DETERMINING THE EFFICACY OF BRIDGING PROGRAMMES IN THE FACULTY OF SCIENCE AT THE UNIVERSITY OF KWAZULU-NATAL

M. Murray
School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal, South Africa
e-mail: *murraym@ukzn.ac.za*

Key words: Heckman model, propensity score.

Summary: The vast majority of mainly black African students enrolling at a higher education institution come from a township school where a lack of resources and teacher training creates an environment of rote learning where students leave with only a superficial understanding of some of the linguistic and numeracy concepts needed to successfully complete a higher education based field of study. To address this problem universities have put in place additional teaching programmes that are designed to help bridge this gap. This paper examines the efficacy of some of these bridging programmes using regression adjustment and propensity score matching methods to control for a possible selection bias that can occur with observational studies. To control for a possible selection bias that may occur when selection into treatment is being determined by a set of confounding variables that may be unobservable, Heckman's switching regression model was also fitted to the dataset that was collected.

1. Introduction

South Africa has an education system that is highly polarized. On the one hand we have a cohort of mainly White and Indian students who are able to attend private or Model C based schools where the additional resources that they have at their disposal allow for the appointment of teachers who are far better qualified to help bridge the articulation gap that exists between a secondary and higher education based education. For the vast majority of mainly African students, however, their education takes place in a vastly under-resourced township school where a lack of teacher training often leads to a superficial engagement with texts and a measure of rote learning that has been designed to deliver the correct 'answer' rather than understand the thought processes behind the derivation of that answer. As a consequence. Scott et al. (2013) have found that amongst all school leavers, only 18% manage to do well enough to qualify for entry into a higher education based institution (HEI). Amongst those that qualify, 33% drop out in their first year of study with only 45% managing eventually to complete their studies. All these figures point not only to an education system that is

AMS: 62P99

in crisis, but also a society that is being populated each year by an ever increasing cohort of students with limited employment opportunities. As Scott et al. (2013) have observed, "*If we want to succeed ... then we need more matriculants with schools and colleges showing us how they are going to improve students. Whereas certain circumstances like family issues, poverty and poor schooling are beyond our control ... we must be willing to do what we can with what's in our control?*" One of the primary objectives of this paper is to help address this problem by looking at the efficacy of two bridging programmes that operate at the University of KwaZulu-Natal (UKZN). Because selection into anyone of these two bridging programmes is not random, one first needs to control for a possible selection bias before one can determine whether the two programmes are effective in helping underprepared students to adjust to a university based education.

Table 1: Graduation rates (%) that have been achieved in degree regulation time: 2006 first-time entering cohort, excluding UNISA.

	African	Coloured	Indian	White	All
3-year degrees	20	20	26	43	29
4-year degrees	30	28	31	47	36
3-year diplomas	16	27	27	38	20
All 3- and 4-year qualifications	20	24	28	44	27

Table 2: Attrition rates (%) by the end of degree regulation time: 2006 first-time entering cohort, excluding UNISA.

	African	Coloured	Indian	White	All
3-year degrees	39	50	37	31	29
4-year degrees	41	47	43	33	39
3-year diplomas	45	45	39	38	44
All 3- and 4-year qualifications	42	47	39	33	40

Source: A proposal for undergraduate curriculum reform in South Africa: The case for a flexible curriculum structure. Report of the Task Team on Undergraduate Curriculum Structure (Scott et al., 2013).

Looking at the graduation and attrition rates that appear in Tables 1 and 2 one can see how racially skewed the performance profile of South African students has become. Given that most of these African students come from a township school, as far back as the early 1980s, universities began to put in place programmes of extended learning that were designed to help these students bridge the gap between what they have been taught and what they need to know in order to successfully complete a degree at a higher education institution. The aim in this paper is to determine whether two such programmes that run in the Faculty of Science at the University of KwaZulu-Natal are successful in achieving this objective.

As a starting point for this paper, some notation will be introduced. A very important distinction between an average treatment effect for the entire population and an average treatment effect for the treated section of our population will then be made. Noting that assignment to treatment is not done in a random manner, several approaches that can adjust for this source of possible bias will then be outlined. Based on the assumption that selection into a bridging programme is being determined by a set of fully observable covariates X that induce a conditional independence assumption (CIA) in the population, the results from a regression based technique will then be compared with those obtained from a propensity score match. If selection into a bridging programme is also being determined by another set of variables that are not fully observable (such as those associated with their socio-economic status) then a modelling approach that makes use of switching regressions will also need to be considered (Terza, 1998).

2. Statistical methodology

Given that one is interested in determining whether a bridging programme is effective in improving the throughput rate of students in a given faculty, let T_i represent a treatment indicator variable which we will set equal to one if student i is allowed to enrol for a bridging programme and set to 0 otherwise. Let Y_i denote a response variable for this paper which we will define more carefully in a later section of this paper. It becomes important now to make a clear distinction between the actual outcome Y_i that one can observe and two potential outcomes that one would like to be able to observe; namely $Y_i(0)$ representing the outcome that student i would record if he/she was not bridged and $Y_i(1)$ representing the outcome that he/she would record if that same student were to be bridged. Being able to observe both outcomes would allow one to identify

$$\Delta_i = Y_i(1) - Y_i(0)$$

as a treatment effect for student i in our dataset. Only being able to observe one of these potential outcomes, however, creates a modelling framework where one can at best hope to identify an average treatment effect

$$ATE \equiv E\{Y_i(1) - Y_i(0)\}$$

that one can associate with an individual that is randomly drawn from the overall population or an average treatment effect

$$ATT \equiv E\{Y_i(1) - Y_i(0)|T_i = 1\}$$

that one can associate with an individual who is randomly drawn from the treated subpopulation. Because one is interested in determining whether a bridging programme is successful (or not) it is the estimated value that is obtained for the average treatment effect on the treated (ATT) that will be relevant for this study.

If every student entering the faculty were to be randomly assigned to a bridging programme (or not), then simply subtracting the average value of the outcomes that are being recorded by the non- bridging students from those of the bridging students would produce the following intuitive estimator for ATE (and ATT); i.e.

$$\text{Naive estimator} = \frac{\sum_{i=1}^{N} T_i Y_i}{\sum_{i=1}^{N} T_i} - \frac{\sum_{i=1}^{N} (1-T_i) Y_i}{\sum_{i=1}^{N} (1-T_i)}. \tag{1}$$

Given the modelling context, however, assignment to a bridging programme is based on a very specific set of criteria being met. For example, the total number of matric points that each bridging student has managed to obtain must be such that it would not have allowed for a normal entry into the faculty. Furthermore, these students must also have come from a disadvantaged school background. As a consequence, the student profile (X) of those that gain entry into the bridging programme (such as their socio-economic status and race) may be very different from those who meet the criteria for a normal entry into the faculty. Given that one is able to observe a significant treatment effect amongst those that are treated, one must now be sure that this effect is caused by the bridging programme rather than by some of the background variables in X that may be exerting a confounding influence both on T and the potential outcomes $Y_i(1)$, $Y_i(0)$ respectively. Essentially, from a statistical point of view, one needs first to control for a possible difference in baseline characteristics between the treated and non-treated groups before one attempts to estimate a treatment effect that can be associated with the bridging programme. This can be done using a regression adjustment method or a propensity score providing all the confounding variables that may be affecting the treatment assignment variable T and the potential outcomes have been fully observed and controlled for in the modelling process. If some of these potential confounders are not observable, then an instrumental variable will have to be found or an explicit model for the treatment allocation process will have to be developed (Terza, 1998; Angrist and Krueger, 2001).

2.1. Regression adjustment based methods

A regression based approach begins by specifying a parametric model for both of the potential outcomes (only one of which will be able to be observed), namely

$$Y_i(0) = \beta_0 + X_i\beta_1 + e_i$$

and

$$Y_i(1) = \beta_0 + \delta + X_i\beta_1 + e_i. \tag{2}$$

Since

$$Y_i = (1 - T_i)Y_i(0) + T_iY_i(1)$$

a model for the outcome that one is able to observe can now be given by

$$Y_i = \beta_0 + T_i\delta + X_i\beta_1 + e_i. \tag{3}$$

Ordinary least squares estimation can then be applied to Equation (3), producing an estimate for the treatment effect δ that will be unbiased providing a sufficient number of covariates have been included in X so as to prevent a correlation between T_i and the error term e_i from arising. Essentially one would like to include enough variables in X so that a conditional independence assumption between $T_i|X$ and $Y_i(0), Y_i(1)|X$ eventually holds true for the population that one intends to study.

2.2. Matching on a propensity score

Providing that a sufficient number of variables have been included in X so that the CIA condition given in Equation (3) becomes true, within each cell defined by a particular outcome of $X_i = x$, a random selection into treatment can now be regarded as having taken place. This result allows the naive estimator that has been given in Equation (1) to be used to produce the following estimate for an average treatment effect for that particular cell

$$ATT(x) = E(Y_i(1)|X_i = x, T_i = 1) - E(Y_i(0)|X_i = x, T_i = 0)$$

noting that the summation in Equation (1) is now taken over all those observations for which X_i has been fixed at a particular outcome x. Since the CIA condition given in Equation (3) ensures that one has

$$E(Y_i(0)|X_i, T_i) = E(Y_i(0)|X_i) \tag{4}$$

and

$$E(Y_i(1)|X_i, T_i) = E(Y_i(1)|X_i), \tag{5}$$

one can write

$$\begin{aligned} ATE(x) &\equiv E(Y_i(1)|X_i = x) - E(Y_i(0)|X_i = x) \\ &= E(Y_i(1)|X_i = x, T_i = 1) - E(Y_i(0)|X_i = x, T_i = 0) \\ &= E(Y_i(1)|X_i = x, T_i = 1) - E(Y_i(0)|X_i = x, T_i = 1) \equiv ATT(x) \end{aligned}$$

and thus obtain the following result

$$\begin{aligned} ATT &\equiv (E(Y_i(1) - Y_i(0)|T_i = 1) \\ &= E[E(Y_i(1)|X_i, T_i = 1) - E(Y_i(0)|X_i, T_i = 1)|T_i = 1] \\ &= E_{X|T=1}[E(Y_i(1)|X_i, T_i = 1) - E(Y_i(0)|X_i, T_i = 0)|T_i = 1]. \end{aligned}$$

ATT can therefore be evaluated by first stratifying the data into cells defined by each particular value of X. Within each cell one can then use Equation (1) to compute an estimate for

$$ATT(X) = E(Y_i(1)|X_i, T_i = 1) - E(Y_i(0)|X_i, T_i = 0)$$

which can then be averaged over the conditional distribution $P(X|T_i = 1)$ that is assumed for the treated observations. Note that Equations (4) and (5) do not require that a specific parametric form be given for those expressions as would be the case with a regression adjustment procedure. The successful implementation of this technique however does require that an appropriate number of treated and non-treated observations be found in each of the cells defined by a specific outcome of X. If some cells have no non-treated observations then the treated observations in these cells will have to be discarded resulting in a smaller sample size being used to estimate ATT. To reduce the estimation bias that results from having to throw away these observations, Rosenbaum and Rubin (1984), proposed that matching be done on a single variable, the so-called propensity score, which under a CIA assumption takes on the following form

$$p(X_i) = Pr(T = 1|X_i). \tag{6}$$

Providing one is able to consistently estimate a propensity score for each individual in our dataset, an average treatment effect can be estimated by first matching individuals that have been treated with those that have a similar propensity score that have not been treated. Since an exact match may not always be possible, one could consider specifying a maximum absolute difference between propensity scores (called a caliper) within which a match would be regarded as being acceptable. Because choosing too small a caliper will reduce the number of matched observations that one can use to compute a treatment effect, kernel and local-linear based methods have also been used to compare each treated unit with a weighted average of the outcomes from all the untreated units with a higher weight being placed on those untreated units whose propensity based scores are closest to that of the treated individual. These locally linear and kernel based methods generally produce estimators for an average treatment effect that have a lower variance than their caliper based counterparts.

For binary treatment variables T, one can show that

$$ATE = E\{Y_i(1) - Y_i(0)\} = E\left\{\frac{T_iY_i}{p(X_i)} - \frac{(1-T_i)Y_i}{(1-p(X_i)}\right\} \tag{7}$$

and

$$ATT = E(Y_i(1) - Y_i(0)|T_i = 1) = \frac{E(T_iY_i)}{P(T_i = 1)} - \frac{E\left(\frac{(1-T_i)}{1-p(X_i)}p(X_i)Y_i\right)}{p(T_i = 1)}. \tag{8}$$

Because the propensity scores are being used as inverted weights in Equations (7) and (8), very small propensity scores will cause the IPW estimators generated from Equations (7) and (8) to become very unstable. To overcome this problem a class of "doubly-robust" estimators have been developed that use a logistic (or probit) model to produce estimated propensity scores $\hat{p}(X_i, \hat{\gamma})$ for each individual in the population. Regression adjustment methods are then applied separately to those observations that have been treated and then to those observations that have not been treated. More specifically, for those observations that have been treated, parameters estimates for a regression model that one can associate with the treated observations are obtained by minimizing

$$\sum_{T_i=1}^{N} \frac{(Y_i - \alpha_1 - X_i\beta_1)^2}{\hat{p}(X_i, \hat{\gamma})} \tag{9}$$

with respect to $\alpha_1, \beta_1)$ over those observations that have been treated. Similarly, parameter estimates for a regression model that one can associate with the untreated observations are obtained by minimizing

$$\sum_{T_i=1}^{N} \frac{(Y_i - \alpha_0 - X_i\beta_0)^2}{1 - \hat{p}(X_i, \hat{\gamma})} \tag{10}$$

with respect to (α_0, β_0) over those observations that have not been treated. An estimate for ATT can then be formed using

$$ATT = N_T^{-1} \sum_{T_i=1}^{N_T} [\hat{\alpha}_1 + X_i\hat{\beta}_1] - [\hat{\alpha}_0 + X_i\hat{\beta}_0] \tag{11}$$

where $(\hat{\alpha}_0, \hat{\beta}_0)$ and $(\hat{\alpha}_1, \hat{\beta}_1)$ denote the parameter estimates that minimize Equations (9) and (10) respectively and where the summation in Equation (11) is now taken over the observations in the treated population only. Although we now have to build a separate model for the treatment allocation variable T and then another model for the response variable Y, only one of these models needs to be correctly specified in order to obtain a consistent estimate for the treatment effect that we want to measure (Lunceford and Davidian, 2004; Bang and Robins, 2005; Funk and Westreich, 2008).

2.3. Heckman's treatment selection model

If selection into a bridging programme is being governed by a fully observable set of covariates X that create a conditional independence property in the population that we are wanting to study, then the results obtained from a regression adjustment can be compared with those from a propensity score match. If some of these variables are unobservable then a method that makes use of instrumental variables will have to be considered (Angrist and Krueger, 2001). Being able to find a set of instruments that are correlated with T but uncorrelated with Y apart from a common effect on Y_i through T_i is often a difficult exercise. To overcome this problem, Heckman (1979) has developed a modelling approach where one first corrects for a possible sample selection bias in one's treatment effect by fitting a probit (or logit) model to one's treatment assignment process T. More specifically, with u_i denoting a $N(0,\sigma^2)$ error term and Z_i another set of observable covariates that uniquely help to determine the assignment to treatment process, the probit model sets $T_i = 1$ if

$$X_i\gamma + Z_i\theta + u_i > 0 \tag{12}$$

and $T_i = 0$ otherwise. The estimated treatment assignment probabilities $\hat{T}_i$ that one obtains from this probit fit can then be substituted as instruments into the following model for the observed response variable Y_i

$$Y_i = X_i\beta + \delta T_i + e_i. \tag{13}$$

Known as a two- stage least squares estimation procedure, the error terms u_i and e_i that appear in Equations (12) and (13) are assumed to have a bivariate normal distribution with

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} | T_i, X_i, Z_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right).$$

Often referred to as being a Heckit or endogenous switching regression model, the estimators arising from this two-step fitting procedure can be shown to be consistent and asymptotically normal. Because of issues associated with identification, one needs to make sure that we can find at least one variable Z_i that affects the treatment assignment process Equation (12) but not the outcome Equation (13) (Sartori, 2003; Briggs, 2004).

3. Results

In order to help bridge the articulation gap that exists because of a township based education, the University of KwaZulu-Natal (UKZN) has put in place several bridging programmes two of which exist for students wanting to enrol for a BSc degree in the College of Agriculture, Engineering and Science. In the BSc 4-year Augmented programme, students are allowed to take two years to complete what for normal entry students would be their first year of study with special classes and extra tuition material being provided to help them cope with the mainstream courses that they are taking. In the BSc 4-year Foundation programme, students spend their first year doing a suite of non-credit bearing courses that have been designed specifically to help prepare them for their following year of study when they will be doing the mainstream courses that a normal first year student entering the faculty would have to do.

In order to be eligible for entry into either one of these two bridging programmes, students must have come from a historically disadvantaged school. In South Africa, schools have been grouped into quintiles based on their socio-economic background with a Quintile 1 school being the most disadvantaged. Prior to 2009 normal entry into the faculty was based on a Matric point score of at least 34 points. Those with a Matric point score of at least 28 Matric points were allowed to apply for entry into the BSc Augmented 4-year programme. Those with a Matric point score of at least 20 Matric points were allowed to apply for entry into the BSc Foundation programme. With the phasing out of the Senior certificate in 2008, entry into the Augmented 4-year programme was based on a total Matric point score of at least 22 points (excluding the Life orientation course) and entry into the Foundation programme was based on a total Matric point score of at least 16 points (excluding the Life orientation course).

This study followed the progress of 5976 students who registered for a degree in the Faculty of Science at UKZN over the period 2004 to 2012. Table 3 indicates that 2511 students gained entry through a bridging course with the remaining 3465 students being allowed a normal entry into the faculty.

Table 3: BSc student enrolment by year of first entry (*n*=5976).

Year of first entry	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total
Non-bridging	253	348	308	339	334	546	512	459	366	3465
4-year Foundation	0	0	0	203	204	195	254	232	277	1365
4-year Augmented	0	0	0	195	180	163	186	194	228	1146

Table 4 contains a listing of some possible confounding factors that we may want to adjust for when attempting to compute an average treatment effect for the bridging programmes that are being run in the Science faculty. The variables that we have labelled Male, African, Residence and Financial Aid all represent binary variables which have been set equal to one if the student is male, is of African origin, has been given some form of residence based accommodation or has been given some form of financial aid during their university based studies.

Prior to 2008, students writing their final school leaving subjects were able to do so at a higher, standard or lower grade level. From 2008 onwards, a National Senior Certificate was introduced where the previously graded levels for each subject were collapsed into a single level paper. To capture this effect in the analysis, the variable labelled Obe in Table 4 is an indicator variable that has been set equal to one if the student has matriculated post 2007.

Table 4: Student demographics based on enrolment figures in the Faculty of Science over the period 2004 to 2012.

Baseline covariates	Bridged Students	Non-bridged students
Male	1444	1771
Female	1067	1694
African	2469	1611
Non-African	42	1854
Residence	1533	865
No-Residence	978	2600
Financial Aid	1517	1333
No Financial Aid	994	2132
Obe	2113	2217
Non-Obe	398	1248

3.1. Matric point score

The results that students obtain for their final school leaving exams are often expressed in the form of a point score for each subject that makes use of the following method of scoring (Table 5).

Table 5: Point scores used for the final school leaving examination marks.

Point Score	Percentage mark
7	80-100
6	70-79
5	60-69
4	50-59
3	40-49
2	30-39
1	0-29

A total of seven subjects have to be written producing a total matric point score that forms an important basis for determining whether a person is eligible for entry into a higher education institution (HEI) or not. Those that fail to gain entry based on their total matric point score may be eligible for entry into a bridging programme if they have come from a historically disadvantaged school background. In South Africa, schools have been grouped into quintiles based on their socio-economic background with a Quintile 1 school being classified as the most disadvantaged. Prior to 2009 normal entry into the faculty was based on a Matric point score of at least 34 points. Those with a Matric point score of at least 28 Matric points were allowed to apply for entry into the BSc Augmented 4-year programme. Those with a Matric point score of at least 20 Matric points were allowed to apply for entry into the BSc Foundation programme. With the phasing out of the Senior certificate in 2008, entry into the Augmented 4-year programme was based on a total

Matric point score of at least 22 points (excluding the Life orientation course) and entry into the Foundation programme was based on a total Matric point score of at least 16 points (excluding the Life orientation course).

Table 6 shows the extent to which the Faculty of Science at UKZN has focussed on admitting students from the lower quintile schools into their bridging programmes. Given that many of these students, with enough bridging support, may eventually be able to complete their studies, it is important that one successfully separate the bridging programme effect from that of the other potentially confounding variables relating to their socio-economic background that may also be helping them to complete their studies at UKZN.

Table 6: Percentage of 5976 enrolments by school quintile.

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Not Bridged	3.25	3.60	7.46	10.65	30.10
Bridged	6.75	7.70	16.56	10.69	3.05

3.2. Response variable

One could consider using the total number of courses that have been failed for the first time as a response variable for this paper. A reviewer of this paper, however has correctly pointed out that some sort of correction will need to be made for the number of years that a student has spent studying for that degree. If one were to focus only on those students who have successfully graduated from their studies, because graduation requires the passing of a fixed number of courses, no such adjustment would be necessary but one would be throwing away a large number of observations from the dataset; namely all those students who have dropped out from their studies or who are still busy with their studies when the data collection process ended. For this reason

$$Y = \frac{\text{Total number of courses passed} - \text{Total number of courses failed}}{\text{Total number of years spent at university}}$$

has been used as a response variable for this paper. Essentially Y represents a per annum based 'rate of progress' with positive valued outcomes for this response variable indicating better performers. For example, a student wanting to finish a 3-year degree typically has to complete a total number of 48 courses whereas as a student enrolling for a 4-year degree typically has to complete a total of 64 courses. Thus, if either student wants to complete their degree in the minimum prescribed period of time (and with no other course failures) then they must record (for the 3-year degree) an outcome

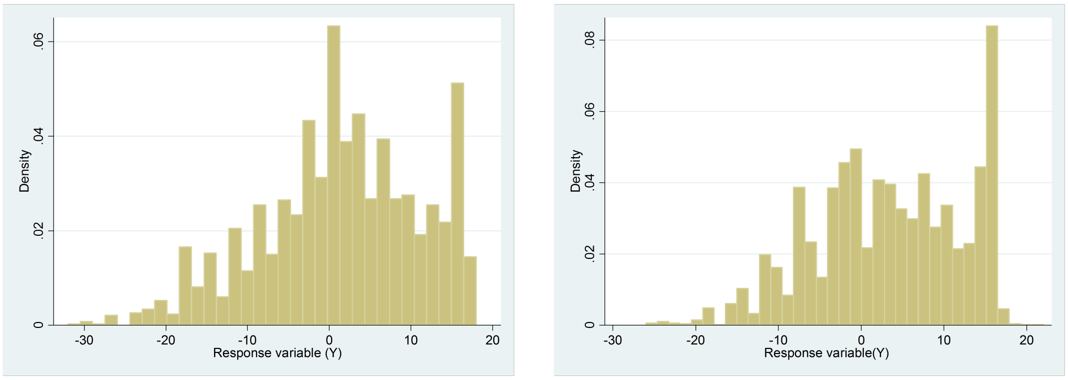
$$Y = \frac{48 - 0}{3} = 16$$

or (for the 4-year degree)

$$Y = \frac{64 - 0}{4} = 16$$

for this response variable.

Some students however may be allowed to register for more than 16 courses in a particular year, pass them all and thus on a year by year basis record a value for Y that is greater than 16 (Figures 1(a) and 1(b)).



(a) (b)

Figure 1: (a) Observed Y responses for students on the bridging programme. (b) Observed Y responses for students not on the bridging programme.

Figures 1(a), 1(b) and Table 7 present some results relating to this choice of response variable for the treated (bridged) and non-treated students. Because the estimated mean response for Y for the bridged students is lower than that for the non-bridged students, one may be inclined to conclude that the two bridging programmes are not actually helping those who come from a township school background to perform as well as their non-bridged counterparts. One should note however that in the context of this study, assignment to a bridging programme was done in a nonrandomized manner. Thus the treatment effect that one observes may no longer be attributable to the bridging programmes themselves but rather to a set of background variables that distinguish a student that is bridged from a student that is not bridged.

Table 7: Descriptive statistics relating to the number of courses that are being failed.

	Mean	Standard Deviation
Bridged students	1.368	9.735
Non-bridged students	3.321	9.001

3.3. Regression adjustment methods

Stata 13 was used to fit the regression model that has been given in Equation (2) to the observed dataset. To illustrate the importance of including all relevant confounders in one's fitted model structure, Table 8 presents some results that were obtained from excluding and then including the total matric point score as an extra variable in X. By not including the total matric point score, the treatment effect associated with bridging can be seen to change from being negative valued (implying that the bridging programmes are not actually helping to reduce throughput rates as measured by the chosen response variable) to becoming positive valued implying that the bridging programmes are actually helping to improve these throughput rates once an adjustment has been made for the confounding variables in that model structure.

Table 8: Parameter estimates obtained from fitting a regression adjustment model to Y.

Parameter	Model 1 estimate	Model 2 estimate
Bridged	-1.145*	2.817*
Obe	-0.534*	1.560*
Male	-0.631*	-0.291
African	-2.756*	-0.639
Residence	-0.676*	-0.683*
Financial aid	1.11e-04*	8.94e-05*
Matric Points	omitted	0.628*
Intercept	4.405*	-20.441*

* denotes significant at 5% level.

Table 9: Treatment effect estimates.

Treatment effect	Estimate	Robust standard error	95% Confidence Interval
ATE	1.032	0.478	[0.095, 1.969]
ATT	3.020	0.339	[2.354, 3.686]

Including all the variables that we have listed in Table 4 produced the estimated treatment effects that are given in Table 9. Providing that enough variables have been included in the model formulation so that a conditional independence assumption becomes valid, the positive effect that has been recorded for ATT suggests, amongst those students that have been bridged, that the bridging programme is actually helping to improve their throughput rates (as measured by our chosen response variable Y). Quadratic and interaction effects were also added to the model with a similar set of results being obtained.

Table 10: Parameter estimates obtained for the subpopulation of bridged students only.

Covariate	Parameter estimate	Standard error
Obe	2.121*	0.479
Matric Points	0.531*	0.057
Male	-0.437	0.386
African	4.012*	1.196
Residence	-0.653	0.446
Financial aid	7.7e-05*	9.7e-07
Intercept	-19.708*	2.171

* denotes significant at 5% level.

Table 10 contains the parameter estimates that result from fitting the model structure given in Equation (2) to the treated subpopulation of bridging students only. A significantly positive estimate for Matric Points suggests (as one would expect) that students with a higher matric point score perform better in this bridging programme than students with a lower matric point score. The positive estimate for Obe suggests that students matriculating under the new single grading system (that was introduced in 2008) appear to be failing less courses than those that have matriculated under the older three-tiered higher, lower and standard grading system. A possible explanation for this result could be that prior to 2008 students with potential would have been forced (in these township schools) to do mathematics and science at a standard or lower grade level which would then have prevented them from being able to gain access to a university based institution. Post 2008, the introduction of a single grading system has now allowed some of these students with potential the opportunity to gain entry into a university where with some extra bridging they can successfully complete their studies.

3.4. Propensity score based methods

Conducting a propensity score based analysis produced the average treatment effects that appear in Table 11. The first row of this table contains an ATT effect obtained using the propensity scores as inverted probability weights (Equation (8)). The second row in this table applies the doubly robust method given in Equation (11) whereas the last three rows of the table implement a propensity score match of each treated (bridged) student with their single nearest neighbour in the control group (NN=1), their two nearest neighbours in the control group (NN=2) and their four nearest neighbours in the control group (NN=4). Because each ATT effect is positive-valued, all these methods of matching indicate once again, but from a different perspective, that the two bridging programmes are indeed helping to reduce the number of courses that these same students would fail had they not been put onto a bridging programme. When using a propensity score, however, it is important to make sure that the support region for the treated and non-treated observations overlap with each other, a result that Figure 2 would seem to suggest is only partially being satisfied. It is for this reason that we have chosen to also include in our analysis Heckman's endogenous switching regression model because it helps to account for a possible bias that may arise because selection into treatment is also determined by additional covariates that we have not been able to observe.

Table 11: Average treatment effects based on the use of propensity score methods.

Method	ATT Estimate	Robust standard error	95% Confidence
IPW	4.169	0.846	[2.512, 5.827]
IPWRA	2.769	0.506	[1.778, 3.761]
NN=1 match	2.173	0.549	[1.097, 3.249]
NN=2 match	2.157	0.551	[1.077, 3.237]
NN=4 match	2.163	0.551	[1.083, 3.243]

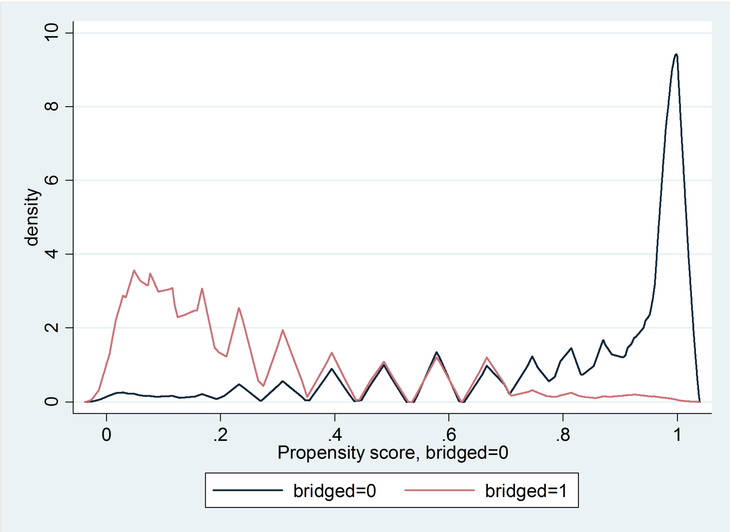


Figure 2: Density plots for the estimated propensity scores associated with the bridged and non-bridged student populations.

3.5. Heckman's switching regression model

Table 12 contains parameter estimates for the treatment assignment process that has been given in Equation (7). Predictor variables that were used for this treatment allocation process include an indicator variable denoting whether a student comes from a quintile 5 school (Quintile5=1) and an interaction term between a student's total matric point score and whether or not they have come from a quintile 5 school. The estimates that we have obtained fully reflect the emphasis that UKZN is placing on selecting students with a lower total matric point score and particularly those that also come from a disadvantaged quintile 5 school background. Table 13 contains a set of parameter

Table 12: Parameter estimates resulting from a probit fit using Equation (7).

	Parameter estimate	Standard error
Matric Points	-0.233*	0.006
Quintile5	1.144	0.636
Matric Points*Quintile5	-0.0745*	0.021
Intercept	7.259*	0.190

* denotes significant at 5% level.

estimates that result from fitting the outcome that has been given in Equation (6) to a very specific choice of covariates that have been included in that equation.

Once again the significantly positive estimate for ATT that appears in Table 14 suggests that the

Table 13: Parameter estimates resulting from a fit based on the model representation that is given in Equation (6).

	Parameter estimate	Standard error
Bridged	3.967*	0.703
Matric Points	0.639*	0.041
Male	-0.370	0.230
African	-0.729*	0.351
Residence	-0.736*	0.312
Financial aid	9.2e-05*	8.2e-06
Intercept	-20.052*	1.622

* denotes significant at 5% level.

Table 14: Average treatment effect for the Heckman selection model.

	Estimate	Robust standard error	95% Confidence Interval
ATT	3.967	0.703	[2.588,5.345]
ρ	-0.101	0.049	[-0.196,-0.004]

LR Test $\rho = 0$: Chi-square value 4.25; p-value 0.039.

bridging programmes are helping students to fail less courses than would normally be the case if they were not being bridged. The LR test statistic that appears in Table 14 fails to reject (at a 1% level of significance) a null hypothesis of independence between the error terms e and u that appear in Equation (8). Such a result would suggest that we have been able to include enough covariates in our earlier models to support the conditional independence assumption that is needed to justify the conclusions that we have been drawing in these earlier sections.

Table 15 contains a collection of different Heckman based model structures that have been fitted to our dataset. From a model selection point of view: whereas model fit in a linear regression setting can be assessed using a R^2 value, no such measure can be used to assess the fit of a Heckman model because of the probit model formulation that is being given in Equation (12). A likelihood ratio (LR) test however can be used to compare the performance of a baseline specification that contains a constant with that of a higher order model that also contains other predictors variables X_i in the specification that has been given in Equation (13). Model fit can also be based on the AIC and BIC values generated by each model structure. The model structure whose results appear in Tables 12 to 14 was chosen because it produced the smallest AIC and BIB value in Table 15.

Focussing on the ATT estimates that appear in Table 15 one can see, once again, that the two bridging programmes are helping students to perform better than they would if they had not been put on one of these two programmes.

Table 15: Average treatment effects for a collection of Heckman selection models.

Covariates (X) in model	ATT (std error)	ρ (std error)	AIC	BIC
Obe, Male,African, Residence, Financial aid, Matric points	4.381(0.701)	-0.124(0.049)	46511	46604
Male,African, Residence, Financial aid, Matric points	3.966(0.703)	-0.101(0.049)	46541	46598
Male,African, Financial aid, Matric points	3.989(0.703)	-0.106(0.049)	46545	46625
Male,African, Matric points	5.079(0.703)	-0.154(0.049)	46669	46743
African, Matric points	5.077(0.696)	-0.154(0.048)	46669	46735
Matric points	5.091(0.676)	-0.154(0.047)	46667	46726

4. Concluding remarks

In order to accept the results that are provided by a propensity score match or a regression adjustment procedure, the conditional independence assumption that has been given in Equation (3) needs to be verified. By fitting Heckman's switching regression model to our data, the LR test for independence that appears in Table 14 has been able to confirm (at a 1% level of significance) that a sufficient number of covariates have been controlled for when using a regression adjustment and/or a propensity score match to estimate the ATT. Because all the estimates for ATT that appear in Tables 9 and 11 are significantly positive, one can conclude that the two bridging programmes are indeed helping students from a disadvantaged school background to fail less courses than they would if they were not being put onto those bridging programmes. If one is more comfortable interpreting one's results using a 5% or 10% level of significance, then the results generated from the fitting of a Heckman model would also suggest that the two bridging programmes are helping students on these programmes to perform better than would be the case if they were not put on these programmes.

Because the two bridging programmes that operate in the Faculty of Science employ a different method of implementation, one may also want to consider conducting a separate analysis for those students who are put on the 4-year augmented programme and comparing these results with students who are being put on the 4-year foundation programme. UKZN also has bridging programmes in the College of Law and Management Studies that one may want to consider.

One could also consider changing one's choice of response variable Y to one where for example we consider graduation within the minimum prescribed period of time or even graduation within the minimum prescribed period of time plus one year. In these cases one would be fitting a binary response model to one's observed data.

Acknowledgement

The author would like to thank the referees for their very useful suggestions and ideas. Their input led to a substantial improvement being made to the paper.

References

ANGRIST, J. D. AND KRUEGER, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives*, **15** (4), 69–85.

BANG, H. AND ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, **61** (4), 962–973.

BRIGGS, D. (2004). Causal inference and the Heckman model. *The Journal of Educational and Behavioral Statistics*, **29** (4), 397–420.

FUNK, M. J. AND WESTREICH, D. (2008). Doubly robust estimation under realistic conditions of model misspecification. *Pharmacoepidemiology and Drug Safety*, **17** (S1), S106.

HECKMAN, J. (1979). Sample selection bias as a specification error. *Econometrica*, **47** (1), 153–161.

LUNCEFORD, J. K. AND DAVIDIAN, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, **23** (19), 2937–2960.

ROSENBAUM, P. R. AND RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, **79**, 516–524.

SARTORI, A. (2003). An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis*, **11**, 111–138.

SCOTT, I., NDEBELE, N., BADSHA, N., FIGAJI, B., GEVERS, W., AND PITYANS, B. (2013). A proposal for undergraduate curriculum reform in South Africa: The case for a flexible curriculum structure. Report of the Task Team on Undergraduate Curriculum Structure Discussion Document Council on Higher Education, Pretoria.

TERZA, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics*, **84**, 129–154.

Manuscript received, 2015-01-05, revised, 2015-03-31, accepted, 2015-05-04.