0.0121	0.0171	–0.711	0.477
TENURE2*100	0.0037	0.0633	0.059	0.953
BREAK	0.1047	0.0889	–1.177	0.239
INACTIVE	0.0015	0.0049	0.302	0.763
ANY QUALN.	0.3482	0.1281	2.719	0.007
DEGREE	–0.1819	0.0878	–2.070	0.038
A-LEVEL	–0.3755	0.1130	–3.322	0.001
Training in previous year:				
SELF	0.2878	0.0805	3.575	0.000
EMPLOYER	0.1199	0.0761	1.575	0.115
SIZE <25	–0.1090	0.0808	–1.349	0.177
SELF-EMP.	–0.0619	0.1419	–0.436	0.663
FULLTIME	0.3345	0.1012	3.307	0.001
SOC=1	–2.3174	0.3742	–6.191	0.000
SOC=2	–2.2375	0.3781	–5.918	0.000
SOC=3	–2.2978	0.3758	–6.114	0.000
SOC=4	–1.9932	0.3681	–5.415	0.000
SOC=5	–2.6493	0.3922	–6.755	0.000
SOC=6	–1.8776	0.3730	–5.033	0.000
SOC=7	–1.9315	0.3769	–5.125	0.000
SOC=8	–2.6415	0.3914	–6.750	0.000
SOC=9	–2.3068	0.3911	–5.898	0.000
AREA=2	0.3681	0.1266	2.908	0.004
AREA=3	0.5521	0.1236	4.466	0.000

AREA=4	0.3439	0.1430	2.406	0.016
AREA=5	0.2203	0.1920	1.148	0.251
AREA=6	0.1995	0.2196	0.908	0.364
AREA=7	0.2927	0.1541	1.899	0.058
AREA=8	0.1889	0.1775	1.064	0.287
JS1=2	0.0724	0.1506	0.481	0.631
JS1=3	–0.3545	0.1308	–2.710	0.007
JS1=4	–0.3617	0.1329	–2.721	0.007
JS1=5	–0.8694	0.1208	–7.200	0.000
JS1=6	–1.0713	0.1361	–7.874	0.000
JS1=7	–1.3476	0.1563	–8.622	0.000
CONSTANT	1.7659	0.4352	4.057	0.000

Notes: BREAK – any break in employment, 1992–96; INACTIVE, time spent in inactive status, 1994–96; SOC – Standard Occupational Classification, 1-digit (reference class=uncodable); JS1 – job satisfaction measure.Table A2 Model with job satisfaction excluded from regressors.

N = 1674; Wald chi-square(37 df)=158.30, p<0.001; Log likelihood = –1036.8

	Coef.	*S.E.*	*z*	*P>|z|*
FEMALE	–0.1199	0.0858	–1.397	0.162
Youngest child:				
3–4	–0.3783	0.2119	–1.785	0.074
5–11	0.1789	0.1520	1.176	0.239
12–16	0.1838	0.2546	0.722	0.470
FEM*3–4	0.1751	0.2763	0.634	0.526
FEM*5–11	–0.3284	0.1907	–1.722	0.085
FEM*12–16	–0.1632	0.2855	–0.572	0.568
NONWHITE	0.2818	0.1340	2.103	0.035
AGE*10	–0.0245	0.0468	–0.524	0.600
TENURE*10	0.0832	0.1655	0.503	0.615
TENURE2*100	–0.0061	0.0062	–0.988	0.323
BREAK	–0.0364	0.0845	–0.431	0.667
INACTIVE	–0.0006	0.0049	–0.121	0.903
ANY QUALN.	0.4526	0.1215	3.724	0.000
DEGREE	–0.2021	0.0846	–2.390	0.017
A-LEVEL	–0.3573	0.1073	–3.330	0.001
Training in previous year:				
SELF	0.2632	0.0774	3.398	0.001
EMPLOYER	–0.0066	0.0727	–0.090	0.928
SIZE < 25	–0.1661	0.0773	–2.149	0.032
SELF-EMP.	–0.2263	0.1382	–1.637	0.102
FULLTIME	0.3566	0.0973	3.664	0.000
SOC=1	–2.2103	0.3932	–5.621	0.000
SOC=2	–2.1078	0.3971	–5.308	0.000
SOC=3	–2.1923	0.3929	–5.580	0.000
SOC=4	–1.814543	0.3861	–4.700	0.000
SOC=5	–2.4699	0.4069	–6.071	0.000

SOC=6	–1.8153	0.3901	–4.654	0.000
SOC=7	–1.6995	0.3939	–4.315	0.000
SOC=8	–2.3722	0.4100	–5.786	0.000
SOC=9	–2.0362	0.4075	–4.997	0.000
AREA=2	0.3798	0.1219	3.116	0.002
AREA=3	0.4844	0.1191	4.065	0.000
AREA=4	0.3389	0.1381	2.454	0.014
AREA=5	0.3401	0.2040	1.667	0.095
AREA=6	0.3051	0.2091	1.460	0.144
AREA=7	0.2429	0.1516	1.600	0.110
AREA=8	0.1783	0.1715	1.040	0.298
CONSTANT	1.0284	0.4353	2.363	0.018

Notes: BREAK – any break in employment, 1992–96; INACTIVE, time spent in inactive status, 1994–96; SOC – Standard Occupational Classification, 1–digit (reference class=uncodable).