Società italiana di
economia pubblica

XXIV
CONFERENZA

ECONOMIA INFORMALE, EVASIONE FISCALE E CORRUZIONE

Pavia, Aule Storiche dell'Università, 24 - 25 settembre 2012

EDUCATION MERGERS AND STUDENTS' DROPOUT DECISIONS

ROSSELLA IRACI CAPUCCINELLO

Education Mergers and Students' Dropout Decisions

Rossella Iraci Capuccinello *

June 30, 2012

Abstract

This paper investigates the causal effect of the choice to enroll in a merged Further Education College on the students probability to drop out.
We use a large administrative data set (ILR) on the population of students enrolled in the Further Education sector. We employ the propensity score matching methodology making use of several types of matching algorithms. We check the overlap and common support region and we asses the quality of our matching through the use of different procedures. In particular, we calculate and confront the standardized bias and the pseudo-R^2 before and after matching and we register a reduction of both measures. This indicates that our propensity score specification brings to a considerable improvement of covariate balance.
Finally, we assess the sensitivity of our estimates to deviations from the identifying assumptions of unconfoundedness and common support. We indirectly verify that our estimates are robust to deviations from the Conditional Independence Assumption through three different approaches. Therefore, we calculate the M-H bounds, we estimate the ATT with the inclusion of a confounder and we estimate the effect of a treatment known to have none through the use of multiple control groups.
We find a negative and significant effect of enrolling in a Further Education College which has recently been merged on the probability of dropping out. This is consistent with the fact that merged colleges can obtain a reduction in costs and a greater diversification of the teaching and service provision probably allowing for the students to choose courses which are more in line with their ability and preferences and for improved tuition and support services.

Keywords: Dropout, mergers, matching models.
JEL Classification: I20, I21, I28.

*r.iracicapuccinello@lancaster.ac.uk, Department of Economics, Lancaster University Management School, Bailrigg Lancaster LA1 4YX, UK.

1 Introduction

This paper investigates the causal effect of the choice to enroll in a recently[1] merged Further Education College on the student's probability to drop out. As pointed out in the previous chapters, dropping out from college implies a high cost for both individuals and the society. In fact, young people risk entering NEET (not in education, employment or training) as a consequence of leaving education before having achieved a qualification. This can obviously have serious long term consequences on the future labour market outcomes of those students, on social cohesion and on the State budget as it implies higher expenditures on social care.
Since the publication of the Foster Report in 2005 a clear commitment to create incentives for Further Education colleges to focus on achievement and progression of their students has been made and this commitment seems to be confirmed by the new Government. At the same time, as can be noted from figure 1, the report created strong incentives for the colleges to merge exploiting both economies of scale and of scope [Foster, 2005]. This, together with the increasing national and international pressure to raise the proportion of students achieving a Further or Higher Education qualification clearly shows the policy relevance of some of our research questions.

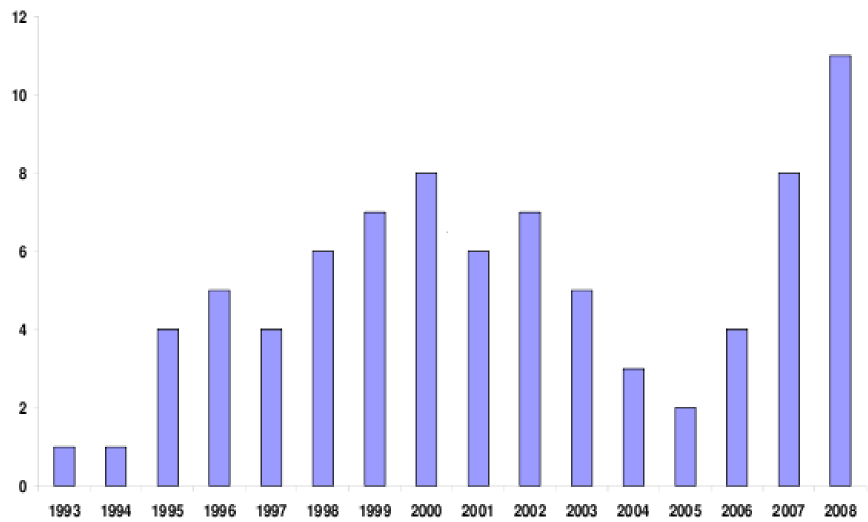

Figure 1: Number of Mergers per year. Source:[Payne, 2008]

[1]Recently merged, in our analysis, indicates a college which has been merged at any point after 1997.

While, traditionally, there has been a certain level of attention in the management sciences literature to the effects of college mergers on college finances, efficiency or organization [Goedegebuure, 1992, Lang, 2002], the effects of these institutional restructuring on student outcomes and, in particular, on the students' probability to dropout has been widely neglected. Therefore, our work is innovative in that it focuses on the effects on students' dropout behavior, thus concentrating on the students rather than on the institution.
However, we refer to the management literature to understand the reasons why colleges merge and the possible link that this type of institutional reorganization has with students' dropout. Payne [2008] and Lang [2002] show that there are two main reasons to advocate the need for colleges to merge. On one hand, there are the economic and financial reasons connected with generating production efficiencies. This is mainly related to the opportunity to exploit economies of scale both in teaching and service provision. However, as noted by Lang [2002], it can be connected with the objective of benefiting from government financial incentives for merged institutions. On the other hand, college mergers can create greater opportunities for the diversification of courses and for the provision of a wider range of services to students. This can be true for both small and large institutions involved as for both of them merging entails a greater potential for diversification.
Our work attempts to shed some light on this topic, trying to determine if the positive effect of mergers on the costs of education provision and on diversification is accompanied by a positive or a negative effect on the individual's probability to dropout. Our expectation is that the increased diversification of the curricula within merged institutions and the possibility to provide more and better infrastructures and services should have the effect of increasing the students retention as they can get a better match between their personal aspirations and learning preferences and the learning provision of the college. Moreover, it is likely that they can also benefit from higher quality support and orientation services.
However, we are aware of the possible increase in dropout rates which could immediately follow a merger due to initial organizational difficulties. In order to minimize the bias due to these organizational problems we consider colleges which have merged from 1997/1998. This means that while we analyze students outcomes in the year 2002/2003, the colleges have merged at any point in time between 1997/1998 and 2002/2003. Therefore, most of them would have already overcome the initial organizational difficulties and would have already reorganized their curricula and started to benefit from both financial and diversification gains.
We use once again the ILR data set, which consists of administrative data on the whole population of students enrolled in the English Further Education sector. Our final sample consists of 447,940 students, 52,360 of whom enrolled

in a Further Education college which has merged from 1997 onwards.
We employ a propensity score matching approach, making use of several types of matching algorithms. Therefore, we match the students enrolled in a merged college with students from the control groups which are similar to them in terms of a propensity score calculated on the basis of gender, maturity, ethnicity, disability, prior attainment and number of colleges in the Local Learning Skills Council. We, then, check the overlap and common support region and we assess the quality of our matching through the use of different procedures. This allows us to conclude that our propensity score specification brings to a considerable improvement of covariate balance.
Finally, we assess the sensitivity of our estimates to deviations from the identifying assumptions of unconfoundedness and common support. We indirectly verify that our estimates are robust to deviations from the Conditional Independence Assumption through three different approaches proposed respectively by Rosenbaum [2002], Ichino et al. [2008] and Imbens [2004].
Our estimates suggest that enrolling in a Further Education College which has recently been merged implies a reduction of the probability of dropping out of 1.6 to 4 percentage points. This is consistent with the fact that merged colleges can obtain a reduction in unit-costs and a greater diversification of the teaching and service provision. Therefore, increasing the probability of students choosing courses which are more in line with their ability and preferences and of colleges to provide them with improved tuition and support services. It is, in fact, very likely that this increased choice of courses available to the students in a single Further Education Institution and the availability of more and higher quality support services will affect the student's probability of dropping out. As a consequence, we can conclude that the current policy of creating incentives for Further Education colleges to merge is likely to have positive effects on students' outcomes in terms of improved student dropout rates.

2 Econometric Methodology

Estimating the effect of attending a Further Education college which has recently been merged on the students' probability of dropping out, implies being able to appraise the counterfactual situation in which those same students are instead attending a college which has not been merged. As it is evidently impossible for the same student to be observed in both situations at the same time, we will have to find a way to overcome what it is usually called the fundamental evaluation problem.
Another problem which may arise in the attempt to estimate our causal effect

is the presence of selection bias. That is the possibility that the students attending a merged college and students attending a non merged one could have significant differences even in the absence of treatment (enrolling in a merged college). This possibility makes unwise to proxi the outcome that students attending a merged college (treated) would have had hadn't they attended a merged college, with the one of students not attending a merged institution.
The econometric methodology used in this paper is propensity score matching. Matching allows to estimate causal treatment effects overcoming the fundamental evaluation problem and the possible existence of selection bias.
The parameter that we estimate in our analysis is the average treatment effect on the treated "ATT", which is defined as

$$\tau_{ATT} = E(\tau|D = 1) = E(Y(1)|D = 1) - E(Y(0)|D = 1) \qquad (1)$$

Therefore the ATT is equal to the difference between the expected outcome of treated individuals who have actually been treated and the expected outcome of treated individuals hadn't they been treated. If we were able to demonstrate that the outcome is independent from selection into treatment we could overcome the problem of not being able to observe the counterfactual. As a consequence, we have to define a mechanism able to describe the process of assignment into treatment.
Matching methods allow to carefully select a group of non-treated individuals similar to the treated ones in all the relevant pre-treatment characteristics (X). Therefore, the difference in outcomes between those individuals and the treated ones will be attributable to the treatment.
Following a proposal by Rosenbaum and Rubin [1983] we will use a *balancing score*. This score will ensure that at each of its values the distribution of X will be the same for treated and untreated individuals. The propensity score (PS) is one of the possible balancing scores and corresponds to the conditional probability of receiving the treatment given the pre-treatment variables:

Rosenbaum and Rubin [1983] also show that the conditional indipendence assumption remains valid if controlling for p(X) instead of X, that is conditionally on p(X) the treatment and potential outcomes are independent:

$$Y(1), Y(0) \perp D|p(X) \qquad (unconfoundedness\ given\ PS). \qquad (2)$$

As noted by Imbens [2004], if condition 2 holds conditioning on the propensity score removes all biases due to observable characteristics X.
The second key assumption about treatment assignment is the overlap or common support condition:

$$0 < P(D = 1|X) < 1 \qquad (overlap). \qquad (3)$$

The basic intuition behind this assumption is that there has to be at least one similar individual in the counterfactual state for each treated one. In other words, for every single value of X the probability of finding a treated and a control individual must be greater then 0 [Heckman et al., 1999].

Now given assumptions 2 and 3, the matching estimator for the ATT is:

$$\tau_{ATT} = E_{p(X)|D=1}\{E[Y(1)|D = 1, p(X)] - E[Y(0)|D = 0, p(X)]\}$$

Thus, computing the ATT entails taking the mean outcome of treated and controls individuals, comparing them for each given value of p(X) in the common support and finally weighting them for the propensity score distribution. All matching estimators can be seen as a special case of the following where the weights W_{ij} take different forms:

$$\tau_{ATT} = \sum_{i \in T} (Y_i - \sum_{j \in C} W_{ij} Y_j) \, w_i \tag{4}$$

T and C indicate respectively the treatment and control individuals, W_{ij} denote the weights assigned to the control individuals when matching with the treated one and w_i represent a re-weighting needed to re-build the outcome distribution for the treated.

3 Data and Estimation Results

We use a large administrative dataset provided by the Learning Skills Council, the Individual Learner Record (ILR). In particular, the cross section data we use, provides complex information about the whole population of students enrolled in Further Education Colleges in the year 2002-2003.
The production of unbiased estimates of a treatment effect through propensity score matching depends heavily from the quality of the data used. As pointed out by Mueser et al. [2007], the use of administrative data to obtain propensity score matching estimates of the average treatment effect on the treated can be a very effective tool. One reason for this is the fact that data on the outcome for both treated and untreated individuals comes from the same source. Another important characteristics is the availability of large datasets.
In our case the use of administrative data allowed us to use the whole population of students and to dispose of a very rich set variables to use in the estimation of the propensity score.
Our final sample includes 447,940 student 52,360 of whom are enrolled in further education colleges which have been merged from the year 1997-1998 onwards.

Table 1: Descriptives.

Covariates	*Completers*	*Dropout*	*Total*		
Age>20	39,399	7,611	47,010		
%	83.81	16.19			
Male	197,997	18,498	216,495		
%	91.46	8.54			
Disability	20,447	1,489	21,936		
%	93.21	6.79			
Ethnic origin: Bangladeshi	5,800	642	6,442		
%	90.03	9.97			
Ethnic origin: Black African	11,983	1,225	13,208		
%	90.73	9.27			
Ethnic origin: Black Caribbean	9,569	1,262	10,831		
%	88.35	11.65			
Ethnic origin: Black Other	7,872	1,073	8,945		
%	88.00	12.00			
Ethnic origin: Chinese	5,224	260	5,484		
%	95.26	4.74			
Ethnic origin: Indian	13,848	783	14,631		
%	94.65	5.35			
Ethnic origin: Pakistani	16,459	1,404	17,863		
%	92.14	7.86			
Ethnic origin: Asian Other	6,175	566	6,741		
%	91.60	8.40			
Ethnic origin: Other	9,884	1,107	10,991		
%	89.93	10.07			
No prior qualification	20,211	2,746	22,957		
%	88.04	11.96			
Prior qualification< 1	1,359	143	1,502		
%	90.48	9.52			
Prior qualification= 1	43,768	4,327	48,095		
%	91.00	9.00			
Prior qualification= 2	142,446	7,770	150,216		
%	94.83	5.17			
Prior qualification= 4 or 5	519	133	652		
%	79.60	20.40			
Prior qualification unknown	191,975	21,530	213,505		
%	89.92	10.08			
Covariates	*Observations*	*Mean*	*Stand. Dev.*	*Min*	*Max*
Number of colleges in the LLSC	447940	10.17779	5.657223	1	24

We decided to restrict our sample only to full time, full year, non working students as students enrolled as part-time or in a short course are likely to present different characteristics from the majority of students and because the determinants of their dropout behavior are different [Montmarquette et al., 2007, Weiler and Pierro, 1988, Warren and Lee, 2003].
We also restricted our sample to students aged 16-24, considering the much higher dropout rate of more mature students. Finally, we have decided to exclude from our analysis the students which have transferred to other courses as their proportion in our data is very limited (less than 2%) and as we don't possess information about where they transferred to.
A description of the main characteristics of those students with respect to the dropout behavior is provided in table 1.

3.1 Estimation of the propensity score

In our analysis, making use of matching estimators implies being able to match the students enrolled in a recently merged college with students that are very similar to them in all relevant observable pre-treatment characteristics except for being enrolled in a non merged institution. However, finding perfect matches on all those characteristics can be a rather complicate task due to the emergence of the curse of dimensionality. The solution proposed by Rosenbaum and Rubin [1983] entails finding close matches to the treated individuals on the basis of a one-dimensional measure: the propensity score.
We estimate the propensity score to enroll in a recently merged institution conditional on a set of pre-treatment characteristics through the use of a probit model. Furthermore, we included only variables which where not influenced by the the student decision to enroll in a merged college.
The pre-treatment characteristics we have chosen to include in our propensity score specification include gender, maturity, ethnicity, disability, prior attainment and number of colleges in the Local Learning Skills Council. Following the suggestion of Rubin and Thomas [1996], Heckman et al. [1998], they were chosen, first of all, on the basis of the existence of a well known relationship with the outcome of interest (dropout probability) but also for their capacity to predict treatment. In fact, as table 2 shows, all of these variables are significantly affecting the probability of enrolling in a merged college.
We have also used the leave one-out cross validation method even though our main discriminant in choosing to include one variable has been the existence of economic theory or empirical findings in support of a relationship between the variable and the probability of dropout. Moreover, the use of administrative data collected by the Learning Skills Council ensures that the information about treated and untreated individuals comes from the same

questionnaire/form[2].

Table 2: Propensity score[a] estimation, year 2002-2003

Covariates	*dF/dx*	*(s.e.)*
Age > 20	$0.025***$	(0.002)
Male	$-0.008***$	(0.001)
Disability	$0.006***$	(0.002)
Bangladeshi	$-0.063***$	(0.003)
Black african	$-0.042***$	(0.002)
Black caribbean	$-0.019***$	(0.003)
Black other	$0.021***$	(0.003)
Chinese	$0.053***$	(0.005)
Indian	$0.036***$	(0.003)
Pakistani	$0.008***$	(0.002)
Asian other	$-0.016***$	(0.003)
Ethnic other	$-0.009***$	(0.003)
No qualification	$0.022***$	(0.004)
Qualif. $<$ *level 1*	$0.066***$	(0.011)
Qualif. level 1	$0.009**$	(0.003)
Qualif. level 2	$-0.037***$	(0.003)
Qualif. level 4-5	$0.045***$	(0.014)
Qualif. unknown	$0.045***$	(0.003)
No. collleges in LLSC[b]	$-0.007***$	(0.000)

[a]Propensity score estimated with a probit
[b]Local Learning Skills Council

3.2 Common support

One of the 2 key assumptions for the validity of the propensity score matching estimation is the overlap or common support condition. This assumption implies that both the students enrolled in a merged institution and the students enrolled in a non-merged one, should have a positive probability of enrolling in a merged college as shown by their propensity score.
To ensure that this condition is met we will, first of all, proceed with a graphical analysis. In fact, plotting the propensity score distribution for treated

[2]Heckman et al. [1999] show that this is an important requirement for the matching methods to be correctly implemented.

and untreated individuals gives an idea of the level of overlap that we have achieved. Figure 2 shows that we have a very good overlap. However, we can notice that in the tails of the distribution there are relevant differences in the density of the propensity score. As for estimating the average treatment effect on the treated, satisfying the common support condition only implies being able to find potential matches for the treated individuals in the untreated group, from our graphical analysis we can conclude that this condition is satisfied.
Another method which we have used to confirm the results of our graphical analysis is the min-max method to discard from our analysis the treated observations that lie outside of the region of common support. This method entails finding the minima and maxima of the propensity score distribution for both treated [0.0191, 0.3433] and untreated individuals [0.0057, 0.3572] and defining our region by selecting the highest of the two minima and the lowest of the two maxima [0.0191, 0.3433]. It can be noticed that, in this case, the region of common support corresponds to the interval showing the distribution of the propensity score for the treated individuals which indicates that we have perfect overlap at least for as much as regards the estimation of the ATT. As a consequence, we can conclude that our estimated effect can be considered very representative. Moreover, having a perfect overlap entails being able to find matches which are very close to the treated ones using matching without replacement [Mueser et al., 2007].

3.3 Matching Quality

When implementing propensity score matching it is of utmost importance to check for the quality of the matching. In practical terms, this means checking for the covariates balance in the matched sample. Obtaining a good covariates balance implies that the marginal distribution of each covariate is very similar for treated and untreated individuals.
The most widely used method for checking the covariates balance is the so called Standardized Bias or standardized difference in means. This method proposed by Rosenbaum and Rubin [1985] entails comparing the standardized difference in means for each of the covariates, between treated and untreated individuals before and after matching.
The standardized bias is computed applying the following formula:

$$SB = 100 \frac{(\bar{x}_{\mathrm{non-merged}} - \bar{x}_{\mathrm{merged}})}{\sqrt{0.5(s^2_{\mathrm{non-merged}} - s^2_{\mathrm{merged}})}} \tag{5}$$

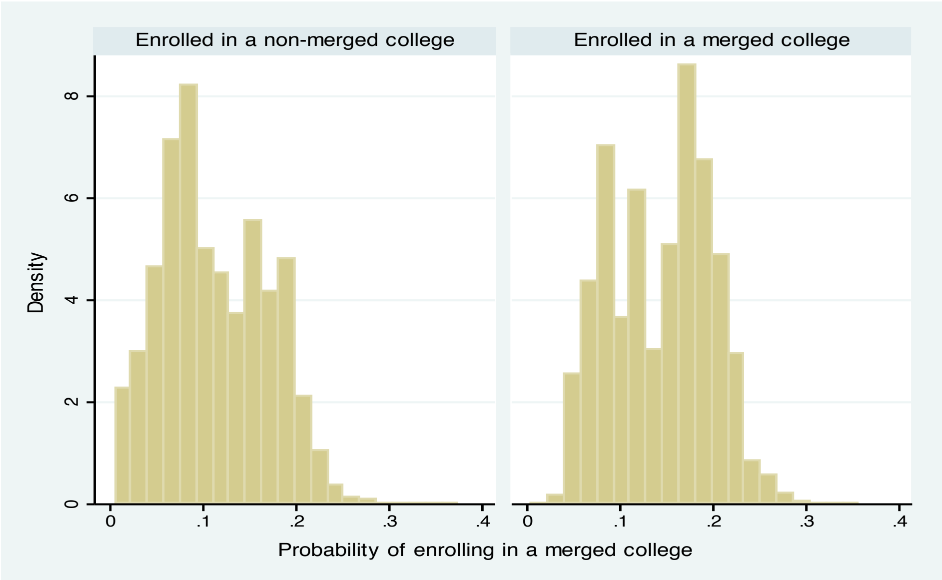

Figure 2: Propensity Score Distribution by Treatment.

where $\bar{x}_{\text{non-merged}}$ and $s^2_{\text{non-merged}}$ are, respectively, the mean and the variance for the students enrolled in a college which has not been merged and $\bar{x}_{\text{merged}}$ and s^2_{merged} are the mean and variance for students enrolled in a merged college.

A reduction in the standardized bias after matching helps proving that the covariates balance is improved by the matching procedures. Rosenbaum and Rubin [1985] consider a standardized difference in means with an absolute value lower than 20 as an acceptable level. Caliendo and Kopeinig [2008] note that it is often not very clear what level of the standardized bias could be considered as sufficient. However, in this application, we achieve levels of the standardized bias which are well under the threshold considered as acceptable by Rosenbaum and Rubin [1985].

Another method suggested by Sianesi [2004] suggests to compare the $Pseudo-R^2s$ before and after matching. In fact, $Pseudo-R^2s$ give a measure of how well the covariates explain the treatment probability. As a consequence if the covariates after matching are well balanced the $pseudo-R^2$ should be quite low.

In our analysis we implement both methods. Firstly, we calculate the Standardized Bias for all of our covariates and with all of the matching algorithms used and then we plot the resulting standardized differences in means for each covariate before and after matching. As can be seen in the graphs 3, 4, 5, 6, 7, 8, 9 and 10 we achieve a considerable reduction in the Bias in all the ATT estimations.

In particular, the graphs 9 and 10 show that the radius estimators achieve the best results in reducing the bias. In fact, *Radius matching* is a type of caliper matching where multiple matches are used within a certain caliper. The use of multiple matches helps keeping the variance low while the imposition of a caliper reduces the bias. In particular, 9 out of 19 covariates have standardized differences in means lower than 1, and 9 between 1 and 2 with only one covariate taking the value of 2.8 [3]. Before matching only two covariates had standardized biases lower than 1, 3 had values between 1 and 2 and the remaining where taking higher values including three of them taking values well higher than the threshold of 20 proposed by Rosenbaum and Rubin [1985]. This noticeable reduction in bias after matching supports our assumption that matching allows us to create two groups of individuals comparable in all aspects except the participation into treatment.
When using the $pseudo - R^2$ method we can see that it decreases from a value of 0.045 for the unmatched sample to a value of 0, 0.002 or 0.003 for the matched one depending on the matching algorithm used. Once again the lowest value of zero is obtained when radius matching with a caliper of 0.005 is implemented.
A third option to check for the covariates balance, would be to implement the stratification test proposed by Dehejia and Wahba [2002]. However, there is no consensus in the literature about the relevance of this type of test as a post-matching check for covariates balance, especially in cases where matching methods not based on stratification are considered.
Overall, we can conclude that we have achieved a good balance of the covariates and that the quality of our matching is exceptionally good.

3.4 The causal effect of enrolling in a merged college

We now investigate the causal effect of enrolling in a merged college on the probability of dropping out from further education. To do so, we proceed with the estimation of the average treatment effect on the treated through the use of different matching algorithms. Table 3 shows the estimated ATT for all of these models including the model for the unmatched sample. This allows us to compare the estimated effect before matching with the effect obtained after implementing matching techniques. In particular, we have estimated the average treatment effect of enrolling to a recently merged institution on the probability of dropping out through the use of nearest neighbor (with and without replacement), Caliper matching (with and without replacement and with a caliper of 0.005, 0.05 and 0.1), multiple neighbor (with 5, 10 and 15

[3] We consider the absolute value of the standardized difference in means.

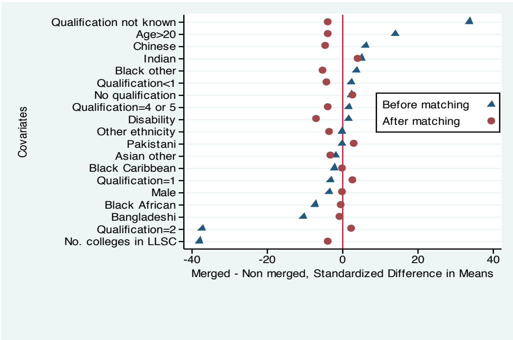

Figure 3: Covariates Balance. Nearest Neighbor, no replacement; Caliper matching(cal=0.1 and cal=0.05), no replacement.

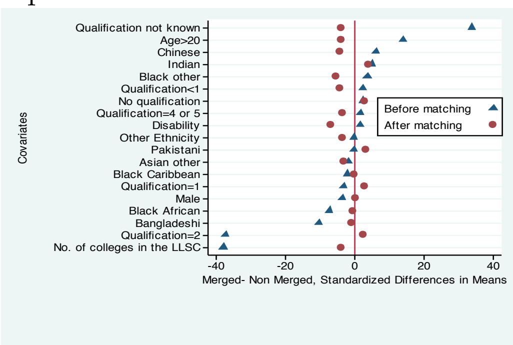

Figure 4: Covariates Balance. Caliper matching(cal=0.005), no replacement.

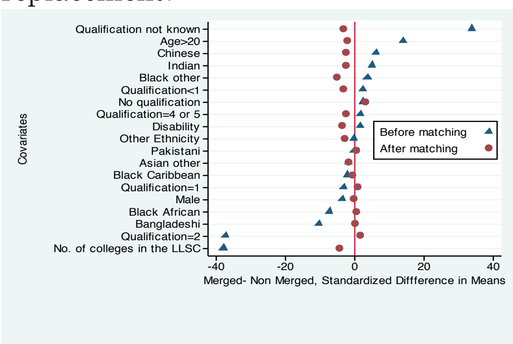

Figure 5: Covariates Balance. Multiple Neighbor (10 neighbors).

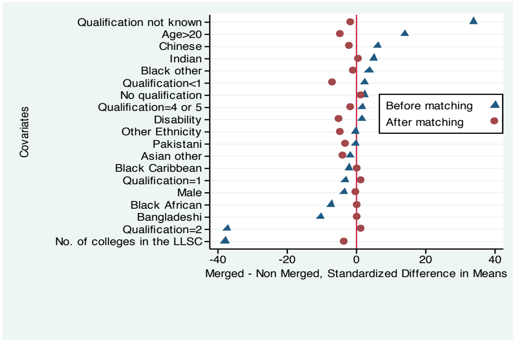

Figure 6: Covariates Balance. Nearest Neighbor and Caliper matching (cal=0.1, cal=0.05 and cal=0.005), replacement.

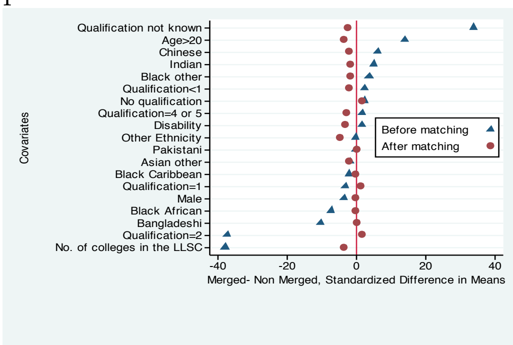

Figure 7: Covariates Balance. Multiple Neighbor (5 neighbors).

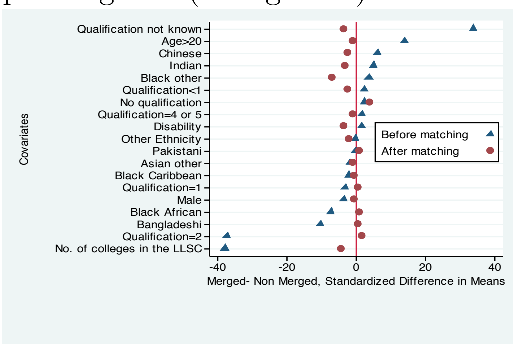

Figure 8: Covariates Balance. Multiple Neighbor (15 neighbors).

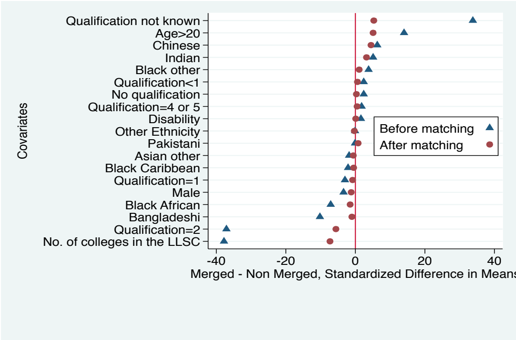

Figure 9: Covariates Balance. Radius matching (cal=0.05).

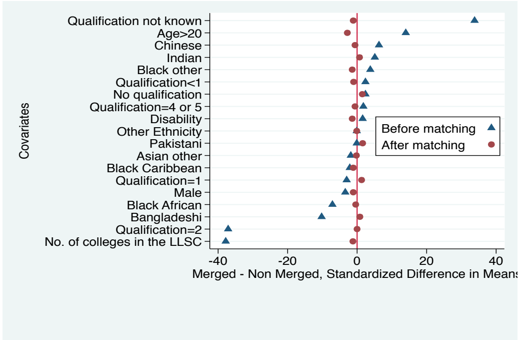

Figure 10: Covariates Balance. Radius matching (cal=0.005).

neighbors) and radius matching (with a caliper of 0.05 and 0.005). We have decided not to implement a stratification and a kernel matching because the size of our data set would make the computations unfeasible[4].

In table 3, we report the ATT, the relative standard error and the number of treated and control observations in the common support. This is important to show that we have perfect overlap having in all cases except one all the observations in the common support.

The nearest neighbor and the caliper matching estimations with replacement all give estimates of the ATT that are not significant and quite low values for the standardized bias. This is in line with the findings of Smith and Todd [2005] about the fact that matching with replacement trades a reduced bias with an increase in variance. In fact, allowing for replacement improves the matching quality as untreated individuals which are very similar to many treated ones are matched several times. On the other hand, this is likely to increase variance as the total number of used controls is decreased.

However, all of the remaining matching estimators give significant estimates of the effect of enrolling in a merged college on the probability of dropping out from further education. In particular, while the before matching model estimates a nearly 1% reduction in the probability of dropping out, the remaining estimators show that enrolling in a recently merged institution reduces the probability of dropping out of 1.6 to 4%.

The nearest neighbor estimator with no replacement gives a significant effect equal to a decrease in the probability of withdrawal of 2.2 percentage points. When using caliper matching the ATT is once more estimated to be nega-

[4]We have a data set of 447,940 observations and these two types of matching techniques are computationally very demanding

tive and equal to a 2.2 percentage reduction [5] in the probability of dropping out and this estimate does not change when we vary the caliper from 0.1 to 0.005. This indicates that overall the control observation which is deemed to be the nearest neighbor is very close to the treated observation to which it is matched. This is a further confirmation of the good quality of our matches. When using the multiple neighbor algorithm we obtain very different estimates depending on the number of neighbors used. Decreasing the number of neighbors from 15 to 5 increases the estimated average treatment effect on the treated from -2.8% to a -4% and the corresponding standardized biases are quite low.

Table 3: Propensity score matching estimation. Average treatment effect on the treated, year 2002-2003

Matching Algorithm[a]	*ATT*	*Stand. Error*	*Observations*[b]	
			Treated	*Controls*
Unmatched	-0.009***	0.001	52,360	395,580
Nearest Neighbor, with replacement	-0.006	0.034	52,360	395,580
Nearest Neighbor, no replacement	-0.022***	0.002	52,360	395,580
Caliper=0.005, with replacement	-0.006	0.034	52,360	395,580
Caliper=0.005, no replacement	-0.022***	0.002	52,357	395,580
Caliper=0.05, with replacement	-0.006	0.034	52,360	395,580
Caliper=0.05, no replacement	-0.022***	0.002	52,360	395,580
Caliper=0.1, with replacement	-0.006	0.034	52,360	395,580
Caliper=0.1, no replacement	-0.022***	0.002	52,360	395,580
Multiple neighbors, N=15	-0.028***	0.009	52,360	395,580
Multiple neighbors, N=10	-0.032***	0.011	52,360	395,580
Multiple neighbors, N=5	-0.040***	0.015	52,360	395,580
Radius, caliper=0.05	-0.016***	0.001	52,360	395,580
Radius, caliper=0.005	-0.020***	0.001	52,360	395,580

[a]Balancing Property and Common Support satisfied

[b]For the unmatched sample total observations are shown. For the matched samples we show only observations on the common support.

Finally, the estimates using radius matching gives an ATT of -1.6% using a caliper of 0.05 and an ATT of -2% using a caliper of 0.005. In this last

[5]As can be seen from table 3 when caliper matching with replacement is implemented the estimated effect is not significant. Also in this case changing the caliper does not affect the estimates.

case, we also obtain the estimate with the lowest standardized bias. We will consider this last estimation as our preferred one. However, the entity of the standardized bias is so low and the estimated effect so near when using different matching algorithms that any of the techniques used with the exception of matching with replacement could be trusted.

4 Sensitivity Analysis

The identification of the average treatment effect on the treated relies on the conditional independence assumption, which is to say all existing selection bias is assumed to be determined by the observable characteristics used as covariates in the propensity score estimation. Thus, our identification strategy is based on the assumption that there are no unobserved factors which are influencing both the probability that a student enrolls in a recently merged college and his/her probability to drop out. This is equivalent to assuming that there is no hidden bias. Since it is not possible to test this assumption directly, we will have to perform a sensitivity analysis in order to assess the robustness of our conclusions to its failure.
The literature [Caliendo and Kopeinig, 2008, Imbens, 2004] shows different strategies, both parametric and non-parametric that can be used in order to check the plausibility of the CIA and the sensitivity of the results to its failure. In this analysis we will implement three of these techniques, all of them non-parametric.
The first one, originally proposed by Rosenbaum [1987] and implemented by Heckman et al. [1997] and by Ichino et al. [2008] relies on the existence of multiple control groups. The basic idea consists in estimating the effect of a treatment that if the unconfoundedness assumption holds should be equal to zero. A good example is a case where it is possible to divide the control units in two groups: the individuals which are eligible but not treated and the individuals which are not eligible. The following step entails defining the treatment variable as 1 if the individual is eligible but not treated and zero if it is not eligible. Subsequently, we estimate the effect of this treatment on the outcome. It is clear that in the absence of unobserved variables affecting both the original treatment and the outcome, the estimated effect of being eligible but not treated should be equal to zero. Imbens [2004] note that not rejecting the null hypothesis that the effect is zero, doesn't necessarily mean that the conditional independence assumption holds. However, finding a zero effect supports its plausibility. In particular, this type of analysis can be considered a strong one in the cases where the two control groups are likely to have differences in the bias.
In our case, we have created two control groups dividing the untreated stu-

dents in two categories: the ones living in an LLSC[6] where at least one merged institution exists, and the ones whose LLSC doesn't have any merged institution. This is, in someway, similar to having two control groups, one formed by non-participants which are entitled to the treatment and one formed by non-participants which are not entitled to it. This similarity holds if we assume that students will find a college in the Local Learning Skills Council where they live. This is true in the majority of the cases, as college students are, normally, living at home while studying and as the LLSC is, normally, a quite big area. However, we are aware that this approach might be problematic for the students living very near to the LLSC borders where it might be easy to find a nearby college which is in a different LLSC. One way to solve this problem would be using the distance between the student home and the colleges to define who can be considered "entitled" and who not. However, we are not currently in a position to be able to create such a variable. We have, then, defined a treatment variable equal to 1 if the control student lives in an LLSC where there is at least one merged college and zero if the control student lives in an LLSC where merged colleges are not available. This two groups are, in our opinion, likely to have different biases because the presence of merged colleges in an LLSC can be considered an indicator of the fact that the Local Learning Skills Council takes a pro-active role in trying to improve the quality of the learning provision. The last step is to estimate the ATT of being an "entitled not treated" on the probability of dropping out. We have estimated this effect with nearest neighbor and with caliper matching and in both the cases we obtained a zero effect.

As a consequence, we can conclude that this analysis provides support for the plausibility of the CIA assumption. However, as we have previously mentioned, it cannot be concluded that the CIA definitely holds. Our second strategy will, thus, try to assess how strongly an unobserved factor would have to influence the treatment probability in order to bring our estimated effect to zero.

To do so, we use the so-called Rosenbaum Bounds [Rosenbaum, 1987]. The idea is that the probability of being treated is a function of both observed and unobserved factors, such that:

$$P_i = P(D_i = 1 | x_i, u_i) = F(\beta x_i + \gamma u_i) \tag{6}$$

where x_i represents the observed factors influencing the probability of being treated for the individual i, and u_i are the unobserved ones. As a consequence, β is the effect of the observed covariates on the probability of participation and γ is the effect of the unobserved factors. If the unconfoundedness assumption holds and there is no hidden bias the value of gamma will be zero

[6] Local Learning Skills Council

and the second term of the equation disappears.
Assuming that the F is the logistic distribution and that the matching procedures gives us couples of matched treated and untreated individuals (i and$j$), the odds ratio of being treated can be written as:

$$\frac{P_i(1-PJ)}{P_j(1-P_i)} = \frac{exp(\beta x_i + \gamma u_i)}{exp(\beta x_j + \gamma u_j)} \tag{7}$$

because of the matching procedures $x_i = x_j$. Thus, we are left with $exp[\gamma(u_i - u_j)]$. This means that the individuals i and$j$ will have different probabilities of receiving treatment and this difference will depend from γ and from the difference in the unobserved factors.
The assumption of unconfoundedness entails that either γ takes a value of zero such that the unobserved factors have no effect on the probability of participation or that the unobserved factors are the same across treated and untreated individuals ($u_i = u_j$).
Thus, as noted by Rosenbaum [2002] the odds ratio of being treated shown in equation 7 are bounded by the values $\frac{1}{e^\gamma}$ and e^γ. Therefore, the only case where the treated and untreated individuals have the same probability of being treated is when $e^\gamma = 1$. In this case, it is clear that there is no hidden bias and the CIA holds. Higher values of e^γ will mean that there is hidden bias. In particular, a certain value of e^γ represents the effect an unobserved factor would have to produce on the odds of being treated in order to explain away the estimated average treatment effect. Becker and Caliendo [2007] have devised a Stata routine which allows to implement this sensitivity analysis exploiting the Mantel-Haenszel test statistics (Q_{MH}). Rosenbaum [2002] shows that for values of $\gamma > 1$ the Q_{MH} test-statistics is bounded by two known distribution Q^+_{MH} and Q^-_{MH}. More specifically, Q^+_{MH} represents the case in which the average treatment effect has been overestimated, while Q^-_{MH} represent a situation where the effect has been underestimated.
Table 4 shows the (Q_{MH}) bounds and their significance level for different values of e^γ and with different types of propensity score matching estimators. The Q_{MH} bounds for $e^\gamma = 1$ show a scenario where the estimated ATT is free of hidden bias in this case the Q_{MH} bounds take a value between 4.2 and 17.9 depending on the type of matching estimator. This shows that there is strong evidence of an effect of enrolling in a merged college on the probability of dropping out.
In the presence of negative selection for which the students that are most likely to enroll in a merged institution have a lower likelihood of dropping out, we would have overestimated the ATT. The Q^+_{MH} is not very informative in this case because with a negative selection bias increasing the odds that a student enrolls in a merged institution (e^γ) would certainly make Q^+_{MH} even more significant. This is in fact shown by the fact that for all of the matching

algorithms used our estimates are robust even to the existence of an unobserved factor which is doubling the probability of being treated[7].
However, if we concentrate on the bound Q^-_{MH} we can see that our results are sensitive to the existence of an unobserved factor which would increase the probability of being treated by more than 50% in all the matching algorithms where we allow for replacement and to one that it is increasing the same probability by more than 75% when we use multiple neighbor with 5 or 10 neighbors. In the remaining cases the unobserved factor would have more than double the participation probability in order for our estimates to be affected[8].
Therefore, we can conclude that our estimates are quite robust to the existence of hidden bias. However, since there is any direct way to test the validity of the CIA, as we are not able to know the potential outcome of treated individuals had they not been treated, this result only makes the unconfoundedness assumption more plausible. Therefore, we will use a third strategy in order to further confirm this plausibility.
Ichino et al. [2008] have proposed a method to assess the sensitivity of the ATT estimates to the failure of the unconfoundedness assumption. The key idea consists in simulating the distribution of a confounder variable and its association the treatment and the outcome given the observed covariates through the definition of a set of four parameters:

$$Pr(U = 1|T = i, Y = j, X) = Pr(U = 1|T = 1, Y = j) \equiv p_{ij} \tag{8}$$

where $i, j \in \{0, 1\}$. Therefore, the p_{ij} gives the probability that $U = 1$ in each of the four categories given by the different values of the treatment and outcome variables. The following step consists in introducing the predicted confounder variable in the estimation of the propensity score. This propensity score will then be used to estimate the average treatment effect on the treated through the use of a matching algorithm. The choice of the parameters defining the distribution of the unobservable implies a certain deviation from the unconfoundedness. In our case, we have assumed that the confounder is distributed as the students' prior attainment (level 2) which is the variable with the highest standardized bias before matching. We have then computed the ATT through the use of the nearest neighbor algorithm with no replacement and repeated the matching estimation 100 times. The resulting ATT is indicating that enrolling in a merged college reduces the probability of dropping out by 2.1%. Therefore, our estimates are not sensitive to the existence of a confounder distributed as the students prior attainment (level 2).

[7] In fact, the estimates for $e^\gamma = 2$ are all significant.

[8] In table 4 we have reported our sensitivity results only for up to a value of $e^\gamma = 2$ however, in most of the cases our estimates were robust even to an unobserved factor increasing the probability of being in a merged college by a factor of 4.

Table 4: Sensitivity to the presence of hidden bias.

Matching Algorithm		$e^{\gamma}=1$	$e^{\gamma}=1.25$	$e^{\gamma}=1.5$	$e^{\gamma}=1.75$	$e^{\gamma}=2$
N. N., with repl.	Q^{+}_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^{-}_{MH}	4.204***	1.632*	0.367	2.137**	3.68***
N. N., no repl.	Q^{+}_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^{-}_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.005, with repl.	Q^{+}_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^{-}_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.005, no repl.	Q^{+}_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^{-}_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.05, with repl.	Q^{+}_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^{-}_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.05, no repl.	Q^{+}_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^{-}_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.1, with repl.	Q^{+}_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^{-}_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.1, no repl.	Q^{+}_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^{-}_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
M.N., N=15	Q^{+}_{MH}	17.897***	25.922***	32.680***	38.579***	43.855***
	Q^{-}_{MH}	17.897***	10.057***	3.729***	1.572**	6.194***
M.N., N=10	Q^{+}_{MH}	16.211***	23.366***	29.401***	34.676***	39.398***
	Q^{-}_{MH}	16.211***	9.231***	3.601***	1.106	5.216***
M.N., N=5	Q^{+}_{MH}	13.748***	19.484***	24.337***	28.588***	32.401***
	Q^{-}_{MH}	13.748***	8.167***	3.678***	0.056	3.326***
Radius, caliper=0.05	Q^{+}_{MH}	6.733***	19.580***	30.228***	39.395***	47.488***
	Q^{-}_{MH}	6.733***	6.021***	16.500***	25.458***	33.334***
Radius, caliper=0.005	Q^{+}_{MH}	6.781***	19.627***	30.276***	39.443***	47.538***
	Q^{-}_{MH}	6.781***	5.972***	16.451***	25.408***	33.282***

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Overall, the different strategies used in this chapter to assess the sensitivity of our estimates to the failure of the unconfoundedness assumption go in the direction of confirming that CIA is plausible.

5 Concluding Remarks

This paper investigates the effect of enrolling in a recently merged college of further education on the students' probability of withdrawal. This is, to our knowledge, the first attempt to evaluate mergers with a focus on the students' outcome and specifically on dropout. Up to this point, decisions about college mergers, either from the college management or from the Government or Government agencies have been taken on the basis of the expected results in terms of efficiency gains, reduction in unit-costs and changes in the diversification of the learning provision. In this analysis, we try to evaluate mergers focusing on their effects on the students rather than on the institution. Therefore, this research is innovative in that it introduces another dimension in the evaluation of the further education sector mergers. This is a quite radical change in focus that allows to have a broader picture of the reasons why colleges should or should not merge.

We do so, through the use of matching methods as they allow to evaluate the effect of enrolling in a recently merged institution on the student probability to drop out overcoming the fundamental evaluation problem. In fact, since we don't have experimental data guaranteeing the random assignment of students into merged and non-merged institution, we have employed a technique that allows us to compare our treated observations with untreated ones similar to them in any relevant aspect except the participation to treatment.

We employ a large administrative data set on the whole population of students enrolled in the Further Education sector in England in the year 2002/2003. The use of this extremely large data set gives us the opportunity to obtain matching estimates of high quality as the availability of an high number of controls makes it easier to find close matches.

We estimate the average treatment effect on the treated through the use of different matching algorithms and we find that enrolling in a recently merged further education college decreases the probability of dropping out of 1.6 to 4 percentage points [9]. We also show that employing matching estimators makes the bias much smaller than when employing a naive estimator. This result seems to confirm that there is no trade-off between the effects of mergers on the merged institution and the effects on the individual students. However, our finding also shows that evaluating the possibility of a merger should take into consideration the expected reduction effect in the students' dropout rates as well as the opportunity to realize efficiency gains, costs reductions and increased curricula diversification.

We implement three types of sensitivity analysis to check whether our estimated average treatment effect is sensible to the failure of the conditional independence assumption. The result of this analysis suggests that the unconfoundedness assumption is plausible, thus confirming that we can interpret the estimated ATT as a causal effect.

The findings of this study have the interesting policy implication that the evaluation of possible mergers should take into account also the effects on students. When the effects on the students' dropout rates are taken into account, the idea of mergers becomes an even more attractive one since we show that enrolling in merged institution is beneficial for the students. The main argument against mergers has been that they cause disruption in the

[9] The estimate varies according to the matching algorithm employed.

students' life. However, we show that at least in the case in which the effect on students outcomes is evaluated over a longer period than just the first few months after a merger takes place, the long-term positive effect of mergers offsets any possible short-term negative effect caused by initial organizational problems.

References

S. O. Becker and M. Caliendo. Sensitivity analysis for average treatment effect. *Stata Journal*, 7(1):71–83, 2007.

M. Caliendo and S. Kopeinig. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72, 2008.

R. H. Dehejia and S. Wahba. Propensity score matching methods for non-experimantal causal studies. *Review of Economics and Statistics*, 84(1): 151–161, 2002.

Foster. Realising the potential; a review of the future role of further education colleges. Technical report, Department for Business, Innovation and Skills, 2005.

L. C. J. Goedegebuure. *Mergers in Higher Education.* Centre for Higher Education Policy Studies, Utrecht, 1992.

J. J. Heckman, H. Ichimura, and P. E. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4, Special Issue):605–654, 1997.

J. J. Heckman, H. Ichimura, J. Smith, and P. E. Todd. Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098, 1998.

J. J. Heckman, LaLonde, and J. Smith. The economics and econometrics of active labor market programs. In O. Ashenfelter and D Card, editors, *Handbook of Labor Economics*, volume III, pages 1865–2097, Amsterdam, 1999. Elsevier.

A. Ichino, F. Mealli, and T. Nannicini. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, 23(3):305–327, 2008.

G. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29, 2004.

D. W. Lang. There are mergers, and there are mergers: The forms of inter-institutional combination. In *Higher Education Management and Policy*, volume 14. OECD/IMHE, 2002.

C. Montmarquette, N. Viennot-Briot, and M. Dagenais. Dropout, school performance, and working while in school. *The Review of Economics and Statistics*, 89(4):752–760, November 2007.

P. R. Mueser, K. R. Troske, and A. Gorislavsky. Using state administrative data to measure program performance. *Review of Economics and Statistics*, 89(4):761–783, 2007.

L. Payne. The evidence base on college size and mergers in the Further Education sector. Technical Report 19, Department for Innovation, Universities and Skills, 2008.

P. R. Rosenbaum. The role of a second control group in an observational study (with discussion). *Statistical Science*, 2:292–316, 1987.

P. R. Rosenbaum. *Observational studies.* NY: Springer, New York, 2nd edition, 2002.

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

P. R. Rosenbaum and D. B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38, 1985.

D. B. Rubin and N. Thomas. Matching using estimated propensity score: relating theory to practice. *Biometriks*, 52(1):249–264, 1996.

B. Sianesi. An evaluation of the Swedish system of active labour market programmes in the 1990s. *Review of Economics and Statistics*, 86(1):133–155, 2004.

J. Smith and P. E. Todd. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353, 2005.

J. R. Warren and J. C. Lee. The impact of adolescent employment on high school dropout: Differences by individual and labor-market characteristics. *Social Science Research*, 32:98–128, 2003.

W. Weiler and D. Pierro. Selection bias and the analysis of persistence of part-time undergraduate students. *Research in Higher Education*, 29(3): 261–272, 1988.