University Quality and Labour Market Outcomes of Canadian Youth*

(work in progress)

Joniada Milla†

March 21, 2012

Abstract

This paper estimates the wage returns to the Canadian university quality by making a comparison between the Maclean's magazine best overall ranking and a quality index constructed by Principal Component Analysis. The main data source is Youth in Transition Survey for the years 1998-2008 and the outcome of interest is the hourly wage rate of Canadian youth. Using OLS, matching methods and dose-response functions we draw some main results from this analysis. First, more low-ability students sort into high-quality universities than the reverse. Second, we find that returns to graduating from a middle-ranking university for women vary from 10% to 15%. Returns for men are statistically zero. Third, returns to university quality are positive for small improvements in ranking within the group of middle-ranking universities. However, they are zero for small improvements in ranking within the group of lowest- and highest-ranking universities.

JEL Classification: I23, J24, J31
Keywords: Returns to Education, University Quality, Wage Rates

*The empirical work of this paper was carried out using Statistics Canada microdata at the University of Guelph Branch Research Data Center (GBRDC) and the results have been approved for release. While the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada.

†Corresponding Author. Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. jmilla@uoguelph.ca, (001)519 824 4120 ext.56539, www.uoguelph.ca/~jmilla/

1 Introduction

In this paper we explore the role that the choice of university has in the level of hourly wages during the initial transition from schooling to labour market following graduation. University quality might be a factor accounting for the wage differential observed through years, yet it is a topic not well analysed for the Canadian universities. While there is a vast literature based on data from the US and Europe, because of the different institutional structure (Europe has a mainly public, tuition free university system, while in the USA there are mainly private universities and tuition fees vary a lot) most of the results may not be generalized to the Canadian case. To our best knowledge, there is only one study conducted on Canadian data, Betts et al. (2007). Further and updated research is needed using Canadian data. Another factor motivating us to analyse this topic is the fact that few Canadian datasets identify the major(s) as well as the institution(s). The data we use is the Youth in Transition Survey-Cohort B (YITS-B), which has information on the participants for the years 1998 to 2008 and is organized in five cycles. An individual's university is directly observed in this micro data. Detailed information is available on the major(s) and institution(s) that the student has ever attended and/or completed. Furthermore, there is information on all the jobs (wages, occupation, industry) the student worked within two years prior to and including the time of the interview of each cycle. We merge with YITS-B the university ranking and other university characteristic variables which we constructed from the yearly published information in the Maclean's magazine and the Canadian Association of University Teachers (CAUT) Almanac.

The literature has been struggling to decide on a single measure of university quality. Some papers use a couple of university characteristics like university rankings, professor/student ratio, professor salaries, etc. Noticing a high correlation among these variables, other papers have used factor analysis to combine them in one. Likewise, we use the principal component analysis[1] to combine 19 different university traits (see Table 9 for a list), which signal different attributes of the universities, into a single index as a measure of quality. These variables date the year in which most of the students graduate from high school and are in the process of applying for

[1]We propose the use of the stochastic dominance approach as a new way of building an index by choosing the characteristics that proxy best for the university quality in a future version of this paper

post secondary education.

Differently from the existing Canadian studies, the availability of a measure for ability in our data as well as a wealth of information on individual and family characteristics, allow us to assume that selection into universities of different quality is based on some observable variables, conditioning on which, sorting into universities happens randomly. In this way, we are able to identify a causal effect of the university quality on hourly wage rates four to six years post-graduation. We use least squares estimator, nearest neighbour and propensity score matching to estimate the quality effect among quality categories defined as low-ranking (bottom quartile in the university quality distribution), middle-ranking (inter quartile range in the quality distribution) and highest-or top-ranking universities (top or fourth quartile of the quality distribution). Finally, in a setting with more than two (high vs. low quality) treatment categories (i.e. continuous treatment variable) we estimate the dose response and treatment effect functions. The dose in our case is graduating from a university with a higher ranking in the hierarchy and the response is the treatment effect observed in the change of hourly wage rates.

Our findings indicate that university quality matters a lot for both genders when we do not control for high school grade point average (GPA), which in turn convey important information about the behaviour process. This is because of two reasons: higher ability individuals go to better schools and some of the observed wage premium these individuals get could be attributed to university quality when it is actually innate ability. Once we control for predetermined individual characteristics, among which high school GPA, we observe that the magnitude shrinks and the statistical significance vanishes for men mainly. We find that university quality returns for women vary from 10% to 15% to having a degree granted from a university falling in the second, third or fourth quartile of the university quality distribution as opposed to the first quartile (omitted category). The returns for men are of similar magnitude, even though not significantly different from zero. Our matching estimates of the return to graduating from a highest-ranking university yield insignificant returns for both genders. This result coincides with the least squares estimates. The reason we think is that the support condition is only marginally satisfied (for men in particular) as well as the choice of the number of matches in nearest neighbour matching and the choice of the bandwidth in the propensity score matching.

Both a high number of matches and high bandwidth lead to higher standard errors. Lastly, we estimate the dose-response function and the treatment effect function in order to analyse returns to small improvements in ranking rather than returns among categories. This technique uses all the observations in the sample rather than only the observations in the first and fourth quartile of the university quality distribution as nearest neighbour and propensity score matching do. The results from the dose-response function indicate that small improvements in ranking have a small but significant effect on wages for the quality index values within the group of middle-ranking universities. More specifically, there are positive returns to university quality for small improvements in ranking as the quality index values range between 3.5-standard-deviation-below and 2-standard-deviations-above the mean value of the university quality index. The returns to small improvements in ranking within the group of lowest-ranking and highest-ranking universities are zero.

The results obtained from this empirical work may be of practical use to the students and their parents as well as to universities. Knowing by how much do wages respond to graduating from a university with a higher ranking is important information in helping students make the right choice for their career. Also, the quality index hints on how should the universities administer and allocate their resources in order to reach for higher levels of ranking.

The analysis in the present paper may be extended further in several aspects. First, we plan to see whether the university quality affects other outcomes as well (other than the hourly wage). Alternative outcomes include probability to be employed, probability to drop out of university, probability to graduate and the probability to continue graduate school. Second, based on the argument by Black and Smith (2006) that ranking could change by field of study, the next step of this analysis is controlling for majors or (if sufficient number of observations) splitting the analysis by field of study.

The paper is organized as follows. Having introduced the topic in this section, we review the existing literature in Section 2 and discuss the data and methodology in Sections 3 and 4, respectively. We analyse the empirical results in Section 5 and conclude in Section 6. Some planned extensions to the present analysis are noted in Section 7.

2 Literature Review

A huge literature analyses the returns to education. Most of it is based on the Mincer (Mincer, 1958) earnings regression, which is a model that specifies logarithmic wages as a function of years of schooling and years of experience as displayed in equation (1) below.

$$log\,\omega_i = \alpha_0 + \alpha_1 S_i + \beta_1 E_i + \beta_2 E_i^2 + u_i \tag{1}$$

where S_i is the years of schooling, E_i is the years of experience and E_i^2 is the experience variable squared. The coefficient α_1 is interpreted as the return to schooling. Card (1999) makes a review of the contributions to this research area. He concentrates mainly on the papers that challenge two main implicit assumptions of the Mincer model: exogeneity of the years of schooling variable and the functional form. Firstly, the education variable in the above set up may capture other confounding effects of unobservable characteristics like the ability of the individual which we cannot measure and thus hides in the error term $u_i = \gamma A_i^* + \epsilon_i$ where A_i^* is a measure of the latent ability and ϵ_i is a independent error term. If there is not any way to control for A_i^* then $Cov(S_i, u_i) \neq 0$. Violation of this orthogonality assumption yields inconsistent estimates and unreliable hypothesis testing. Researchers have applied different methods to solve this problem. Some assume "selection on observables" and in that case the above equation takes the following form

$$log\,\omega_i = \alpha_0 + \alpha_1 S_i + \gamma A_i^* + X\beta + \epsilon_i$$

where A_i^* is a proxy measure of the latent ability (e.g.: high school grades, standardized test scores) and X includes all control variables (respondent's own background characteristics, experience and experience squared, family, friends and high school characteristics). In the information space of X and A_i^* the assumption $Cov(S_i, \epsilon_i) = 0$ holds and S_i is no longer endogenous in the empirical model. Several other papers, due to data unavailability, deal with selectivity on unobservable variables by instrumental variable techniques[2] to isolate the returns to education

[2]Starting with Card(1995) a number of papers assume "Selection on Unobservables" and use proximity to college as an instrumental variable (IV) for years of education completed. But that might be a weak instrument for the university quality. So, Long (2008) uses the average quality of the nearby colleges within a certain radius of the student as the instrument for the quality of the college at which the student attends. On how to construct this see pg.592.

on logarithmic wage from other confounding effects[3]. Secondly, the assumption of the linear functional form of the Mincer equation is likely not to hold. The effect of education for the years 8, 12, 16 (coinciding to the completion of elementary school, high school and college or university) on the wage rate might be nonlinear - this is commonly known as the “sheepskin effect”. Some non-linearities in those specific years of schooling might exist due to the fact that having completed a certain level of education and having obtained the diploma/certificate/degree documenting it, matters differently in the determination of a higher wage by the employee. What about the prestige of the institution that grants the degree? Will that induce an additional increase the wage rate of the employer beyond the education level attained? This is where the topic discussed in this paper fits in the labour literature. Hence the above equation becomes

$$log\,\omega_i = \alpha_0 + \alpha_1 S_i + \alpha_2 Q_i^* + \gamma A_i^* + X\beta + \epsilon_i \tag{2}$$

where Q_i^* indicates the latent university quality variable. Our purpose is to examine the returns to the quality of the university degree attained, thus estimating parameter α_2. The research dedicated to analysing the returns to university quality is extensive using US data, less so for European data and fairly new on Canadian data. Among the many relevant papers are Eliasson (2006); Chevalier and Conlon (2003); James et al. (1989); Brewer et al. (1998); Horstschraer; Suhonen (2011); Heckman et al. (2003). The prominent papers in the US literature are Black and Smith (2004, 2006); Black et al. (2005); Long (2008, 2010); Monks (2000); Dale and Krueger (2002). Black and Smith (2004, 2006) use NLSY the 1979 cohort and see the effect of the 4-year college quality on the hourly 1989 and 1998 wage rate. These two papers raise the issue of measurement error of the proxies used for the latent quality variable. They try to fix this issue by building a quality index using factor analysis and principal component analysis. Another way of dealing with measurement error is through instrumental variables. Black and Smith (2006) derive a GMM estimator which they prefer best as opposed to factor analysis because it makes direct use of the covariance matrix between the proxy variables. They find an average impact of 0.039 on the logarithmic hourly wage rate caused by one unit increase in the quality index. Black

[3] Instrumenting variables for S_i usually used are distance to school, education of the parents, the education of the partner/spouse.

and Smith (2004) in a matching framework, where the quality variable is a binomial indicator of attending a high quality[4] four-year college, find an impact of 12-14% increase in the log hourly wage rate. Long (2008) criticizes this method reasoning that the amount of the observations not used (pertaining to the inter-quartile range) is big which reduces the sample size a lot and thus the efficiency in estimation. The other critique is related to the fact that the "*estimates refer to discrete moves from one group of universities to the other and do[es] not allow the estimation of the effect of moving up the quality distribution within a group of colleges*" (Long, 2008, pg.594). In his 2010 paper Long (2010) looks into the trend of how the effect of years of education and four-year college quality changes over a period of 30 years (1970-2000) by using three different datasets that cover each of the three decades. He decomposes the analyses by gender and race and finds that the changes in the years of education and quality effects on a set of outcome variables are heterogenous among subgroups but mainly increasing through years for some of them. Black et al. (2005) also conduct a through-time analysis of the university quality on wage rates and find that it is quite stable during the time span 1987-1998 with men benefiting more than women (except in 1989). They also consider a few other labour market outcomes apart from the logarithmic hourly wage rate such as educational attainment, graduate school attendance, labour force participation, hours of work during the last year, marital status, number of children and spouse earnings.

Long (2010) enters years of schooling and four-year college-quality variables separately. Meanwhile, Black and Smith (2004, 2006) and Black et al. (2005) argue that if years of schooling, even though a controversial variable, is not included it might bias the results. They show this by presenting both the results with and without the years of schooling. In this version of the paper we do not include the years of schooling in our specification. We plan to add the estimation results with the years of schooling as a robustness check in the next updated version of the paper.

Holmlund (2009) summarizes mainly the methods used and results of the studies which use European data. In this paper the author contributes to the literature through a analysis using a very big Swedish dataset on individuals and university characteristics and by employing quartile

[4] A four-year college here is defined as being "high quality" if it falls in the fourth quartile of the distribution of the quality index built by principal component analysis as opposed to falling in the first quartile.

regressions. She finds that the returns to university quality are higher for the individuals who belong in the top quartiles.

A recent paper, Lang and Siniver (2011), questions the validity of assuming "selection on observable variables" and draws attention to not merely the difference in wages among graduates of universities with different status "elite" vs. not "elite" higher education institutions), but on the mechanisms that generates these differences. They find that when hired immediately after graduation, the students of the top university that they study are paid much higher wages than the graduates from the college that they study (both granting 4-year undergraduate degrees). This is because having no information on the skills of these students, they are considered as representatives of each institution. In this way the good student from the college is paid much less than he deserves, and the bad student from the university is paid much more than he deserves. However, with time, the employee "learns" about the skills and the good students from each institution have similar wages, and bad students have similar wages. So the "elite" factor vanishes after a couple of years of experience.

To our knowledge, the only paper that attempts to estimate alumni's wage rate returns to university traits in Canada is Betts et al. (2007). They use data from the National Graduates Survey and pool together three cross sections for the years 1982, 1986 and 1990. This dataset lacks a measure for the latent ability of the participants. In the absence of this important variable that could help in addressing the selection issue, the authors use a fixed effect model which "to the extent that the most able students in a province always attended universities A and B over the eight-year period under study,[...] sweep[s] *average* ability of the university's student body out of the wage equations" (Betts et al., 2007, pg.10). The results are interpreted as "something approaching a causal effect of [university] resources on student outcomes" (Betts et al., 2007, pg.10). The outcomes of interest are labour force participation and annual earnings five years after graduation.

Differently from Betts et al. (2007), our analysis is based on one cross section of data. There are a few strengths in the dataset that we use. The availability of a measure for ability as well as a wealth of information on individual and family characteristics, allow us to assume that selection into universities of different quality is based on some observable variables, conditioning on which, sorting into universities happens randomly. In this way, we are able to identify a causal effect of the university quality on hourly wage rates 4 to 6 years post-graduation. Apart for being a very recent dataset, YITS-B allows the identification of the universities and field of study attended, switched, and/or graduated, the occupation and the industry an individual has worked in. In this way, we could match the university one has attended to external (not within YITS-B) data on this university's characteristics. Our measure of university quality is based on 19 variables (see Table 9 for definitions). These variables date the year in which most of the students graduate from high school and are in the process of applying for PSE. The outcome variable we consider here is the log hourly wage rate as of December 2007. We chose to concentrate first on this outcome, as we aim to capture the returns of university quality to productivity rather than labour market participation. We may extend our analysis on considering the later as well as other outcomes on which university quality might have an impact.

3 Data

The main dataset that we use is the Youth in Transition Survey Cohort B (YITS-B). Students of age 18 to 20 in December 1999 were surveyed every two years until December 2007, and each survey asks questions that aim revealing information referring to the past two years from the date of the interview. Detailed information on the sample size, time of the interview, reference time and age of the participants can be found in Table 8. In the first wave of data the students were 18-20 years old and this time corresponds to the age range in which we expect most of them to have graduated from high school and enrolled in a PSE institution. By the third wave the age range is 22-24. By this age, we expect the students to have graduated from at least a PSE program and be in the job market. During the four years that the fourth and the fifth wave of the survey cover we expect most of the individuals to have started a full time job and to have settled in life. In the fifth wave the number of observations drops from 22,378 to only 9,934. Meanwhile the age range is 26-28 years old.

Our analysis uses data of the participants of YITS-B who have a Bachelor's degree or equivalent[5] as of December 2007. This subsample contains 2,520 observations, 59% (or 1,485) of which are women and 41% (or 1,035) are men. Because the individuals must have an overall post-secondary status "graduate, non-continuer" as of December 2007, the subsample shrinks further to 2,026 (60% or 1220 women and 40% or 806 men). Of the observations deleted, 494 were those people that graduated from BA program but are continuing another BA program or a post graduate program. Since they are still students, they are not counted in the labour force. Hence, 20.4% of all participants in cycle 5 of the survey have completed and attained one BA degree from a Canadian university and are part of the labour force. YITS-B was suitable for our purpose because it contains a wealth of information on the respondent, family, high school and friends, and especially detailed information about the PSE programs attended identifying the type of the degree, the type of the institution granting the degree, the code classification of this institution as well as the field of study.

Having these data available we could merge YITS-B with the university characteristics from external sources. Most of the data on university quality indicators are from the publicly available data in the university Ranking Issues of the Maclean's magazine published every year by the end of November. Maclean's publishes a overall ranking as well as the components used in the calculations to derive the overall rankings. The rankings were based on 24 indicators until 2006[6] collected from 47 universities across Canada. The detailed definitions and the source of every university quality indicator are provided in Table 9. This data was complemented with data from the CAUT Almanac and the Tuition and Living Accommodation Costs data set that Statistics Canada releases and it is the only data that contains tuition for each field of study by university in Canada.

[5] This number does not include the individuals who attained university diploma or certificate below Bachelor's (undergraduate level) and those who attained a first professional degree. Because their wage structure is different from a regular BA degree, they are excluded from the sample. Also the sample excludes those individuals that have attained a MA or a PHD degree. The reason for this is that most likely their wages will be higher when compared to any BA graduate thanks to their post-graduate degree. Including these might confound the university quality effect with that of a higher degree. We choose to drop these observations firstly and then include them back in the sample as a robustness check exercise. After controlling for the post-graduate degree attainment as well as the quality of the last university attended, it is interesting to see how our estimates change.

[6] In 2007, they changed the methodology to be based on third-party data only, and this necessitated reducing the number of indicators to 14. (Mary Dwyer Senior Editor (Universities), Maclean's)

4 Methodology

In this paper we split the analysis by gender because as Lefebvre and Merrigan (2010, pg.3) note "[...] males and females face different labour markets and occupy different types of jobs". We use two measures for university quality: the Maclean's Best Overall Ranking and the index we construct using the Principal Component Analysis (PCA). The Maclean's University Ranking variable is categorical ranging in value from 1 to 47, whereas our index is a continuous variable. Both these measures are constructed as a weighted average of the university characteristics; the only difference is the way the weights are assigned to each. The Maclean's Best Overall Ranking is constructed by weighing 10% Student Awards, 10% Student Faculty Ratio, 8% Awards per Full Time Faculty, 6% SSH Grants, 6% Medical Science Grants, 6% Total Research Dollars, 6% Operating Budget, 6.5% Scholarship & Bursaries, 6.5% Student Services, 5% Expenses, 5% Acquisitions, 4% Holdings per Student, 1% Total Library Holdings, 20% Reputation. Cramer and Page (2007) criticize the university ranking published by Maclean's as not realistic. The above individual university characteristics have a high correlation with each other, and so including all of them together in the model would introduce multicollinearity and would drive their coefficients to zero (Black and Smith, 2004). A way to avoid it is combining the individual characteristics into one index. A suitable technique to build such an index is Principal Component Analysis (PCA). PCA yields linear orthogonal combinations of the individual characteristics by assigning weights to each. These weights are determined by the solution of an optimization problem which seeks to maximize the extent that the index accounts for the covariance between the university characteristics. PCA may create as many orthogonal combinations, known as components, as there are inputs, in this case the university characteristics. Starting with the first component, the extent of covariance accountability that the component captures decreases in the second one and so on. Within a component the variable contributing most to the covariance is weighted highest. We use the first principal component (FPC) of the orthogonal transformation as our quality index. This is an efficient and optimal way of combining many university characteristics into one without worrying about the multicollinearity when these, otherwise, enter together in a regression equation.

As it is common in this literature, we start the analysis with the ordinary least squares

(OLS) estimates. We then see how the estimates are sensitive to different specifications as we control for more information. Next, we consider matching techniques to estimate returns to university quality: nearest neighbour matching and propensity score matching. Lastly, Long (2010) criticizes this method by saying that the binomial quality measure coupled with the matching methodology results in an efficiency loss due to the huge reduction of the sample size. In order to use all the data in our sample we estimate the dose-response and treatment effect functions.

For the part of the analysis where we use matching techniques for estimation, the treatment variable is binary, $H_i \in \{0, 1\}$, and defined as follows.

$$H_i = \begin{cases} 1 & \text{if } Q^* \geq 75^{th} \text{ percentile of the } pdf(Q^*) \\ 0 & \text{if } Q^* \leq 25^{th} \text{ percentile of the } pdf(Q^*) \end{cases}$$

In words, the treatment variable which is our university quality indicator, H_i, takes a value of one if the individual has graduated from a university that falls in the top (or fourth) quartile of the university quality distribution; we will call them high-quality universities. H_i takes a value of zero if the individual has graduated from a university in the bottom (or first) quartile of this distribution; we will call them low-quality universities. Further, let the outcome be $log\, \omega_i$ representing the log hourly wage of each individual at cycle 5 of our data by when the age range of the participants is 26-28 years of age. The wage rate post-dates graduation from a university program by four to six years. The potential outcome (log hourly wage rate), which is a different notion than the observed outcome, for each treatment state is

$$log\, \omega_i = \begin{cases} log\, \omega_{1i} & \text{if } H_i = 1 \\ log\, \omega_{0i} & \text{if } H_i = 0 \end{cases}$$

Our coefficient of interest is the average treatment effect on the treated (ATT) defined as

$$ATT = E\left[log\, \omega_{1i} - log\, \omega_{0i} \mid H_i = 1\right]$$

An alternative way of formulating ATT is:

$$ATT = E\left[log\,\omega_{1i} \mid H_i = 1\right] - E\left[log\,\omega_{0i} \mid H_i = 1\right]$$

So, ATT is the average log hourly wage difference between those that graduated from a high-quality university and the average log hourly wage that these same individuals would have had if they had graduated from a low-quality university. The later is unobservable because we can not observe one same individual in both states, and thus we can not see both potential outcomes of an individual in the treatment and non-treatment case. So, $E\left[log\,\omega_{0i} \mid H_i = 1\right]$ can not be observed; it is commonly known as the counterfactual. We can only estimate the counterfactual by $E\left[log\,\omega_{0i} \mid H_i = 0\right]$ and thus estimate ATT as the difference between the average outcome of the treated (high-quality university graduates) and of those who were not treated (low-quality university graduates). However, this is only possible at a cost. As shown in (Angrist and Pischke, 2008, pg.11) the equation below clearly displays this cost: the selection bias.

$$\begin{aligned}\underbrace{E\left[log\,\omega_i \mid H_i = 1\right] - E\left[log\,\omega_i \mid H_i = 0\right]}_{\text{Observed Difference in Average Outcome}} \quad = \quad & \underbrace{E\left[log\,\omega_{1i} \mid H_i = 1\right] - E\left[log\,\omega_{0i} \mid H_i = 1\right]}_{ATT} \\ & \qquad\qquad + \\ & \underbrace{E\left[log\,\omega_{0i} \mid H_i = 1\right] - E\left[log\,\omega_{0i} \mid H_i = 0\right]}_{\text{Selection Bias}}\end{aligned}$$

Selection bias derives from the fact that students with certain attributes and background self-select into university education, and moreover self-select into high-quality universities. To better see this, consider the bivariate distribution of ability measure (high school GPA) and university quality by quartile in Table 1. This way of showing selection is borrowed from (Black and Smith, 2004, pg.107). Table 1 displays only the entries corresponding to the first and fourth quartile, while the entries corresponding to “Total” are the column and row totals which include the observations in the omitted quartiles as well. In this table, for each cell the first number indicates the number of observations and the second number indicates the percentage of observations from the whole sample that falls in the cell. The bivariate distribution by quartile that would suggest an equal frequency in each cell determined by the quartile of the ability and quality measure

Table 1: Frequency by ability and university quality top and bottom quartiles

Quality index	Ability		
	First Quartile	Fourth Quartile	Total
Women			
First Quartile	77	34	236
	8.39	3.7	25.71
Fourth Quartile	51	47	229
	5.56	5.12	24.95
Total	244	166	918
	26.58	18.08	100
Men			
First Quartile	78	18	173
	13.47	3.11	29.88
Fourth Quartile	29	28	123
	5.01	4.84	21.24
Total	199	94	579
	34.37	16.23	100

Note: University quality measure here is the the PCA index.
The first number in each cell is the observation number.
The second number is the percentage of observations that fall in each cell.
The table excludes the corresponding numbers for the second and third quartiles.

would have about 6.25% of the observations in each cell. However, this is not the case for both genders. All cell percentages are different from 6.25% and the differences seem bigger for men. As in Black and Smith (2004), we notice asymmetry in sorting, but differently from them: In our data we see a higher portion of low-ability students in high-quality universities than high-ability students in low-quality universities. These numbers suggest that low-ability students target more the high-quality universities than the high-ability students do. One main point to take away from this table is that conditioning only on ability, the number of observations in each cell reflects a noticeable selection.

Self selection results in correlation between the potential outcomes and the treatment reflected as a difference between $E\left[log\,\omega_{0i} \mid H_i = 1\right]$ and $E\left[log\,\omega_{0i} \mid H_i = 0\right]$. Notice that if

$$E\left[log\,\omega_{0i} \mid H_i = 1\right] = E\left[log\,\omega_{0i} \mid H_i = 0\right] \tag{3}$$

then selection bias would be zero and ATT can be easily estimated as the observed difference in log hourly wages,

$$\underbrace{E\left[log\,\omega_i \mid H_i = 1\right] - E\left[log\,\omega_i \mid H_i = 0\right]}_{\text{Observed Difference in Average Outcome}} = \underbrace{E\left[log\,\omega_{1i} \mid H_i = 1\right] - E\left[log\,\omega_{0i} \mid H_i = 1\right]}_{ATT}$$

A solution to the selection problem is random assignment of students to universities of different quality. This would make the two groups (treated and untreated) comparable and make possible the calculation of the counterfactual as in equation (3). Random assignment can be guaranteed when the data are experimental and the researcher has direct control on assigning the treatment randomly. In the case of non-experimental data (e.g. survey data), researchers are able to assume that selection into universities is dependent only on some characteristics which can be observed or measured like family background, own attributes, past academic performance, etc. This is commonly known as the unconfoundedness or selection-on-observables or conditional independence assumption (CIA). In notation: $log\,\omega_h \perp H \mid X$ for all $H \in \{0, 1\}$. What this says is that treatment is assigned "as if randomly" after we condition on sufficient variables based on which the individuals self-select or are selected by the universities. Thus, even though before conditioning on X, a matrix containing predetermined characteristics of individual i, we most likely have

$$E\left[log\,\omega_{0i} \mid H_i = 1\right] \neq E\left[log\,\omega_{0i} \mid H_i = 0\right]$$

Under CIA, after conditioning on X we have,

$$E\left[log\,\omega_{0i} \mid X_i,\, H_i = 1\right] = E\left[log\,\omega_{0i} \mid X_i,\, H_i = 0\right]$$

So, we can easily estimate the average treatment effect on the treated as

$$ATT = E\left[log\,\omega_{1i} \mid X_i,\, H_i = 1\right] - E\left[log\,\omega_{0i} \mid X_i,\, H_i = 0\right]$$

Having stated this result, what the nearest neighbour (NN) matching method does is finding for each treated individual at least one untreated individual that has the same values of X as the treated individual and calculate the difference in their hourly earnings. After doing this for each treated individual, ATT is just the mean of all these differences. One issue with NN matching is what the literature refers to as "curse of dimensionality". The more variables you include in X, the more you guarantee that CIA holds, however as the number of these variables increases the bigger the number of cells defined by the values of X get and each cell of the multivariate distribution of the treatment and X becomes less and less populated and some of these cells are even empty. When this happens, the calculation of the counterfactual is not possible. Differently from NN matching, propensity score matching[7] (PSM) aiming to overcome the "curse of dimensionality" issue, calculates the counterfactual by matching the individuals on the probability of getting the treatment, known as the propensity score. In this way matching is done based on only one variable and it is less likely to have empty cells (shown by Rosenbaum and Rubin, 1983). For the PSM estimator, the CIA is represented as

$$log\, \omega_H \perp H \mid s(H, X) \text{ for all } H \in \{0, 1\}$$

where $s(H, X)$ is the propensity score and is defined as the conditional probability of receiving treatment H having certain pre-treatment characteristics X.

There are several advantages in using matching methods relative to least squares (OLS) regression. First, least squares regression assumes the causal effect of the treatment is constant for each individual, while matching techniques estimate this effect for each individual i in the sample, and report and average of these effects. Second, unlike OLS, matching disposes of the assumption that the relationship between the treatment and the outcome of interest is linear. Third, the balancing property in OLS is assumed, whereas in matching we can explicitly test for it (Rosenbaum and Rubin, 1983, see). For a technical and detailed description on the matching techniques see Rosenbaum and Rubin (1985); Abadie and Imbens (2006); Cochran and Rubin (1973); Dehejia and Wahba (1999); Heckman et al. (1997, 1998b,a,c); Imbens (2000); Lec (2001); Rosenbaum and Rubin (1983, 1985); Rubin (1974, 1980).

Building on Rosenbaum and Rubin (1983), Hirano and Imbens (2004) introduce the estimation of the propensity score in the case of a continuous treatment, named generalized propensity

[7] We use the *psmatch2* command in Stata of Leuven and Sianesi (2003).

score (GPS). They then calculate the dose response function and treatment effect function based on the GPS. In the case where the treatment is a continuous and normally distributed variable, say Q^*, the CIA in notation is

$$log\,\omega_{q^*} \perp I(Q^* = q^*) \mid s(q^*, X) \text{ for all } q^* \in Q^*$$

where $I(.)$ is an indicator function and $s(q^*, X)$ is the GPS. The estimation of the dose-response function is done in two steps. First, the conditional expectation of the outcome as a function of two scalar variables (university quality index and GPS level) is estimated,

$$\beta(q^*, s) = E\left[log\,\omega \mid Q^* = q^*, GPS = s\right]$$

Second, in order to estimate the dose-response function at each treatment level, the conditional expectation of the outcome estimated in step one is averaged over the score of GPS calculated at each particular level of the treatment, i.e.

$$\mu(q^*) = E\left[\beta(q^*, s(q^*, X))\right]$$

A version of this method was coded and provided for use as a package in STATA by Bia and Mattei (2008). For a detailed description of the functional form of $s(h, x)$, $\mu(q^*)$ and $\beta(q^*, s)$ see Hirano and Imbens (2004) and Bia and Mattei (2008).

5 Empirical Results

Table 2 below displays the mean value and the variance of the earnings per hour at the start and end of the present or most recent working position held, segregated by gender and age. Notice that the mean hourly wage of the women is higher than that of men when they are firstly hired, however, men seem to get higher wage increase than women during the tenure and thus end up having higher wages by the end of their tenure or when last in the job. The variation in the hourly wage rate per age group is modestly higher for women than for men. The descriptive

Table 2: Hourly wage rate by gender and age

Variable	Age	Mean	Std. Deviation	Obs. No.
Females				
Hourly wage at start job	26	23.117	31.415	376
	27	22.041	26.123	442
	28	23.640	31.673	400
Hourly wage at end job	26	26.236	31.194	376
	27	25.471	25.351	442
	28	27.971	31.512	400
Males				
Hourly wage at start job	26	22.199	27.359	279
	27	22.330	26.403	305
	28	21.743	20.217	218
Hourly wage at end job	26	27.065	28.164	279
	27	27.512	26.555	305
	28	28.683	21.541	218

statistics for the control variables and the university characteristics used in this paper can be found in the Appendix in Tables 11 and 12.

Figure 1 below provides a visual representation of the relationship between the log hourly wages (most current or when last in the position) and the university quality measured by the Maclean's ranking (blue line) and the index constructed as the first principal component (FPC) of the PCA (red line). These two equations are obtained by graphing the smoothed values from a kernel-weighted local mean regression using an Epanechnikov kernel[8]. The graph displays a clear positive and concave relationship between university quality measures and hourly wage rates. The increase is steeper as the quality variables increase up to their mean value (quality variables are standardized here with mean zero and unit standard deviation) and then starts to flatten as we go up to higher values of the ranking. Notice an increase in the slope of the FPC index plot (blue line) for the highest values of the index pertaining to the top ranking universities. We show later in the paper that as we condition on ability measure and other control variables this is not the case any more. The Maclean's magazine ranking shows a much steeper wage profile than our FPC index. This might be because it is a categorical variable

[8]Note: Stata command lpoly is used with degree of polynomial zero and rule of thumb bandwidth after both quality variables are standardized (mean zero, unit standard deviation). RDC vetting procedures do not allow for a scatter plot of the raw data to be released.

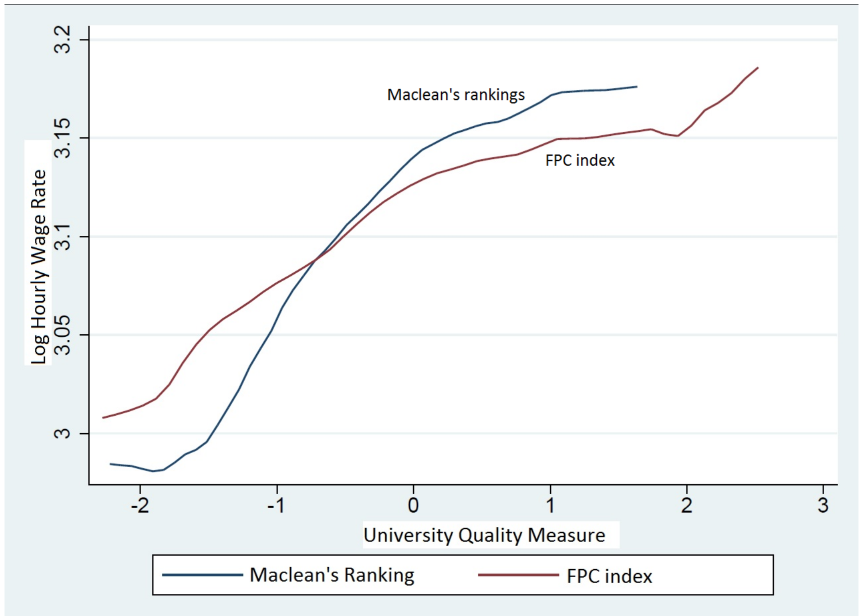

Figure 1: Both Maclean's and FPC index quality measure vs. hourly wages (lpoly)

(ranging from 1-47). Another reason might be related to the way the index is constructed: it might not properly reflect the quality because the weighting is almost equally assigned to each of the university characteristics.

Tables 3 and 4 show the ordinary least squares (OLS) estimates of the university quality effect on hourly wage rates by gender using the Maclean's Best Overall Ranking and the FPC index respectively. Each of the columns in these two tables represent a regression. As we go from the first to the fifth column we add more control variables so that we address the selection issue. The change in the magnitude and more so the change in the significance level of the estimated parameters reflects it.

The quality variable in this specification is composed of three dummy variables indicating the quartile of the university quality distribution. The omitted category is attending a university that falls in the first quartile of the distribution of the quality measure. There are several interesting results from these two tables. Column 1 in each table shows the results for the specification conditioning on only university quality. Returns to graduating from a university that belongs in higher quality categories changes in a non-monotonic fashion. except for men

Table 3: OLS estimates of return to university quality - Maclean's Ranking

Y=log(hourly wage)	(1)	(2)	(3)	(4)	(5)
Ability Measure		✓	✓	✓	✓
Individual Characteristics			✓	✓	✓
Parental Characteristics				✓	✓
Experience					✓
Women					
Second Quartile	0.108**	0.082*	0.075	0.079	0.106**
	(0.049)	(0.049)	(0.054)	(0.053)	(0.044)
Third Quartile	0.129***	0.102**	0.123**	0.115**	0.104**
	(0.046)	(0.046)	(0.053)	(0.053)	(0.044)
Fourth Quartile	0.077*	0.055	0.065	0.061	0.047
	(0.044)	(0.044)	(0.048)	(0.048)	(0.040)
Men					
Second Quartile	0.149**	0.117*	0.007	-0.001	0.058
	(0.063)	(0.063)	(0.073)	(0.073)	(0.069)
Third Quartile	0.141**	0.098*	0.066	0.059	0.086
	(0.059)	(0.059)	(0.067)	(0.068)	(0.062)
Fourth Quartile	0.155***	0.114**	0.001	-0.007	0.004
	(0.054)	(0.055)	(0.062)	(0.062)	(0.057)

Note: 918 obs for women and 579 for men. Standard errors in parenthesis.
***Significance at 1%, **Significance at 5%, *Significance at 10%

Table 4: OLS estimates of return to university quality - FPC index

Y=log(hourly wage)	(1)	(2)	(3)	(4)	(5)
Ability Measure		✓	✓	✓	✓
Individual Characteristics			✓	✓	✓
Parental Characteristics				✓	✓
Experience					✓
Women					
Second Quartile	0.076*	0.062	0.107*	0.111**	0.071
	(0.044)	(0.044)	(0.057)	(0.057)	(0.047)
Third Quartile	0.153***	0.138***	0.161***	0.152***	0.091**
	(0.048)	(0.047)	(0.053)	(0.053)	(0.044)
Fourth Quartile	0.054	0.037	0.078	0.068	0.032
	(0.048)	(0.047)	(0.055)	(0.055)	(0.046)
Men					
Second Quartile	0.108*	0.089	0.051	0.059	0.084
	(0.057)	(0.057)	(0.068)	(0.068)	(0.063)
Third Quartile	0.138**	0.094	-0.020	-0.021	-0.018
	(0.059)	(0.059)	(0.067)	(0.067)	(0.062)
Fourth Quartile	0.179***	0.129**	0.024	0.019	0.036
	(0.062)	(0.063)	(0.069)	(0.070)	(0.065)

Note: 918 obs for women and 579 for men. Standard errors in parenthesis.
***Significance at 1%, **Significance at 5%, *Significance at 10%

when using FPC index (second panel, Table 4). Notice that the returns to attending a university in the fourth quartile versus the first is higher for men than attending a university in the second or third quartile versus the first. This observation is reflected in Figure 1 as an increase in the slope past the value of 2 in the FPC index (red line). However, as we control for additional variables (columns (2) to (5) in Tables 3 and 4) in our OLS specifications we see that these results do not persist and that they were driven by non-random sorting. Our preferred specifications are column 4 and 5 in both Table 3 and 4. Since one might argue that post-graduation experience is not that important in early careers, we added the fifth specification to see how the results change in its presence. When using Maclean's Ranking as a measure for quality, for females the returns to university quality are on average about 10 to 12% for attending a university in the second or third quartile of the university quality distribution rather than in the first quartile. Returns to attending a university in the fourth quartile are not significantly different from zero. Using our FPC index (table 4), reveals higher returns (11-15%) for females when we do not control for experience, however lower returns (9%) when we control for post-graduate experience. Notice, however, that the returns to graduating from a university belonging to the third quartile of the quality distribution are robust to any of the five specifications considered in the case of females. These returns are statistically significant, positive and ranging between 10-15%. In both tables the wage rate of men seems to be unaffected by the quality of the university they graduate from. We think this might be due to the fact that the number of men in our sample is 35% smaller than that of women. Even though not explicitly obvious in the OLS specifications, the number of men that attained a degree in the first quartile might be insufficient to identify our parameters of interest.

Next we use matching techniques to estimate the return to university quality. In a matching framework we face a trade off between the balancing property and the support condition, or differently known as the "curse of dimensionality". So, we face a trade off between conditioning on many variables which increases the likelihood that Conditional Independence Assumption (CIA) holds, but on the other hand the more variables we condition on the fewer will be the number of individuals to compare in each cell and so the support condition is likely to fail. The estimates of returns to university quality on log hourly wage rate using nearest neighbour matching are displayed in the tables below for each of the quality measures, Maclean's Ranking in Table 5 and FPC quality index in Table 6. In this specification the treatment variable has

Table 5: NN matching estimates of returns to university quality on log hourly wage

Average Treatment Effect for the Treated - Maclean's Ranking	(1)	(2)	(3)	(4)
Ability Measure	✓	✓	✓	✓
Individual characteristics		✓	✓	✓
Parental Charactersitics			✓	✓
Bias correction				✓
Both	0.090***	0.034	0.061	-0.035
	(0.035)	(0.047)	(0.042)	(0.042)
Women	0.060	0.084	0.110**	0.027
	(0.044)	(0.055)	(0.051)	(0.051)
Men	0.123**	0.057	0.063	-0.039
	(0.057)	(0.068)	(0.066)	(0.069)

Note: 529 observations of women and 338 observations for men.
Number of matches is 4. Standard errors in parenthesis.
***Significance at 1%, **Significance at 5%, *Significance at 10%

Table 6: NN matching estimates of returns to university quality on log hourly wage

Average Treatment Effect for the Treated - FPC Index	(1)	(2)	(3)	(4)
Ability Measure	✓	✓	✓	✓
Individual characteristics		✓	✓	✓
Parental Charactersitics			✓	✓
Bias correction				✓
Both	0.098**	0.046	0.046	-0.169***
	(0.039)	(0.052)	(0.046)	(0.052)
Women	0.051	0.083	0.068	-0.932***
	(0.046)	(0.062)	(0.055)	(0.131)
Men	0.176	0.061	0.061	-0.051
	(0.069)	(0.077)	(0.074)	(0.082)

Note: 441 observations of women and 248 observations for men, 689 in total.
Number of matches is 4. Standard errors in parenthesis.
***Significance at 1%, **Significance at 5%, *Significance at 10%

to be dichotomous. In our case the treatment variable is the dummy variable taking a value of one if the individual graduated from a university falling in the fourth quartile of the quality distribution and zero if it falls in the first quartile. Thus we drop from this analysis all the observations belonging to the inter-quartile range and we are left with 529 observations for women and 338 for men. For each specification we match on four nearest observations when calculating the counterfactual. Notice that as we go from column (1) to (3) conditioning on more variables, the standard errors of each estimate increase. Even though the magnitudes of the return to quality is not too different from the OLS specification, due to much higher standard errors the estimates are not statistically different from zero. This might be the consequence of the support condition satisfied only modestly, (i.e. there are not enough observations in the data to build the counterfactual), or the number of matches (set to 4) introduces a lot of noise in estimates due to bad matches.

Hoping to overcome this issue, rather than conditioning on a multi-dimensional vector of control variables we use propensity score matching (PSM) that conditions only on the probability to get the treatment or propensity score, i.e. in our setting to have a degree from a high-quality university. PSM not only reduces the dimension of the variable based on which matching is done, it also releases the linearity assumption present in OLS. We enforce of common support condition and use Epanechnikov kernel with Silverman's rule of thumb bandwidth. The estimates using this method are displayed in Table 7 below. In all the cases (both genders, women and men) the

Table 7: PSM estimates of the university quality returns (Top vs. Bottom Quartile)

Average Treatment on the Treated			
	Women	Men	Both
Maclean's Ranking	0.006	0.006	-0.014
Standard Error	(0.071)	(0.065)	(0.046)
Observations	477	261	780
FPC Index	-0.106	0.023	-0.338
Standard Error	(0.84)	(0.090)	(0.098)
Observations	314	192	563

Note: Propensity score calculation uses Epanechnikov Kernel and Silverman's bandwidth.

balancing property is satisfied even though the histograms of the propensity score for the treated and untreated in Figures 2, 3, and 4 in the Appendix indicate that this is true onll marginally especially in the case of men. This causes high standard errors, which are comparable to those we obtained in the nearest neighbour matching estimates. Thus, no surprise that none of the estimates in Table 7 is significant. One other reason, might be the selection of the bandwidth. Here, the rule of thumb bandwidth is used. The bandwidth selection might yield better results if chosen by cross validation.

Thus, both nearest neighbour matching and propensity score matching confirm the results of OLS regarding the returns to graduating from a highest-ranking university versus a lowest ranking university[9] that they are statistically zero.

Next, we estimate the dose-response functions and treatment effect function which are analogous to propensity score matching but allow for a continuous treatment variable. In this part of our analysis, we do not split the sample by gender and use only our preferred measure of quality, the FPC index. The dose-response procedure conducts a balancing test which is satisfied in all specifications that we consider below. The dose in this case is a unit higher in the quality index, which indicates a higher ranking in the university that granted the degree. The estimated results are shown in Figures 5, 6, 7, 8 in the Appendix. The left panel in each figure is the graph of the dose response function, indicating the mean log earnings per hour per each quality measure value. The right panel in each figure is a graph of the treatment effect function, indicating the mean increase as we go up the hierarchy of university ranking. The top (green) and bottom (red) lines are the upper and lower bound of 95% confidence interval constructed by using bootstrap standard errors and accounting for the fact that GPS is an estimate and thus introduces noise in estimation. From Figure 5 to Figure 8 we add sets of conditioning variables (personal characteristics, family background, ability measure, experience) as noted in the title of each figure. In Figure 5 we see that the average expected log hourly wage rate increases with a steeper slope in the lower values of the quality index. The slope is less steep for values above the mean (of zero) and actually turns negative after a value of five. The dose response graph shows the same trend no matter on what variables we condition the calculation of the GPS. This is not true for the treatment effect functions. Notice that as we include more controls in the calculation of the GPS, the portion of the function whose confidence interval does not include zero becomes wider. Thus, accounting for more pre-treatment characteristics actually identifies some differences in the marginal increase in hourly wages per one standard deviation increase in the university quality index. For this reason let us concentrate mainly in Figure 8. The results from the treatment effect function indicate that small variability in ranking have a low

[9]A multivariate matching exercise is needed at this point in order to see whether the robust results that OLS reveals for the "Third quartile" quality returns are still valid when we use matching methods. We are working on producing those estimates.

but positive and statistically significant effect on wages for the quality index values belonging in the range of 3.5-standard-deviations-below and 2-standard-deviations-above the mean. The expected hourly earnings increase at an increasing rate as the university quality index increases up to the mean value within this range, but increase at a decreasing rate as the quality index increases away from the mean.

Thus, there are positive returns to university quality for small improvements in the ranking for middle-ranking universities. These returns are zero for small ranking improvements within the group of lowest-ranking and highest-ranking universities.

6 Conclusion

In this paper we estimate the returns to university quality in the post-graduation hourly wages of Canadian Youth of age 26 to 28 years old by making a comparison between the Maclean's Best Overall Rankings and a quality index constructed as the principal component of the Principal Component Analysis of university characteristics. The analysis is split by gender. Our main data source is Youth in Transition Survey covering the years 1998-2008. University characteristics data are constructed from the Maclean's magazine and CAUT Almanac publicly available data. Several main findings emerge from the analysis in this paper. Firstly, we observe asymmetric sorting in that more low-ability students sort into high-quality universities than high-ability students into low-quality universities. Second, our findings indicate that university quality matters a lot for both genders when we do not control for high school grade point average (GPA), which in turn convey important information about the behaviour process. This is because of two reasons: higher ability individuals go to better schools and some of the observed wage premium these individuals get could be attributed to university quality when it is actually innate ability. We find that university quality returns for women vary from 10% to 15% to having a degree granted from a university belonging to the second or third quality quartile, respectively, as opposed to the first quartile. The returns for men are of similar magnitude, even though not significantly different from zero. Third, our matching estimates yield insignificant returns for both genders regarding returns to graduating from a top quality university (fourth quartile quality distribution). This result coincides with the OLS estimates. The reasons might be due to the support condition only marginally satisfied in our data as well as the choice of the number of matches in nearest neighbour matching and the choice of the bandwidth in the propensity score matching. Lastly, we estimate the dose-response function and the treatment effect function in order to analyse returns to small improvements in ranking rather than returns among categories. This method makes use of all the observations in the sample rather than only the observations in the first and fourth quartile of the university quality distribution as nearest neighbour and propensity score matching do. The results indicate that returns to university quality for small improvements in ranking have positive returns as ranking improves within the group of middle-ranking universities. These returns are zero for small ranking improvements within the group of lowest-ranking and highest-ranking universities.

7 Further Research

This version of the paper is only a first draft. We are working on complementing further the analysis in several ways.

First, based on Dale and Krueger (2002) (who state that the quality index is unlikely to have only one dimension and that the university rankings are different depending on the field of study) and on Betts et al. (2007) (who find that controlling for field of major is important in reducing the bias in the estimates) we also intend to split the analysis by major or control for it in our regressions. Second, additional to personal and family characteristics we will add the school characteristics and peer effect variables as controls in our main specifications. This will increase the likelihood that the assumption of selection on observable characteristics holds. Third, the sample size in the OLS and matching results shrinks mainly due to listwise deletion. Black and Smith (2004); Black et al. (2005); Black and Smith (2006) avoid the potential selection bias created by listwise deletion in the control variables, as it might lead to non-random deletion of the entries, by recoding the missing values as zero and including a dummy variable indicating this. While in this version of the paper we do not correct for this, we plan to apply the same method and expect that this will increase the sample size. Fourth, based on Long

(2010) apart from ability, there might be other omitted variables in the Mincerian equation such as “ambition” and “taste for education”. We might use the aspiration variable in YITS, indicating the education level one “would like” to complete, and the responses to other relevant questions in the survey that tease out the taste for education. Moreover, Monks (2000) states that not only students select into universities but also universities select the students through the admission process. This selectivity is based on academic ability (like test scores, high school GPA) and ability to pay tuition (net family income). So, the authors include these to account for non-random selection into universities. While YITS-B does not have data on the family income, it has detailed information on how one is financing the PSE tuition and fees (amount from parental contribution, employment, bursaries and scholarships). We could use these variables to account for this other potential source of selection. Fifth, we are planning to explore different ways of constructing the university quality index. One might use stochastic dominance techniques. Sixth, using dynamic matching techniques we plan to incorporate in this analysis the next step after graduating from a university program, which is the choice between continuing graduate studies or labour market. Seventh, a last extension to this paper would be analysing the university quality effect in other variables such as the probability to be employed, the probability to graduate or drop out of university and the probability to pursue graduate studies.

Acknowledgements

I am thankful to Thanasis Stengos, Miana Plesca, Michael Hoy and Louis Christofides for their support, guidance and advice while working on this project. I thank Fraser Summerfield for helpful comments and suggestions. All errors and omissions are solely my own.

Appendix

Table 8: Timing of cycles for YITS - B

	Obs	Participants Age	Refence Time Period	Time of the Interview
Cycle 1	22,378	18-20	Jan1998-Dec1999	Jan2000-Apr2000
Cycle 2	18,779	20-22	Jan2000-Dec2001	Jan2002-Apr2002
Cycle 3	14,817	22-24	Jan2002-Dec2003	Jan2004-Apr2004
Cycle 4	12,435	24-26	Jan2004-Dec2005	Jan2006-Apr2006
Cycle 5	9,946	26-28	Jan2006-Dec2007	Jan2008-Apr2008

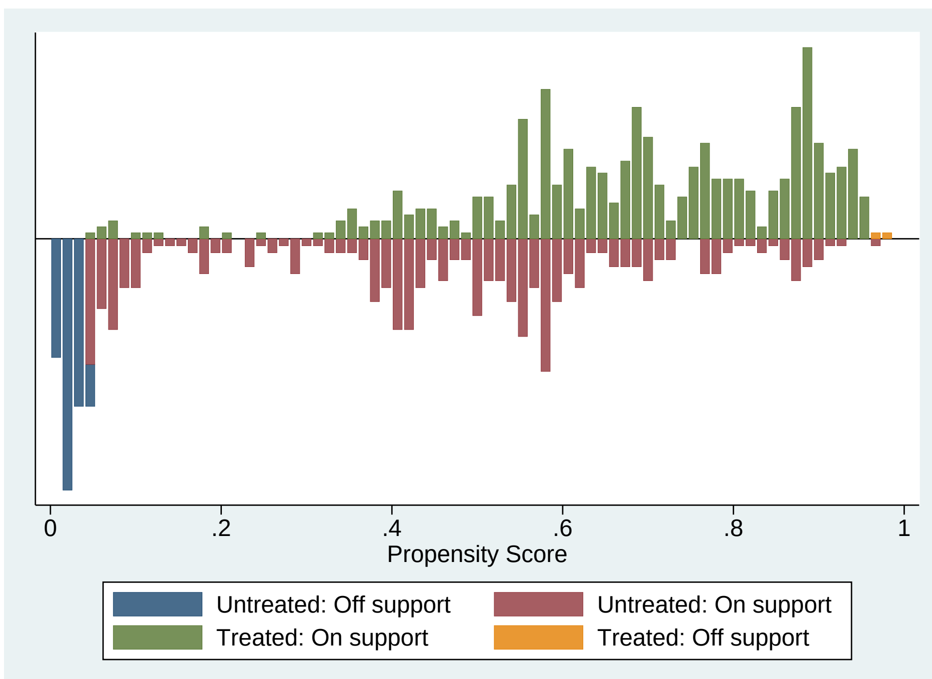

Figure 2: Propensity score histogram for both genders

Table 9: University Quality Variable Definitions

Year	Variable Name	Definition
Maclean's		
	Proportion who graduate	Percentage of full-time second-year undergraduates who completed their degree within one year of the expected graduation date.
	Classes Taught by Tenured Faculty	Percentage of first-year classes taught by tenured or tenure-track professors
	Faculty with PhDs:	Percentage of full-time faculty with a PhD degree
	Average Entering Grade	Students are enriched by the input of their peers. Here are the average final-year grades of freshman students entering from high school or Quebec's CEGEP system.
	Student Awards	The five-year tally of the number of students, per 1,000, who have won national awards.
	Faculty Awards	The five-year tally of the number of full-time professors, per 1,000, who have won national awards.
	Faculty Social Sciences and Humanities Grants	Below are the average size and number of peer-adjudicated research grants from both the Social Sciences and Humanities Research Council and the Canada Council. The size of grants is listed per eligible full-time faculty member; the number of grants is per 100 eligible full-time faculty members. The ranking reflects a weighted average of the two.
	Medical Science Grants	Here are the average size and number of peer-adjudicated research grants from both the Natural Sciences and Engineering Research Council and the Medical Research Council. The size of grants is listed per eligible full-time faculty member; the number of grants is per 100 eligible weighted average of the two.
	Operating Budget	These figures show the size of operating expenditures per weighted full-time-equivalent student
	Student Services	Percentage of total operating expenditures devoted to student services
	Scholarships & Bursaries	Percentage of total operating expenditures devoted to scholarships and bursaries
	Holdings per Student	These figures show the number of volumes in all number of volumes in all campus libraries, divided by the number of full-time-equivalent students.
	Acquisitions	To gauge the currency of resources, Maclean's measures the proportion of the library budget allocated to updating the university's collection.
	Expenses	A measure of financial commitment, this indicator shows the percentage of the university budget devoted to maintaining library services.
	Alumni Support	Maclean's measures the percentage of alumni who made gifts to the university over a five-year period.
	Value Added	Which universities get top marks for going the distance with their students? In this attempt to find an output measure, consulting statisticians from McDougall Scientific Ltd. juxtaposed two sets of figures. The first includes measures related to the incoming student: average entering grade and the percentage of the entering students with averages of 75 per cent or higher. The second examines two measures of student achievement: the proportion who graduate and student awards. Finally, the statisticians identified those schools with the greatest difference between the two figures.
	Founding Year	The year the university is founded.
CAUT Almanac		
	Tuition	Full time and Part time tuition and fees for each university
	Student Faculty Ratio	The ratio of the number of full-time tenured faculty members to the number of students enrolled in an university.

Table 10: Variable Definitions- YITS-B

Cycles	Variable Name	Definition
Dependent Variable		
1	Log hourly wage	Logarithmic hourly wage when last in the job, cycle 5, age 26-28, year 2008.
Personal Characteristics		
	Overall high school GPA	The overall high school grade point average (GPA).
	Age	Age of the respondent.
	Rural Dummy	Indicator of rural or urban geography of the most recent residence of the survey participant. This is derived based on the Statistical Area Classification (SATYPE) 2001 Census geography.
	Number of Children	Number of dependent children of the respondent.
	Citizen Dummy	Indicator variable takes the value 1 if the respondent is a Canadian citizen and 0 otherwise.
	Tenure	Number of months worked at last job during the reference time in each cycle.
	Experience	Number of months worked after graduation from a Bachelor's degree.
	Marital Status	A dummy variable is generated for each "married", "living with partner" and "separated, divorced or widowed". The omitted category is being single.
5	Residential Province Dummies	A dummy variable is generated as an indicator variable for each of the Canadian provinces. The omitted category is Ontario.
Parental Variables		
	Father Education Dummies	A dummy variable is generated for each of the education levels "High school", "College", "University and Professional Degrees", "Graduate Degree". Omitted category is "less than high school education".
	Mother Education Dummies	A dummy variable is generated for each of the education levels "High school", "College", "University and Professional Degrees", "Graduate Degree". Omitted category is "less than high school education".
Quality Measures		
	FPC Index	First Principal Component of the Principal Component Analysis (PCA) of the university characteristics shown in table 9
	Best Overall Ranking	This is an index built and reported by Maclean's magazine in Canada. The Maclean's best overall ranking is constructed by weighing 10% Student Awards, 10% Student Faculty Ratio, 8% Awards per Full Time Faculty, 6% SSH Grants, 6% Medical Science Grants, 6% Total Research Dollars, 6% Operating Budget, 6.5% Scholarship & Bursaries, 6.5% Student Services, 5% Expenses, 5% Acquisitions, 4% Holdings per Student, 1% Total Library Holdings, 20% Reputation.

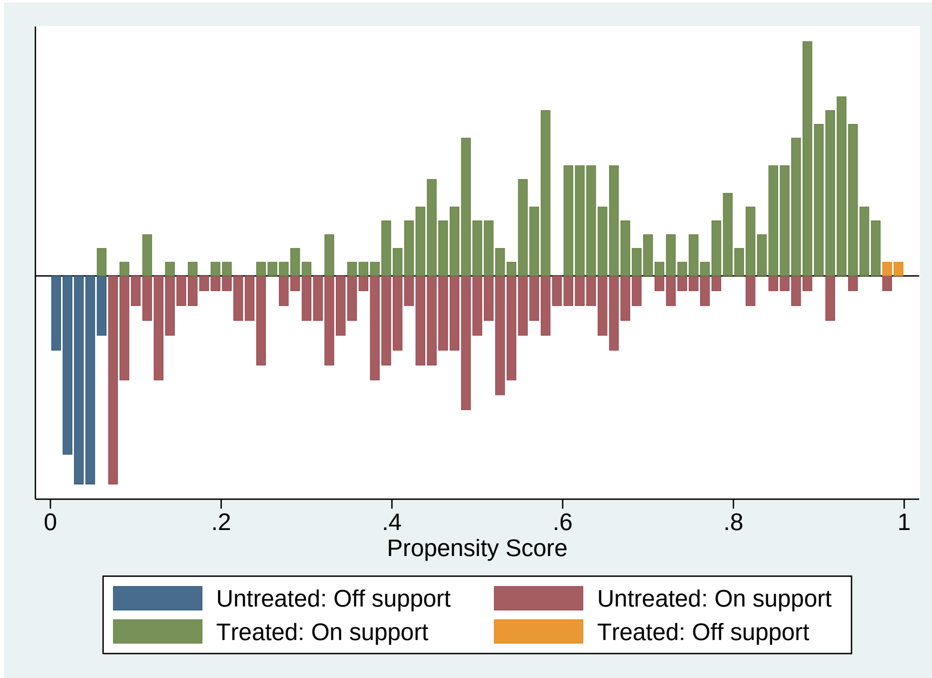

Figure 3: Propensity score histogram for women

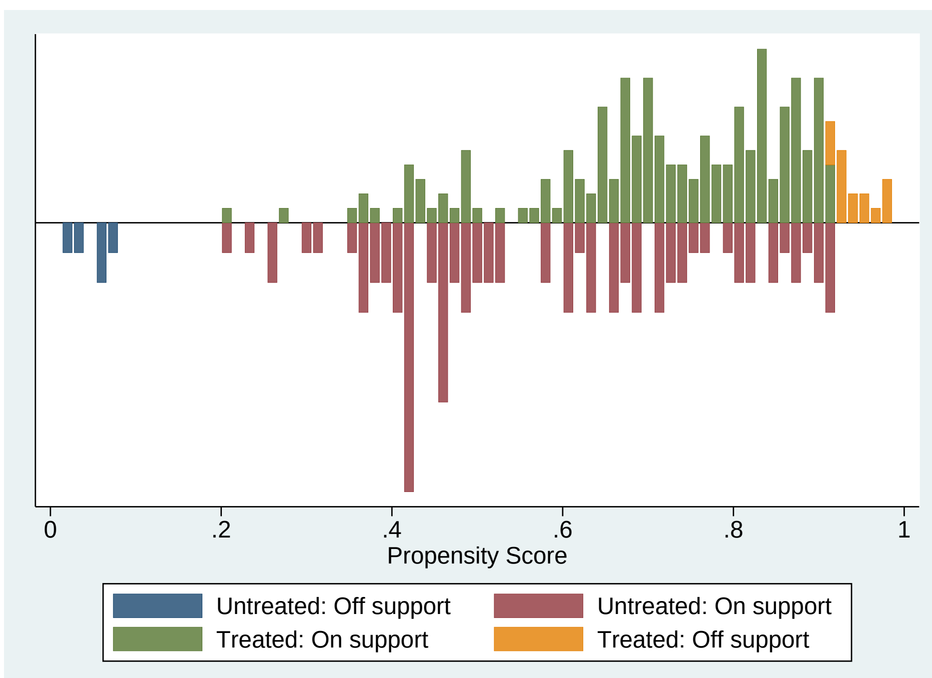

Figure 4: Propensity score histogram for men

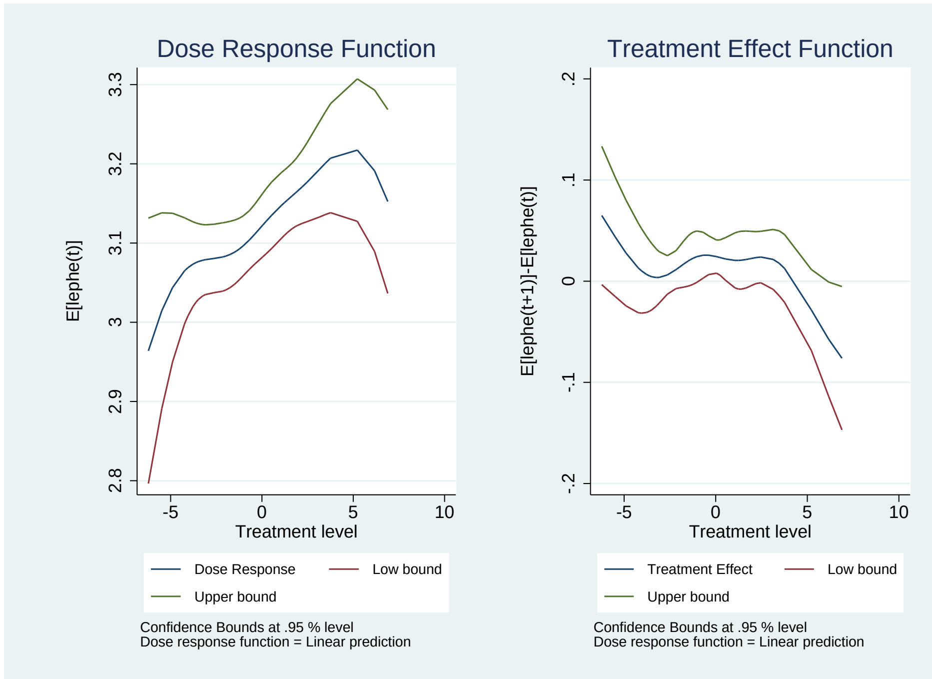

Figure 5: Both Genders: Conditioning on personal characteristics

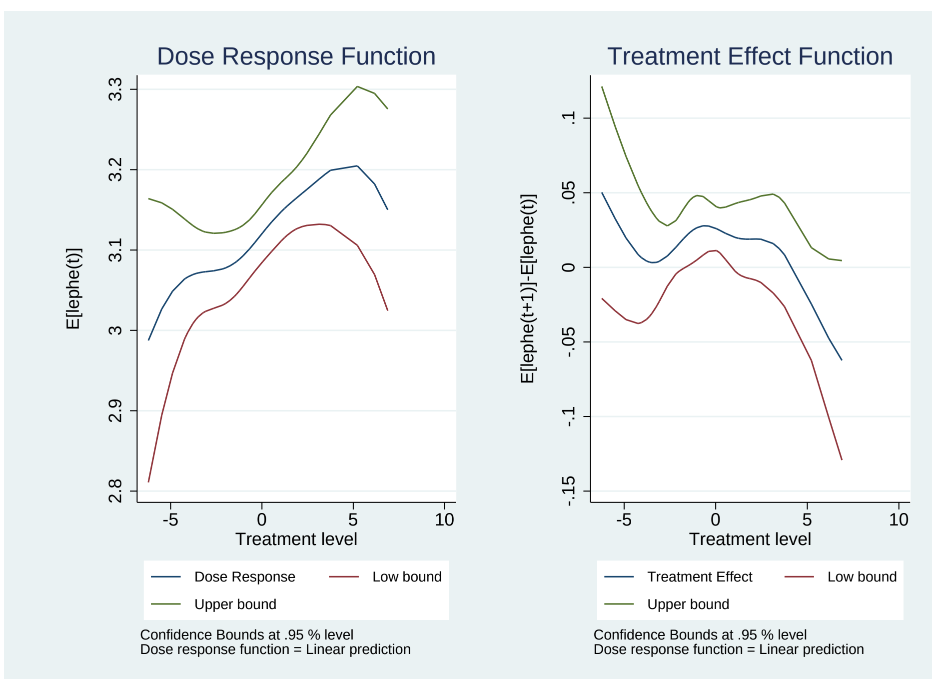

Figure 6: Both Genders: Conditioning on personal characteristics, family background

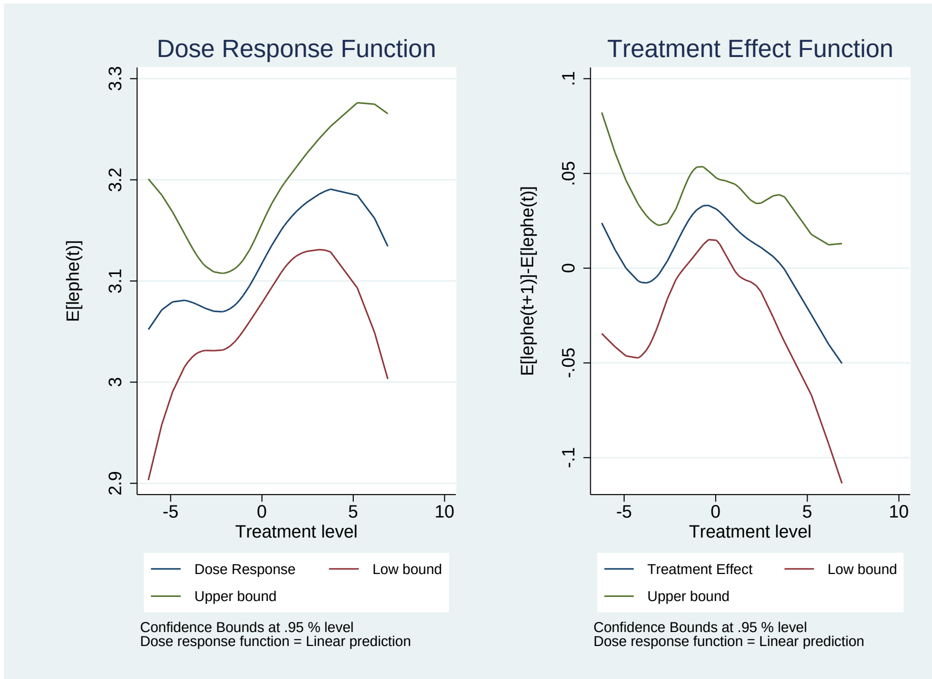

Figure 7: Both Genders: Conditioning on personal characteristics, family background, ability measure

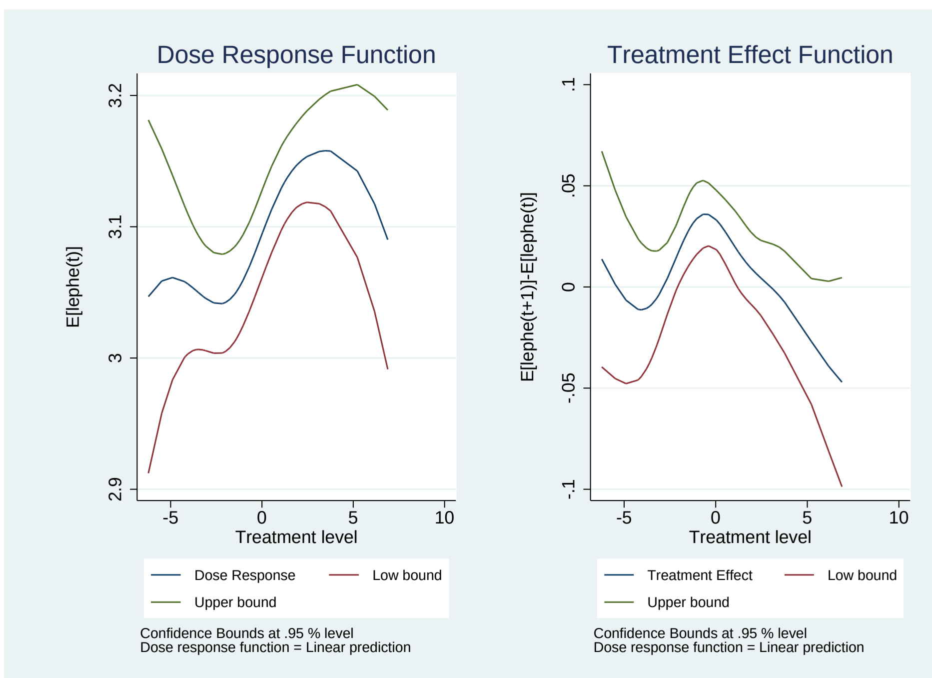

Figure 8: Both Genders: Conditioning on personal characteristics, family background, ability measure, experience

Table 11: Descriptive statistics of Controls

Variable	Mean	Standard Deviation	Obs. No.
Females			
Age (cycle 5)	27.020	0.798	1220
Rural Dummy (cycle 5)	0.172	0.377	1206
Number of dependent children (cycle 5)	0.251	0.593	1212
Citizen of Canada dummy (cycle 5)	0.962	0.191	1220
Visible minority dummy (cycle 1)	0.056	0.230	1218
Married dummy (cycle 5)	0.343	0.475	1220
Living with partner (cycle 5)	0.226	0.419	1220
Separated/Divorced/Widow dummy (cycle 5)	0.012	0.107	1212
Earning per hour start	22.880	29.657	1220
Earning per hour end	26.507	29.296	1220
Tenure cycle 2	24.874	14.740	1172
Tenure cycle 3	24.780	13.347	1191
Tenure cycle 4	25.629	12.149	1202
Tenure cycle 5	26.780	10.506	1196
Female parent college dummy	0.236	0.425	1220
Female parent BA or professional degree dummy	0.181	0.385	1220
Female parent MA or PhD dummy	0.041	0.198	1220
Male parent college dummy	0.191	0.393	1220
Male parent BA or professional degree dummy	0.211	0.409	1220
Male parent MA or PhD dummy	0.068	0.252	1220
Males			
Age (cycle 5)	26.923	0.784	806
Rural Dummy (cycle 5)	0.121	0.326	786
Number of dependent children (cycle 5)	0.133	0.428	798
Citizen of Canada dummy (cycle 5)	0.938	0.241	806
Visible minority dummy (cycle 1)	0.092	0.289	805
Married dummy (cycle 5)	0.244	0.430	806
Living with partner (cycle 5)	0.211	0.408	806
Separated/Divorced/Widow dummy (cycle 5)	0.006	0.079	797
Earning per hour at start job	22.076	25.150	806
Earning per hour at end job	27.651	25.825	806
Tenure cycle 2	22.210	13.461	770
Tenure cycle 3	22.224	13.021	772
Tenure cycle 4	24.388	12.061	789
Tenure cycle 5	26.734	10.704	800
Female parent college dummy	0.216	0.412	806
Female parent BA or professional degree dummy	0.248	0.432	806
Female parent MA or PhD dummy	0.076	0.265	806
Male parent college dummy	0.200	0.400	806
Male parent BA or professional degree dummy	0.259	0.439	806
Male parent MA or PhD dummy	0.109	0.312	806

Table 12: Descriptive Statistics of university characteristics

Variable	Mean	Standard Deviation	Obs. No.
Library Acquisitions 1999	38.990	5.346	1610
Library Expenses 1999	6.189	1.010	1610
Library Holdings 1999	216.213	76.913	1610
Operating Budget 1999	7886.222	7956.312	1610
SSHR number of grants per 100 full-time faculty 1999	14.222	7.491	1610
SSHR grants dollar amount of full-time faculty 1999	2682.480	1849.377	1610
Medical Science number of grants per 100 full-time faculty 1999	84.957	35.145	1587
Medical Science grants dollar amount of full-time faculty 1999	33095.880	18407.550	1587
Scholarships and bursaries 1999	5.007	1.806	1610
Student awards 1999	3.701	1.969	1610
Student faculty ratio 1999	0.199	0.045	1610
Student services 1999	4.377	1.599	1610
Average high school grade of entering cohort 1999	80.276	3.065	1610
Full-time tuition 1998	13659.900	7925.208	1610
Part-time tuition 1998	5620.965	5446.415	1610
Faculty awards 1998	2.831	2.206	1610
Faculty with PHD 1999	88.544	8.025	1610
Founding year of university	1898.367	72.754	1610
Percent Graduating 1999	77.181	8.364	1592
Percentage classes taught by tenures faculty 1999	58.074	12.109	1610
Tuition and Fees 1999	3497.553	915.487	1610
Value Added 1999	16.285	5.321	1534

Source: Author's calculations of the data from Maclean's Magazine and CAUT Almanac after having merged them with YITS-B.

References

Econometric Evaluation of Labour Market Policies, chapter Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption. Heidelber ca/Springer, 2001.

Alberto Abadie and Guido W. Imbens. Large sample properties of matching estimators for average treatment effects. *Econometrica*, 2006.

Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2008.

Julian Betts, Christofer Ferrall, and Ross Finnie. The role of university characteristics in determining post-graduation outcomes: Panel evidence from three recent canadian cohorts. Technical report, Analytical Studies, Research Paper Series, Statistics Canada, Catalogue no. 11F0019MIE, no.292, 2007.

Michela Bia and Alessandra Mattei. A stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, 2008.

Dan Black, Kermit Daniel, and Jeffrey Smith. College quality and the wages in the united states. *German Economic Review*, 6(3):415–443, 2005.

Dan A. Black and Jeffrey A. Smith. How robust is the evidence on the effects of college quality? evidence from matching. *Journal of Econometrics*, 121:99–124, 2004.

Dan A. Black and Jeffrey A. Smith. Estimating the returns to college quality with multiple proxies for quality. *Journal of Labour Economics*, 24 (3):701–728, 2006.

Dominic J. Brewer, Eric R. Ride, and Ronald G. Ehrenberg. Does it pay to attend an elite private college? *Journal of Human Resources*, 1998.

David Card. The causal effects of education on earnings. In O. Ashenfelter and David Card, editors, *Handbook of Labour Economics*, volume 3, chapter 30, pages 1801–1863. Elsevier Science B.V, 1999.

Arnaud Chevalier and Gavan Conlon. Does it pay to attend a prestigous university? *Center for Economics of Education, London School of Economics and Political Science, Houghton Street, London WC2A 2AE*, 2003.

W. Cochran and Donald B. Rubin. Controlling bias in observational studies. *Sankyha*, 1973.

Kenneth M. Cramer and Stewart Page. Cluster analysis and rankings of canadian universities: Misadventures with rank–based data and implications for the welfare of students. *Applied Multivariate Research*, 12(3):183–198, 2007.

Stacy Dale and Alan B. Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 2002.

R. H. Dehejia and S. Wahba. Causal effects in non-experimental studies: Re-evaluating the evaluation of training programmes. *Journal of the American Statistical Association*, 1999.

Kent Eliasson. The role of ability in estimating the returns to college choice: New swedish evidence. *Umea Economic Studies no. 691. Umea University.*, 2006.

James J. Heckman, H. Ichimura, and Petra E. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 1997.

James J. Heckman, H. Ichimura, J.A. Smith, and Petra E. Todd. Characterising selection bias using experimental data. *Econometrica*, 1998a.

James J. Heckman, H. Ichimura, and Petra E. Todd. Matching as an econometric evaluation estimator. *Review of Economic Studies*, 1998b.

James J. Heckman, R. J. LaLonde, and J. A. Smith. The economics and econometrics of active labour market programmes. In O. Ashenfelter and David Card, editors, *The Handbook of Labour Economics*, volume 3A. 1998c.

James J. Heckman, Lance J. Lochner, and Petra E. Todd. Fifty years of mincer earnings regressions. Technical report, National Bureau of Economic Research, Inc., 2003.

Keisuke Hirano and Guido Imbens. The propensity score with continuous treatments. In *Missing Data and Bayesian Methods in Practice: Contributions by Donald Rubin's Statistical Family.* Wiley, 2004.

Linda Holmlund. The effect of college quality on earnings: Evidence from sweden. *Department of Economics, Umea University. Mimeo*, 2009.

Julia Horstschraer. Quality signals in action? the importance of rankings and excellence labels for university choice of high-ability students. ZEW, Manuscript.

Guido W. Imbens. The role of propensity score in estimating dose-response functions. *Biometrika*, 2000.

Estelle James, Nabeel Alsalam, Josef C. Conaty, and Duc-Le To. College quality and future earnings: Where should you send your child to college? *AEA Papers and Proceedings, Economics of Eduucation Industry*, 1989.

Kevin Lang and Erez Siniver. Why is an elite undergraduate education valuable ? evidence from israel. *Labour Economics*, 2011.

Pierre Lefebvre and Philip Merrigan. Labour outcomes of graduates and dropouts of high school and post-secondary education: Evidence for canadian 24- to 26 -year-olds in 2005. *Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi Working Paper 10-45*, 2010.

Edwin Leuven and Barbara Sianesi. Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *"http://ideas.repec.org/c/boc/bocode/s432001.html".This version 4.0.4 10nov2010 E. Leuven, B. Sianesi.*, 2003.

Mark C. Long. College quality and early adult outcomes. *Economics of Education Review*, 2008.

Mark C. Long. Changes in the returns to education and college quality. *Economics of Education Review*, 2010.

Jacob Mincer. Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302, 1958.

James Monks. Thereturns to individual and college characteristics, evidence from the national longitudinal survey of youth. *Economics of Education Review*, 2000.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983.

Paul R. Rosenbaum and Donald B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 1985.

Donald B. Rubin. Estimating causal effects of treatments in randomised and non-randomised studies. *Journal of Educational Psychology*, 1974.

Donald B. Rubin. Bias reduction using mahalanobis-metric matching. *Biometrics*, 1980.

Tuomo Suhonen. *Does University Quality Affect Studend's Subsequant Earnings? Evidence from Finland Using Field-of-Study-Specific Quality Measures.* PhD thesis, School of Business and Economics, University of Jyvvaskyla, Finland, 2011.