UNB
EST. 1785
UNIVERSITY OF NEW BRUNSWICK

NB-IRDT
New Brunswick Institute for Research, Data and Training

USING ADMINISTRATIVE AND SURVEY DATA TO ESTIMATE RETURNS TO HIGHER EDUCATION IN CANADA

Ted McDonald, PhD
Bethany Daigle, MA
Pablo Miah, MA, MSc

PROJECT TITLE

Using administrative and survey date to estimate returns to higher education in Canada

PRINCIPAL INVESTIGATORS

Dr. Ted McDonald
Director, NB-IRDT
Professor, Department of Economics, University of New Brunswick

RESEARCH TEAM

Bethany Daigle, Research Assistant, NB-IRDT
Pablo Miah, Data Analyst, NB-IRDT

PARTNERS

This work was supported by funding from the Labour Market Information Council

HOW TO CITE THIS REPORT

McDonald, T., Daigle, B., & Miah, P. (2019). Using administrative and survey data to estimate returns to higher education in Canada. Fredericton, NB: New Brunswick Institute for Research, Data and Training.

TABLE OF CONTENTS

Abstract 3
Executive summary 4
Introduction 6
Literature review 9
Survey data 9
Census data 14
Administrative data 18
Currently and potentially available datasets 20
PSIS / RAIS 20
LAD / IMDB 21
CENSUS 22
LISA 22
GSS 24
PIAAC 25
Recommendations 27
Immediate recommendations 27
Short-term recommendations 28
Intermediate-term recommendations 30
Long-term recommendations 30
Summary and discussion 32
References 33
Appendix 36

ABSTRACT

In November 2018, Statistics Canada released The Education and Labour Market Longitudinal Platform (ELMLP), a program that enables researchers to link three administrative datasets: The Postsecondary Student Information System (PSIS), the Registered Apprenticeship Information System (RAIS), and the T1 Family File tax records (T1FF) (LMIC, 2018).[1] The ELMLP provides a new opportunity for researchers to compile more comprehensive information on Canadians' transitions from postsecondary education (PSE) into the labour market; however, the platform faces some limitations, including a lack of information on earnings comparison groups. We conduct a survey of the existing literature and currently available datasets and attempt to address limitations by identifying and isolating the most accurate and effective methods for calculating returns to PSE.

We recommend obtaining control groups from the datasets already linked through the ELMLP and creating aggregated cohorts of earnings profiles based on observable characteristics across multiple datasets. We also suggest using administrative tax records to estimate individuals' participation (or non-participation) in higher education. Finally, we propose expanding the information currently accessible through the ELMLP with additional datasets, including the Census, LAD, LISA, GSS, PIAAC, IMDB, and others. Over time, linking these datasets to the ELMLP would allow researchers to generate estimates on the earnings of high school graduates, high school non-completers, and PSE graduates who earned their credentials abroad—all cohorts which are currently not covered by the ELMLP. Accessing this and other missing information will enable the development of more applicable control groups and thus the production of more accurate estimations of the return to PSE in Canada.

[1] The ELMLP can be accessed through Statistics Canada's Research Data Centres. To learn more about the initial release of this platform, see LMIC (2018), found at https://lmic-cimt.ca/wp-content/uploads/2018/12/LMI-Insights-No.-4.pdf.

EXECUTIVE SUMMARY

Education is often described as a valuable asset – for the individuals who attain it and for society at large. Existing literature backs up this assumption and almost exclusively finds positive returns to investment in postsecondary education (PSE). One study even claims that "[t]he strong positive relationship between educational attainment and labour market earnings is a finding so consistent and universal across developed countries as to constitute a stylized fact" (Boothby & Drewes, 2006, p. 3). The widespread acceptance of this 'fact' is evident in the increasingly large investments that students, governments, and other institutions continue to make in PSE. However, while research confirms positive returns to these investments, there is no consensus on the most reliable methods for calculating return. Yet, while the majority of existing reports implement different sources of data, control groups, and systems of measurement, researchers are moving toward more aggregate approaches.

Echoing the work of Finnie and Usher (2005) and Frenette et al. (2014), Statistics Canada recently compiled and released The Education and Labour Market Longitudinal Platform (ELMLP), a program linking three administrative datasets: The Postsecondary Student Information System (PSIS), the Registered Apprenticeship Information System (RAIS), and the T1 Family File tax records (T1FF) (LMIC, 2018).[2] The ELMLP effectively promotes the combination of multiple datasets as the most comprehensive way to compile information on Canadians' transitions from higher education into the labour market. However, the platform faces certain limitations, including a lack of earnings comparison groups.

By analyzing the existing literature and currently available datasets, we compile a list of recommendations we identify as the most accurate and effective methods for addressing this limitation and calculating returns to PSE; we classify these as immediate, short-term, intermediate-term, and long-term suggestions. In particular, we propose analyzing additional datasets to generate estimates on the earnings of high school graduates, high school non-completers, and PSE graduates who earned their credentials abroad—all cohorts which are currently not covered by the ELMLP, but which could be used in the development of earnings comparison groups.

Our immediate recommendations include using the datasets already linked through the ELMLP to define control groups of individuals who pursued but did not complete PSE, as well as drawing on observable characteristics across datasets (such as age, sex, region, etc.) to define aggregate cohorts of earnings profiles. These can then be used as comparison groups in a synthetic cohort analysis following the approach of Green and Worswick (2012). Our primary short-term recommendation involves isolating individuals without evidence of higher education in administrative tax records to obtain control groups of individuals whose highest level of education is secondary school or less. Finally, in the intermediate- and long-term, we also suggest expanding the information currently accessible through the ELMLP with additional

[2] The ELMLP can be accessed through Statistics Canada's Research Data Centres. To learn more about the initial release of this platform, see LMIC (2018), found at https://lmic-cimt.ca/wp-content/uploads/2018/12/LMI-Insights-No.-4.pdf.

datasets, some of which have already been linked with tax records; some that are accessible but have not yet been linked; and some that Statistics Canada may eventually collect and link.

We argue that incorporating data from the Canadian Census, LAD, LISA, GSS, PIAAC, IMDB, and other databases would provide researchers access to information missing from the ELMLP, enabling the formation of more applicable control groups and thus producing more accurate estimations of the return to PSE in Canada.

INTRODUCTION

Education is often presented as one of the most valuable things a person can attain. Typically, parents want a good education for their children, and governments want a well-educated populace. This is evident in the large investments that individuals, governments, and other institutions continue to make in postsecondary education (PSE). For individuals, the completion of PSE can lead to increased earnings, a wider range of employment opportunities, reduced unemployment, further education, higher social status, and, in general, the prolonged enjoyment of learning (Boudarbat, Lemieux, & Riddell, 2010; Stager, 1996). When a population is more highly educated, society itself benefits from greater levels of technological progress, a more informed electorate, increased government tax revenue, and higher quality leadership in business and politics, as well as reduced crime and improvements in health and longevity (Boudarbat et al., 2010; Dee, 2004; Stager, 1996; Stark, 2007).

In Canada, the federal government has consistently described a more educated workforce as "essential to Canada's global economic competitiveness" (Kirby, 2007, p. 5), and provincial governments have echoed this sentiment. A report on PSE in four Canadian provinces[3] identifies higher education as a valuable tool for "individual growth and fulfillment" and "social and economic progress," while further stating that the new economy "demands" higher enrolment in PSE (Kirby, 2007, p. 8). As expected, PSE enrolment has increased, with the number of Canadians (aged 20-24) in postsecondary institutions growing from 28% in 1990 to 41% in 2005. By 2007, Canada had the highest percentage of PSE attainment of all 30 OECD[4] countries, and nearly half the country's working-age population had completed either college, vocational school, or university (Kirby, 2007).

However, as with any investment, PSE can be costly—for individual students, tax-payers, and the government. Students face expenditures on tuition fees and books, as well as net earnings foregone while not participating in the labour force. Meanwhile, the larger economy is faced with the costs of government grants, tax exemptions, expenditures made by universities, and the value of the output forgone while its students are outside the labour force (Stager, 1996).

Given the large investment that PSE requires, the expected return to the completion of PSE is a key factor in students' enrolment decisions and government policies. Of course, financial return is not the only element influencing decisions to pursue or fund PSE. In some cases, it may be of little import. Yet, as suggested by Canadian reviews, economic globalization is leading people to view PSE through a more utilitarian and market-oriented lens (Kirby, 2007). Decisions whether to pursue PSE are increasingly intertwined with expectations of financial return. For policy-makers, information on the monetary returns associated with PSE is integral to spending procedures (Vaillancourt, 1995). Finally, PSE continues to grow more expensive over time. The average tuition fee for undergraduates rose by over 367% from 1990 to 2018—more than 5 times the rate of inflation over the same period. These growing costs are associated with a significant rise in the number of student loan borrowers, the average amount of student loan debts, and

[3] Alberta, British Columbia, Newfoundland and Labrador, and Ontario
[4] Organization for Economic Co-operation and Development

incidents of student loan repayment difficulties (Kirby, 2007). In the face of steeply rising costs, it is increasingly important that individuals contemplating PSE enrolment have an idea of the kinds of returns they can expect from their investment.

Fortunately, there is no shortage of literature on the returns to PSE. Studies overwhelmingly show a strong correlation between PSE completion and benefits such as earnings gains, increased human capital, and higher private, fiscal, and social rates of return. Boothby & Drewes (2006) even claim that "[t]he strong positive relationship between educational attainment and labour market earnings is a finding so consistent and universal across developed countries as to constitute a stylized fact" (p. 3).

However, although most studies reach a similarly positive conclusion,[5] there is a great deal of variation in the sources of data and methods used to measure correlation. Some reports base their findings on survey data,[6] whereas others use census data.[7] Some supplement their calculations with administrative data or other secondary sources,[8] and others rely solely on one primary dataset.[9] Likewise, the standard against which returns are measured varies between reports.[10] Finally, each study employs its own formulas for calculation, controls for different variables, and produces findings for different study periods and labour market experiences.

Following Finnie and Usher (2005) and Frenette, Grekou, and Wannell (2014), researchers are increasingly recognizing the value of combining sources and methods of research in producing more comprehensive outcomes (LMIC, 2018). As a result, in November 2018, Statistics Canada released The Education and Labour Market Longitudinal Platform (ELMLP), a program that enables researchers to link three administrative datasets: The Postsecondary Student Information

---

[5] While the existing literature shows positive returns to the completion of PSE, some studies find that rates of return decrease in size as individuals pursue increasingly higher levels of PSE (Dodge & Stager, 1972; Vaillancourt, 1995).

[6] Bar-Or et al. (1995); Brand & Xie (2010); Caponi & Plesca (2009); Dee (2004); Dodge & Stager (1972); Ferrer & Menendez (2009); Lemieux (2006); Vaillancourt (1986); Walters et al. (2004).

[7] Autor (2014); Boothby & Drewes (2006); Boothby & Drewes (2010); Boothby & Rowe (2002); Boudarbat et al. (2010); Ferrer & Riddell (2002); Stager (1996); Stark (2007); Vaillancourt (1995).

[8] Brand & Xie (2010); Dee (2004); Finnie & Usher (2005); Lemieux (2006); Stager (1996); Stark (2007).

[9] Autor (2014); Bar-Or et al. (1995); Boothby & Drewes (2006); Boothby & Drewes (2010); Boothby & Rowe (2002); Boudarbat et al. (2010); Caponi & Plesca (2009); Dodge & Stager (1972); Ferrer & Riddell (2002); Handa & Skolnik (1975); Vaillancourt (1986); Vaillancourt (1995); Walters et al. (2004).

[10] For example, different studies compare the earnings of

- university graduates and high school graduates (Bar-Or et al., 1995; Boothby & Rowe, 2002; Boudarbat et al., 2010; Stager, 1996)
- college graduates and high school graduates (Autor, 2014; Brand & Xie, 2010)
- PSE graduates by field/level of study and high school graduates (Boothby & Drewes, 2006; Caponi & Plesca, 2009; Stark, 2007; Walters et al., 2004)
- PSE graduates alone by field/level of study (Boothby & Drewes, 2010; Dodge & Stager, 1972; Handa & Skolnik, 1975)
- or PSE graduates, those with some PSE, and those with no PSE (Ferrer & Riddell, 2002; Vaillancourt, 1986).

System (PSIS), the Registered Apprenticeship Information System (RAIS), and the T1 Family File tax records (T1FF).[11]

The ELMLP provides a new opportunity for researchers to compile more accurate information on Canadians' transitions from postsecondary education to the labour market; however, the platform faces some limitations, including insufficient information to create multiple earnings comparison groups. We conduct a survey of the existing literature and currently available datasets and attempt to address the limitations facing the platform by identifying and isolating the most accurate and effective methods for calculating returns to PSE.

Our immediate recommendations include obtaining control groups from the datasets already linked through the ELMLP and creating aggregated cohorts of earnings profiles based on observable characteristics across multiple datasets. We also suggest using administrative tax records (such as the LAD) to estimate individuals' participation (or non-participation) in higher education, as this could be accomplished over a relatively short period of time.

Finally, we propose expanding the information currently accessible through the ELMLP with additional datasets, including the Census, LAD, LISA, GSS, PIAAC, IMDB, and others. While linking these datasets to the ELMLP would be a lengthier process, the outcome would allow researchers to generate estimates on the earnings of high school graduates, high school non-completers, and PSE graduates who earned their credentials abroad—all cohorts which are currently not covered by the ELMLP. Accessing this and other missing information will enable the development of more applicable control groups and thus the production of more accurate estimations of the return to PSE in Canada.

[11] The ELMLP can be accessed through Statistics Canada's Research Data Centres. To learn more about the initial release of this platform, see LMIC (2018): https://lmic-cimt.ca/wp-content/uploads/2018/12/LMI-Insights-No.-4.pdf.

LITERATURE REVIEW

Finnie and Usher (2005) describe an ideal source of data for measuring the quality of PSE. This Canadian database links individuals' beginning characteristics, institutional inputs, learning outputs, and final graduate outcomes (p. 29). However, at the time of their report, no such database existed. Years later, Frenette et al. (2014) made a similar argument, suggesting that the linkage of survey and administrative data from Statistics Canada and Human Resources and Skills Development Canada (HRSDC) would improve research on the labour market outcomes of adult learning. In 2018, these combined sentiments came to realization through the implementation of the ELMLP. However, despite the new opportunities this platform provides, it still faces problems and limitations. We review the literature on the return to PSE to strategically point out potential limitations that could impact the effectiveness of the ELMLP, and we offer recommendations for improvement.

In reviewing the existing data, this report surveys both Canadian and international literature. While we prioritize Canadian studies because of their commentary on Canadian databases, we find it beneficial to look at the international literature as a means of fostering the improvement of Canadian data. For instance, Canadian researchers Caponi and Plesca (2009) describe American ability-related measures such as the Armed Force Qualifying Test and the Scholastic Assessment Test as "ideal candidates" (p. 1102) for their study; and although Canada did not have a similar dataset at the time, such recognition of effective data sources elsewhere is helpful in directing the types of data Canadian institutions should aim to accumulate going forward.

When calculating returns to PSE, the data used in national and international studies can be divided into three categories: survey data, census data, and administrative data. Empirical studies on the return to PSE typically employ at least one of these classes of data, although none of the literature reviewed to date incorporates all three.

SURVEY DATA

Survey data includes information collected from a sample of the population, rather than from every member of the population. Although survey data may not represent the population as thoroughly or accurately as census data, it nonetheless presents some benefits for researchers. Due to its smaller sample size, survey data is much easier and less expensive to collect than census data, and surveys can be administered and updated much more frequently as a result.

Because surveys can be produced in greater number, they are also more likely to address topics not included in census collections, such as family background information allowing researchers to better control for selection bias (Caponi & Plesca, 2009; Ferrer & Menendez, 2009). This advantage is accompanied by a problem, however, as the great variety in survey topics can also result in inconsistency (Bar-Or et al., 1995; Boudarbat et al., 2010; Caponi & Plesca, 2009), making it difficult to compare survey data over a large study period.

In a fairly early study, Dodge and Stager (1972) challenge the hypothesis that returns to PSE were high during the 1960s. In their cost-benefit analysis, the authors calculate the net present value of investment in PSE according to costs of graduate education and PSE graduates' earnings differentials every year until retirement (at age 64).[12] They use data from the 1967 Highly Qualified Manpower Survey to estimate the year of graduation at each level of PSE (the mean current age minus the number of years employed), as well as the earnings of scientists, engineers, and MBAs. The estimated educational cost per student is taken from the government of Ontario's formula for university grants in 1966-67. Dodge and Stager compare the estimated earnings premiums of PSE graduates working in public and private sectors, as well as the returns to graduate education in science, engineering, and business; undergraduate education in Ontario; and graduate education abroad.

Vaillancourt (1986) uses micro-data from Statistics Canada's 1981 Survey of Consumer Finance and Financial Statistics of Education to estimate the future earnings of males without university schooling and those with at least some PSE. The author notes a serious limitation in the data at the time, as the 1981 datasets do not provide standardization of schooling characteristics, such as field and length of studies and degrees held. Moreover, Vaillancourt suggests that the best approach to estimating lifetime age-earnings profiles would be to use individual panel data collected over the working lives of individuals; however, in 1986, no such data was available in Canada. Thus, using cross-sectional data and OLS estimates to calculate age-earnings profiles for men who were 18 years old in 1981, Vaillancourt's study assumes the resultant profiles remain stable in real terms over time and finds that private and public rates of return to investment in PSE range from 7-14% and 6-10%, respectively.

Bar-Or, Burbidge, Magee, and Robb (1995) use micro data from the Canadian Survey of Consumer Finances (SCF) to calculate the changes in return to a university education between 1971 and 1991. Controlling for years of work experience and comparing the full-time earnings premiums of a university education to that of 11-13 years of education, Bar-Or et al. find that the declining returns during the 1970s did not rebound over the study period. However, they acknowledge that their findings could be misleading due to limitations in their data source, control group, and methodology. For instance, they find a lack of consistency: data from the 1970s relates to families, whereas data from the following decades relates to individuals.

Yet, Bar-Or et al. still find the SCF to be a more accurate source than the Canadian Labour Force Survey (LFC). Unlike SCF, the latter does not provide information on earnings or income for the survey year. Nonetheless, although Bar-Or et al. propose a control group of high school graduates, SCF data makes it difficult for the authors to identify high school graduates "cleanly," leading to the formation of an alternate control group of "those who have completed 11-13 years of schooling with no postsecondary education" (p. 766). Finally, while they show that their results do not depend on "(1) ratios of means or ratios of medians, (2) weighted or unweighted data, or (3) grouped data or standard regression techniques" (p. 787), Bar-Or et al. conclude that results can be volatile across years, and so there is a need to deduce trends from longer periods of time.

[12] Dodge & Stager (1972) also calculate the depreciation rates for university land, buildings, and equipment, as well as property taxes, local realty rates, and the value of the schools' physical assets (p. 184).

Walters, White, and Maxim (2004) look at the influence PSE has on the earnings, employment outcomes, and school-to-work transitions of Aboriginals with postsecondary credentials in Canada. Using data from the 1995 National Graduates Survey (NGS) (which surveyed class of '95 graduates again in 1997), Walters et al. compare the labour market outcomes of Aboriginal, visible minority, and other Canadian graduates from university, college, and trades school. However, despite identifying the NGS as "the largest and most comprehensive survey available in Canada addressing issues relating to the labour market experiences of recent postsecondary graduates of various programs" (p. 287), the authors fault the database for its lack of distinction between status and non-status First Nations, as well as its failure to address characteristics regarding health, culture, and ability.

Finally, their study attempts to fill the gaps in previous works by distinguishing between university, college, and trades graduates. It also deviates from previous research reliant on frequencies and cross tabulations and instead uses regression in analyses that control for background characteristics—particularly socio-demographic control variables such as age, marital status, region, language, number of children, and parental education. Nonetheless, they argue the need for further study to compare outcomes of PSE graduates with those who have only some postsecondary schooling, or those without any PSE, to "better contextualize the employment outcomes of Aboriginal university graduates" (p. 296).

Caponi and Plesca (2009) measure the return to PSE relative to a control group of high school graduates. They also attempt to solve limitations they associate with the works of Ferrer and Riddell (2002) and Boothby and Drewes (2006)[13] by controlling for unobserved characteristics, specifically innate ability, in selecting higher education. The authors opt to use data from the 1994 Canadian General Social Survey (GSS) rather than the Canadian Census because the former includes family background information that can be used to fill a gap in ability measures in Canadian data.

Although the data from GSS cycles can face limitations depending on the yearly topic of focus, Caponi and Plesca argue that the few cycles on Education, Work, and Retirement contain more information than other datasets on individual and family characteristics, including educational and working familial histories. Using data on parental education to control for selection bias, the authors determine that propensity score matching is a much more reliable estimator than the IV estimator. They find that despite evidence of positive ability selection into all levels of PSE, the difference in returns to university, college, and trades education only decreases slightly after controlling for ability selection (with a gap of 20% remaining between men with university education and men with college or trades education).[14] Because ability selection accounts for such a small proportion of the differences in return by level of education, Caponi and Plesca propose, for a future project, a full estimation of a structural model that accounts for the present value, costs, and benefits of human capital accumulation.

[13] Ferrer and Riddell (2002) and Boothby and Drewes (2006) are discussed in detail on pages 15-16.
[14] For a thorough discussion of the formulas and methodologies used and considered by this report, see Caponi and Plesca (2009), pp. 1108-1113.

Ferrer and Menendez (2009) respond to an assumption common to most analyses of returns to PSE: the assumption that schooling proceeds in "a linear and uninterrupted fashion from primary school to the highest level of education the individual attains in her lifetime" (p. 5), which they find convenient but not accurate. The authors therefore compare the returns to PSE graduates who delayed their PSE to the returns to non-delayers. They use data from the 1997 Canadian National Survey of Graduates and the 2000 Follow-up Survey, which include a large sample of less traditional educational choices, as well as information on degrees obtained in 1995, pre-enrollment and post-graduation activities (including data on jobs held and wages), and additional education obtained after graduation. In short, this study compares 1997 log wages between single-degree continuing graduates, single-degree delayed graduates, multiple-degree continuing graduates, and multiple-degree delayed graduates—all from the same 1995 graduate cohort—to estimate levels of return for delayers and non-delayers.

Lemieux (2014) considers three potential explanations for the positive returns to PSE, as well as differences between returns to various workers with PSE: 1) education makes workers more productive; 2) education helps workers achieve higher-paying jobs in which output relies more heavily on skill; and 3) workers are more productive and earn higher premiums when their jobs relate to their field of study. Lemieux uses data from the 2005 NGS and the 2006 Canadian Census, as both include information on occupation and field of study. However, he acknowledges, the NGS has the advantage of measuring "relatedness" between subjects' jobs and their education, but it is limited in forming a control group of high school graduates because it does not include information on individuals without PSE.

Meanwhile, the Census contains more consistent, representative information about the earnings of the Canadian population, but it does not consider the "relatedness" variable between jobs and fields of study. To measure relatedness for high school graduates – which are missing from the NGS – Lemieux (2014) proposes constructing a control group either by using a dummy variable of 1 (for high relatedness) or 0 (for low relatedness) or by adopting the low level of relatedness seen in the humanities (51.4%) as indicative of the general skills acquired in high school. Notably, Lemieux does not control for ability in this study but suggests a future study in which additional factors (including ability) are considered.

AMERICAN STUDIES

Grubb (1993) uses the National Longitudinal Survey of the class of 1972 (NLS72), and follow-up surveys until 1986, to measure returns to PSE. Grubb identifies NLS72 data as "superior to most other data for examining returns to postsecondary education" (p. 366) due to the large amount of information it contains on individuals' characteristics. However, it faces limitations in containing data on only one cohort of students during the 1970s. Moreover, because PSE transcript data between 1979-86 is incomplete, the study is dependent on supplementary self-reported schooling data, which contains an additional element of bias. Identifying a linear functional form as preferable to a semi-log functional form, Grubb distinguishes between the returns to credential completers and non-completers.

To control for ability – which the NLS72 does not measure after 1972 – the author includes ability/achievement measurements prior to enrolment in PSE. This includes TEST (a literal test measuring mathematical and verbal abilities in 1972), self-reported high school grades, and SES

(a measure of socioeconomic status). However, Grubb identifies some problems inherent in certain ability measures. For instance, the measurement of grade point average during PSE is limited to only those with PSE and by the tendency of students to avoid courses they may struggle with or drop those they are failing.

Kane, Rouse, and Staiger (1999) implement a "method of moments" technique that uses differences in wages, test scores, and other variables to identify measurement error in self-reported and transcript-reported schooling, and they consider the implications these errors have for OLS and IV estimates of the return to PSE. The authors identify the Census bureau's question regarding level of education as preferable to that of the NLS72. Whereas the Census asks participants to report years of schooling attended and whether they completed the highest grade listed, the NLS survey asks, "Since leaving high school, about how many credits which can be used for a 4-year college Bachelor's degree had you earned by October 1979?" (p. 20).[15] In applying their method of moments technique to the NLS's educational measures, Kane et al. use data from the Post-secondary Education Transcript Survey (PETS) and find that over 90% of PSE graduates with associate's or bachelor's degrees accurately report their degree attainment, but only approximately half of the individuals with 1 or 2 years of college credits give an accurate report. These measurement errors result in OLS estimates understating the returns per year of schooling and overstating the effect of degrees on return and lead to additional biases in IV estimates.

Dee (2004) measures the civic returns to PSE attainment. This study uses data from the High School and Beyond (HS&B) longitudinal study and from the 1972-2000 General Social Surveys (GSS) to estimate the effects of college entrance on voter and volunteer participation and on the frequency of newspaper readership, respectively. The HS&B study began in 1980 with a group of high-school sophomores who answered a series of interview questions as they aged, including a civic-related interview in 1992. Dee uses standardized test scores from the sophomore survey, along with students' responses to questions about correcting social and economic inequality, to predict future civic involvement. However, because the HS&B study faces limitations by only identifying civic returns at the PSE level and providing no measures of the degrees of civic engagement, the author also draws on data from the GSS—a collection of nationwide surveys in which respondents provide information about their education and voting participation.

Dee uses the local proximity of 2-year colleges and measurements of newspaper readership as proxies for degrees of civic awareness and finds that 2-year college availability has significant effects on college entrance, which increases voter participation by approximately 17-22 percentage points. The author also finds that additional schooling significantly increases frequency of newspaper readership, which serves as a proxy for civic awareness. Though, it should perhaps be noted that in an increasingly digitalized age, newspaper readership may no longer be the most accurate proxy to use.

[15] Kane et al. (1999) note that because the question asked by the NLS-72 is not directly comparable to that asked by the CPS, they cannot directly test the validity of the former.

Lemieux (2006) uses data on male hourly wages from the May and Outgoing Rotation Group (ORG) supplements of the Current Population Survey (CPS) to correlate wage inequality with increasing returns to PSE. More specifically, the author identifies trends in wage inequality by comparing data from the 1972-1974 May CPS and the 2003-2005 ORG CPS. He finds that the standard (log) wage equation does not sufficiently reflect the fact that returns to PSE vary across all workers, so he turns to a human capital model with heterogenous returns, which suggests that most of the increase in wage inequality is due to "dramatic increases in the return to postsecondary education" (p. 199). Lemieux finds that changes in wages occur especially among workers with PSE, whereas relative wages remain stable for workers without PSE, suggesting a high return to education.

Brand and Xie (2010) consider variations in the economic returns to a college education across a selection of the population of the United States. Whereas economic scholars typically support the "positive selection hypothesis," which suggests that the individuals most likely to select into college (i.e., socially advantaged individuals) also tend to benefit the most from college, Brand and Xie conjecture the opposite—that the individuals who are least likely to receive a college education will benefit the most. They refer to this assumption as the "negative selection hypothesis" (p. 274) and judge between the two theories by invoking an ignorability assumption[16] and then looking at the earnings effects of completing college through "propensity score strata using an innovative hierarchical linear model" (p. 273). In other words, the study departs from assumptions of population homogeneity and instead groups individuals according to their estimated propensity, or likelihood, to complete college. Brand and Xie estimate binary logistic regressions predicting propensity for completing college between college graduates and non-college graduates (high school graduates without PSE completion).[17]

Further, they draw on two sources of data: The National Longitudinal Survey of Youth and the Wisconsin Longitudinal Survey, which allows them to replicate their analysis for two different cohorts (national and regional) and therefore test the robustness of their findings. The study finds evidence for negative selection for both cohorts. However, Brand and Xie find evidence for positive selection when they restrict their models to more limited sets of covariates, suggesting that missing certain key variables can lead to a distortion in the pattern of results.

CENSUS DATA

In Canadian studies, census data typically refers to the information gathered by the Census of Population, which is conducted by Statistics Canada every five years. Although the long form of this census is not dissimilar to survey data – representing only a random sample of the population – the more common short form includes information about every member of the Canadian population. The long-running, consistent nature of census databases makes them valuable, easily comparable sources of information (Boudarbat et al., 2010). However, the systematic data collected typically does not allow researchers to control for the innate abilities of graduates or

[16] The "ignorability assumption" is also referred to as "unconfoundedness" or "selection on observables," and it states that "potential outcomes are uncorrelated with treatment status, conditional on observed covariates" (Brand & Xie, 2010, p. 276).

[17] To see the multiple formulas used by Brand and Xie (2010) in their analyses, see pp. 275-281.

for the non-monetary benefits associated with PSE completion (Boothby & Rowe, 2002; Stager, 1996).

CANADIAN STUDIES

Deviating from the survey data used in his 1986 study, Vaillancourt (1995) calculates both public and private rates of return to education in Canada by level and field of study. The author uses data and earnings equations from the 1986 Census individual microdata file to project lifetime earnings profiles, and he uses both gross and net earnings to calculate public and total rates of return, respectively, while combining this data with the costs of schooling during the study period.[18] Vaillancourt finds that the highest public and private rates of return result from high school completion, and rates of return consequently decline with higher levels of schooling.

However, he also identifies some problems inherent in his methodology. Not only does the study lack a consideration of other employment rewards and non-employment investment benefits (which results in an underestimation of returns), but it also uses single-year results to represent a lifetime of employment income. Moreover, the census data collected does not identify the number of individuals still attending school, and the decision to include only individuals working full-time and earning positive income excludes those who work part-time by choice—an important facet to consider, as full-time employment increases with schooling. Finally, the conditions of the study period for this report no longer represent the current labour force, market, or employment climate.

In his estimations of returns to investment in university programs in Ontario, Stager (1996) compares the mean earnings of university graduates and high school graduates by five-year age-groups. Using data from the 1991 Population Census of Canada and the 1987 Profile of Post-Secondary Students in Canada, the author compares the benefits of PSE with the cost. There are some limitations to this method, including criticisms that benefit-cost analysis does not allow for "innate ability" (p. 6); it excludes non-monetary and external benefits; and it does not control for different labour market structures. Nonetheless, Stager compares the earnings differentials between university and high school graduates and concludes, first, that in 1990 the private rates of return to PSE ranged from 7% (humanities) to 21% (medicine) and, secondly, that major changes in tuition prompt little change in return because tuition fees comprise such a small portion of the cost students pay for PSE.

Boothby and Rowe (2002) use Statistics Canada's LifePaths model[19] to simulate the lifetime earnings of pairs of high school and PSE graduates. Using 1991 Census data along with data on the earnings streams of high school graduates, the direct costs of university education, and the earnings streams of university graduates, the authors measure individual rates of return by level and field of study.[20] Boothby and Rowe simulate the lives of pairs of individuals who are identical until one enrolls in PSE, at which point a high school graduate clone becomes a control for a PSE graduate. Their study shows a wide range of outcomes, such as median rates of return ranging

[18] For the formulas used, see Vaillancourt (1995), pp. 533 and 540.
[19] For more information on the function of the LifePaths model, see Boothby and Rowe (2002), pp. 15-23.
[20] For the formula used to compute the private internal rate of return to a post-secondary degree, see Boothby and Rowe (2002), pp. 4-5.

from 5% to 23% for undergraduate degrees and 0% to 20% for community college. This variance, they claim, is likely due to individual characteristics that might affect earnings potential, or decisions regarding time use, or even just sheer luck. Along with the LifePaths model's inability to account for such factors, the authors also identify other limitations within the model, such as the exclusion of the tax and transfer system and lack of allocation for time unemployed and time out of the labour force.

Using data from the 1996 Canadian Census, Ferrer and Riddell (2002) empirically compare two models for measuring the impacts of PSE on labour market earnings: Mincer's (1974) human capital earnings function (HCEF) model, which expresses individual earnings in terms of years of schooling completed, and a "credentialist" model, which holds that degrees completed matters, rather than the years spent attaining them. Ferrer and Riddell use a hybrid model[21] to compare three samples of individuals: high school graduates and dropouts; non-university PSE graduates and dropouts, and university graduates and dropouts—each group having attended school for the regular amount of years associated with each degree. Their results show that employed university graduates earn 54% more than high school graduates and 37% more than college and trade-school graduates. They conclusively reject the separation of the HCEF and credentialist models and argue that both dimensions (years of schooling and degrees completed) should be included in empirical studies.

Boothby and Drewes (2006) use Census data from 1980-2000 to measure the evolving earnings premiums to undergraduate, college, and trades education relative to the earnings of a control group of individuals whose highest level of education is high school completion. The authors use ordinary least squares estimates of a HCEF model to develop weekly earnings differentiations for the various levels of education.[22] They also note the importance of distinguishing between earnings premiums and rates of returns to investment in education, given the different costs associated with trades, college, and university educations.

Stark (2007) computes private returns to university graduates in Canada according to field of study and level of education by considering the impact of direct and indirect costs of PSE, as well as income taxes. Moreover, he analyses tuition increases to estimate their probable impact on rates of return. The author uses a combination of census and survey datasets, drawing on earnings data from the 1996 Census of Canada and data on university fees from Statistics Canada's Tuition Fees and Living Accommodation Costs at Canadian Universities Survey, 1995-96. However, Stark acknowledges limitations inherent in the datasets, such as a lack of data on earnings of high school graduates, which he instead computes from the 1996 public microdata file. Before the OLS regression for earnings,[23] Stark calculates after-tax earnings for a representative childless individual by drawing on the federal income tax schedule EI and CPP contributions, as well as federal surtax and GST credit (though without including deductions for RRSP).

---

[21] The hybrid model formula for HCEF functions supplemented by dummy variables is located in Ferrer and Riddell (2002), pp. 890, 892.

[22] A formula for a hybrid credentialism/years of schooling HCEF is found in Boothby and Drewes (2006), pp. 6-7.

[23] For the formula used, see Stark (2007), p. 9.

He also makes assumptions regarding the age at which an individual undertakes PSE and the duration of the various levels of programs, further estimating an "alternative earnings stream" for each level of university education as "the earnings of a person who had the credentials necessary to pursue the investment, but did not do so" (p. 5) (I.e., a person who graduated a level below them). To finally estimate rates of return, Stark compares the earnings estimate of a (non-medical) bachelor's degree to a high school diploma; a master's degree to a bachelor's degree; a PhD to a master's degree; and a medical degree to a high school diploma.

Boothby and Drewes (2010) supplement their previous study (2006) by using data from the 2006 Census of Population to measure financial returns to different types of PSE in Canada from 1980-2005. They examine whether these reflections of supply and demand indicate that Canada has produced too many postsecondary graduates or, more specifically, too many graduates from community colleges or trades schools. The authors measure financial returns and account for foregone earnings and length of time invested in education by assuming individuals undertake full-time studies at age 18 and proceed to graduation without any interruptions—spending 4 years on a bachelor's degree, 3 years on a community college diploma, and 2 years on a trades certificate. By comparing the computed earnings premiums for graduates of universities, community colleges, trades schools, and high schools, Boothby and Drewes (2010) determine that there are consistently high rates of return to higher education for all measured groups except female graduates of trades schools. However, they acknowledge the need for a "comprehensive understanding of the length of studies and the variable enrolment costs across jurisdictions" (p. 5).

Boudarbat et al. (2010) similarly respond to and update the findings of earlier studies. More specifically, they attempt to reconcile the different conclusions reached by Burbidge et al. (2002) and their own earlier work (2006) while measuring the evolution of returns to PSE between 1980 and 2005. Attributing the differences between previous studies to the use of different data sources, Boudarbat et al. employ data from the 1981, 1986, 1991, 1996, and 2001 Censuses, as well as master file data from the 2006 Census. While recognizing some limitations to Census data – such as limited information on annual hours of work and changes in the 2006 Census regarding information on income, earnings, and number of years of schooling – the authors advocate for the Census as the "best data source for documenting trends in the wage structure in Canada" (p. 67) because of its long-running, consistent information on years of schooling, degrees received, earnings, work experience, and economic characteristics of individuals. In contrast, the Survey of Consumer Finances (SCF), which was discontinued in 1997 and more recently used by Burbidge et al. (2002), uses inconsistent questions on educational attainment and does not contain information on years of schooling.

In measuring the wage premiums to attending university against entering the workforce after completing high school, Boudarbat et al. stress the importance of controlling for differences in labour market experiences and age-wage differentials, which they do by presenting raw and adjusted measures of returns to education. Acknowledging that their available data cannot control for unobserved differences such as motivation, ability, and perseverance, the authors nonetheless find that controlling for work experience and using Canadian Census data are responsible for the differences between their current results and earlier studies.

Reano-Alvarez (2012) uses data from the long form questionnaire (2B) of the 2006 Canadian Population Census to measure the return to receiving higher education outside Canada. However, he acknowledges that this data faces some limitations, including random sampling and lack of information on location of study for those who did not complete education beyond the high school level. Nonetheless, by using weighted data, Reano-Alvarez evaluates "the extent to which background variables such as labor and socioeconomic characteristics account for the variability in earnings attributed to study location" (p. 4). The study exclusively analyzes a subset of the population aged 25-59 years, with PSE, whose field of study is included in the Canadian government's list of targeted occupations. Using OLS regression with employment income as a dependent variable, and studies in Canada as the control group, the author finds negative returns to investment in education outside Canada—even if that education is attained in other English-speaking countries. However, for immigrants, he finds that the negative effect on earnings decreases over time spent in Canada.

AMERICAN STUDIES

In a study focused on rising income inequality in America, Autor (2014) looks to widening gaps in returns to education as an explanation. The author calculates earnings premiums from 1979 to 2012 by drawing on Census Bureau P-60 (1979-1991) and P-25 (1992-2012) data on the median earnings of full-time workers by education level, which he converts into 2012 dollars by using the CPI-U-RS price series and into household income by comparing husband-wife earnings from each educational group. The comparison of the median earnings gap between high school graduates and 4-year college graduates shows that from 1979 to 2012, this gap approximately doubled. Autor attributes this growth in inequality not only to the rising relative wages of PSE workers, but also to the declining real earnings of high school graduate workers.

ADMINISTRATIVE DATA

Unlike survey and census data, administrative data is not collected primarily for research but rather serves an administrative purpose. Gathered by government departments, postsecondary institutions, and other organizations, this data is used for running programs, delivering services, registering events, administering benefits, and so on. Moreover, it can include a wide assortment of variables, such as high school grade averages, application files, tax records, and the data stored in postsecondary information systems. Although administrative data is often a cost-effective alternative to direct data collection, it nonetheless can pose problems for researchers, as they have little to no control over the quality and timeliness of the data (Statistics Canada, 2017).

As Finnie and Usher (2005) propose, administrative data is typically implemented alongside other classes of data as valuable supplemental information. For example, Boothby and Rowe's (2002) report uses both 1991 Census data and administrative files, constructing an individual longitudinal data set from the Community College Student Information System/University Student Information System (CCSIS/USUS) maintained by Statistics Canada. In a similar manner, Dodge and Stager (1972) supplement their estimation of earnings premiums with a comparative estimation of the cost of university in 1966-67 by drawing on the Government of Ontario's university grant formula. However, the majority of reports on the return to PSE exclude

administrative data from their calculations, drawing primarily on either survey data, census data, or both.

Straying from the practice of empirical, formula-based analysis, Frenette et al. (2014) focus on how research on returns to adult learning could be improved through better use of the survey and administrative data resources of Statistics Canada and HRSDC. The authors argue the benefits of linking resources within and between each database, identifying the advantages and limitations inherent in each.[24] However, they acknowledge, linking these currently existing resources will not eliminate research limitations. Adult training occurring outside PSE institutions and/or HRSDC-administered programs would still be absent from the datasets. Moreover, creating control groups for comparison to program participants would be extremely difficult, as tax-based administrative files tend to lack sociodemographic information, including age and occupation variables. For example, HRSDC tends to only collect data for program participants, and it is possible that some non-participants did not qualify for the same program and therefore would not form an adequate control group.

Our review of other currently and potentially available datasets offers various alternatives for addressing these limitations. Moreover, this supplemental data could be linked to the ELMLP for use in the development of suitable control groups.

[24] For specific examples of the types of data linkages proposed, see Frenette et al. (2014), pp. 25-26.

CURRENTLY AND POTENTIALLY AVAILABLE DATASETS

A suitable control group for the PSE graduates in the ELMLP would contain individuals who have not received any higher education, including forms of PSE not covered by PSIS or RAIS. Such a comparison group could be developed through the analysis of more datasets containing information on the education levels of Canadian residents, as well as by linking these panels to historical tax records. This combination of linked datasets could address some of the limitations currently facing the ELMLP and can help researchers develop and identify larger, more accurate control groups across Canada.

PSIS / RAIS

In its present form, the ELMLP links the Postsecondary Student Information System (PSIS), the Registered Apprenticeship Information System (RAIS), and T1FF data. A control group could be obtained by using these linked datasets to identify tax filers without any PSE or postsecondary training by selecting a group of individuals in the T1FF but not in either PSIS or RAIS. This would involve the assumption that tax filers missing from the PSIS database did not attend or complete any PSE/training in Canada. The PSIS could also help identify the immigration status (Canadian citizen, permanent resident, student visa, etc.) of select individuals for the duration of their study period in Canada.

However, this method faces a significant limitation: Both PSIS and RAIS only consider PSE/training that occurred in Canada. Therefore, immigrants and Canadian-born residents who attended or completed PSE outside Canada would be indistinguishable from those with a high school diploma or less. While this method still allows the investigation of returns to Canadian PSE against a combined control group of Canadian high-school graduates and immigrants with any level of education (obtained outside of Canada), the findings would not be suitably robust.

This limitation could be partially addressed through an additional linkage: incorporating data from the Longitudinal Immigration Database (IMDB) into a reverse linkage of the ELMLP datasets.[25] The IMDB is a comprehensive dataset that links administrative data from multiple other datasets. This includes landing records from Immigration, Refugees and Citizenship Canada (IRCC), T1FF, the Immigrant Landing File (ILF), and the Non-Permanent Resident File (Temporary Residents). Information available from the landing records would make it possible to distinguish between immigrants and Canadian-born residents with a high school diploma or less. Furthermore, the IMDB already includes the Integrated Permanent and Non-permanent Resident File (PNRF), which contains the self-reported full education profiles of permanent and temporary (who became permanent) residents upon landing. This would be useful in identifying landed immigrants with international credentials only—individuals missing from the linked PSIS/RAIS/T1FF dataset. This data could also be conducive to the development of additional comparison groups based on level and location of education. Since a linkage between T1FF and IRCC landing records already exists in IMDB, it should be available soon for analysis.

ADVANTAGES

[25] This would involve linking PSIS/RAIS data to T1FF, whereas the ELMLP currently links T1FF data to PSIS/RAIS.

Since PSIS and RAIS are both administrative databases, the education data they contain are not self-reported and thus are likely to be more accurate and complete than education data derived from surveys. These datasets can be used in various analyses to incorporate data on complete education profiles (in Canada), such as highest education level, number of years in postsecondary education, withdrawal from an educational program prior to completion, and current education status. Additionally, researchers could use data from PSIS/RAIS to analyze how returns to education vary across cohorts based on program level (I.e., Masters, Bachelors) and to indicate the immigration status of students in Canada.

DISADVANTAGES

One of the main disadvantages of the PSIS/RAIS datasets is the lack of available information on PSE attained outside Canada. Data on educational attainment is available only for education undertaken at a Canadian institution. This especially limits data on the education levels of immigrants arriving in Canada as adults, as most of their PSE is likely to have been obtained outside of Canada.

LAD / IMDB

The Longitudinal Administrative Databank (LAD) is a longitudinal file (1982-2016) comprised of a 20% sample from T1FF records and the Longitudinal IMDB, which consists of information on immigrants who landed in Canada post-1980. Linking LAD with ELMLP would enable researchers obtain control groups from tax files by analyzing the age, education deductions, and tax credits of tax-filers. If certain individuals file income tax after the age of 18 without claiming any education deductions or tax credits at any point, it could be assumed that they did not receive any PSE in Canada. While this excludes data on immigrants who completed their full education abroad, this missing information could be supplemented by observing the education profiles available through IMDB landing records. Beyond this, any immigrant who filed for education tax credits could be assumed to have attended some level of PSE in Canada. Additionally, Canadian residents who attended university (bachelor's degree or higher) outside of Canada could be indirectly identified if they filed for education deductions/tax credits but are missing from the PSIS database.

ADVANTAGES

The LAD provides a method for distinguishing between Canadian-born residents with high school credentials or less and immigrants with non-Canadian PSE credentials. Using data on immigration status, landing year, and the self-reported education levels of immigrants, it would be possible to conduct a longitudinal analysis to estimate returns to PSE.

DISADVANTAGES

The LAD is limited to information on immigrants who landed in Canada since 1980; therefore, any immigrant landing prior to 1980 would be indistinguishable from Canadian-born residents. Furthermore, it is not possible to differentiate between individuals with high school education and those with less. Lastly, Canadian-born residents with overseas PSE credentials who did not file for education deductions or tax credits or studied for level lower than the bachelor's degree would remain unobserved. These limitations could lead to less robust estimations.

CENSUS

LMIC (2018) proposes the integration of data from the Canadian Census into the ELMLP. The long form of the 2016 Census has the advantage of providing the largest sample size from which to obtain control groups, and it includes information on immigrant status, educational attainment, and whether an individual's highest level of education was completed outside of Canada. Moreover, both short- and long-form Census data have been linked with T1 and CRA records for the purpose of gathering income information.

ADVANTAGES

The Education section of the Census provides valuable information on whether an individual completed high school or an equivalent level, as well as any PSE, including trades certificates, non-university diplomas, and university degrees. It also identifies the field of study for the highest certificate attained, which could be used to develop various control groups for analyses of return by field of study. Finally, Census data specifies whether a person's highest education credentials have been attained in Canada or abroad.

DISADVANTAGES

The Census only includes information on the status of an individual's education (such as the highest level or degrees completed) during or prior to the latest census year. Since the most recent data available is from the 2016 Census, the education level of students who had not completed their PSE by that time would not be indicated by the Census. Moreover, other educational qualifications apart from the (self-reported) highest level of education completed are not reported. Only degrees in medicine, veterinary science, dentistry, or optometry stand as exceptions to this rule. Thus, the full education profile of individuals is not observed, in addition to probable biases in how respondents assess the highest level of education they obtained.

LISA

The Longitudinal and International Study of Adults (LISA) collects data on the jobs, education, health, immigration status, and families of participants (aged 15 and over) across Canada. Most importantly for this study, LISA asks multiple and detailed questions regarding respondents' educational levels, expected educational completions, and indicators for multiple certifications. LISA also covers fields of specialization and whether respondents completed their education in Canada or abroad. This biennial survey began in 2012, with the latest cycle completed and available in 2016; therefore, the LISA dataset could be used to develop control groups of

respondents with different levels of education for three different reference periods: pre-2012 (Wave 1), 2012-2014 (Wave 2), and 2014-2016 (Wave 3).

LISA has already been linked to T1FF, which provides historical tax data (1982-2015) for respondents. Overall, the historical linkage rate between the two datasets is high – over 90% in most cases – and is stable when LISA respondents are likely to file a tax return. However, the linkage rate of a given year decreases if the sample includes either respondents under the age of 20 or immigrants who arrived within three years of the study. This issue can be resolved by dropping either or both groups from the sample and only keeping individuals who are likely to file taxes in a given year. Linking LISA and T1FF datasets can also result in the creation of long, balanced panel data ranging from 5 to 35 years.

While there are multiple ways of extracting a control group from LISA, the most direct method would include the creation of cohorts according to highest level of education and period of attaining it. This could include groups with less than a high school diploma, high school graduates, and so forth. Alternately, various cohorts and control groups could be determined according to the number of years of formal education (starting from grade 1). In this scenario, anyone with less than 12 years of formal education could be considered to be without a high school diploma; those with 12 years of education could be considered high school graduates; and those with more than 12 years could have a high school diploma and university degree(s) or other forms of PSE. When combined with other variables, the number of years of formal education could also be used to indirectly determine whether respondents left university prior to completion; however, no indicators directly identify incomplete PSE.

ADVANTAGES

LISA contains variables identifying individuals with multiple degrees or certifications. In LISA 2014 (Wave 2), all completed education levels during the reference period are recorded. For the PSE sections, details such as program duration, year of graduation, and location are collected for the most recent degree. However, separate sections on the highest level of education in Canada and other countries could be used to indicate any other degrees attained before or during the reference years. Similarly, in LISA 2016 (Wave 3), an additional section on the history of individuals' PSE collects data on the three most recent degrees or certificates obtained.

Analyses based on Canadian-born and landed immigrants can be conducted using LISA, since the survey contains variables pertaining to this subject. Moreover, the dataset also makes it possible to identify any Canadian-born individual who studied or at least attained their highest education level overseas.

Since the linkage of T1FF and all three LISA waves is already available, the dataset should be available immediately. The high linkage rate makes it possible to obtain longer historical tax-filer information for the control group, going all the way back to 1982. Moreover, a balanced panel dataset is a useful tool for estimating returns to PSE, as it opens the possibility of conducting several analytical approaches. For example, a 30-year panel has a 74.3% linkage rate and over 6,500 respondents, while a 5-year panel has nearly an 88% linkage rate and over 16,500 respondents.

DISADVANTAGES

The sample size for each wave is approximately 30,000 individuals, which could result in a smaller control group than other datasets in questions, such as the Census and PSIS. Moreover, unless studies exclude individuals who are unlikely to file taxes, the linkage rate between LISA and T1FF decreases. While the exclusion of these groups could lead to a very high linkage rate and a longer, more balanced panel of data, it nonetheless could also mean losing some data on education levels in the control group, and resultant estimations may not be robust.

GSS

The General Social Survey (GSS) is administered annually, and every cycle includes a targeted questionnaire related to various topics such as caregiving, families, time management, social identity, volunteering, and victimization. The GSS contains comprehensive socioeconomic data on participants' age, sex, education, income, immigrant status, and so forth; and, with the exception of 2002 and 2007, the survey targets the general population aged 15 years and older. Moreover, the Education section of the GSS contains information on highest level of education; years of primary and secondary education; and current student status. The survey also provides information on respondents' immigration status, year of landing, and age, which can be used to derive education location with some possibility of error.

These variables could be used to create control groups consisting of high school graduates and of individuals without a high school diploma. Several GSS cycles could also help identify PSE non-completers—individuals who attended a post-secondary institution without graduating. However, it would not be possible to determine the number of years spent in PSE before exiting.

In the 2015 and 2016 cycles of the GSS, there were no questions related to income. Therefore, to obtain the income information of respondents, Statistics Canada linked the GSS to T1FF data. The 2016 GSS has been linked to the 2015 T1FF, gaining information on 90.6% of respondents, and the 2015 GSS was linked to the 2014 T1FF, attaining the information of 88.6% of respondents. Additionally, linked records between the 2011 GSS, 2013 T1FF, and 2014 Immigrant Landing File are available. To attain data for previous records, one could request the linkage of a GSS control group to T1FF records through probabilistic matching.

ADVANTAGES

The GSS provides detailed accounts of the education profiles of respondents. This includes highest education level, years of primary and high school, and current student status. Some cycles (e.g., 2009, 2010, 2011, 2014, and 2016) include information on the highest education levels of respondents' parents and spouse, and others (2009, 2010, and 2011) contain indicators for only partially completing PSE. The indicators for immigration status and year of landing are useful in classifying Canadian-born citizens who studied abroad and landed immigrants that studied in Canada. Recent datasets are already linked with T1FF, and upon confirmation from Statistics Canada, they should be available for request immediately.

DISADVANTAGES

More recent GSS cycles (2008-2011, 2013, and 2016) have indicators for the country of origin for highest education level completed, but these face data quality issues, since the indicator is not consistently present in all recent years. An alternate option for imputing the location of highest education level is to use respondents' immigration status, year of landing, and age. However, the size of the control groups obtained may be quite small, based on the approximate sample size of 25,000 individuals. In particular, the 2011 and 2012 GSS cycles target only individuals aged 45 years and over, providing an even smaller sample size and restricting age cohorts for control groups.

PIAAC

The Programme for the International Assessment of Adult Competencies (PIAAC) is a survey measuring adults' proficiency in literacy, numeracy, and problem solving in 'technology-rich' environments. The PIAAC is an OECD initiative and has been conducted in over 40 countries, with the latest Canadian cycle (2012) targeting adults aged 16 to 65 years. Along with proficiency measurement, the PIAAC collects information on the age, education, and employment status of respondents. In particular, the survey's Education and Training section contains information on highest level of education and asks specific questions pertaining to high school education.

Data from the PIAAC can be used to create several cohorts among individuals whose highest level of education is high school or lower. Further distinctions can also be made according to categories such as specialization and the location of high school and/or highest education level (I.e., Canada or other countries). Finally, the survey identifies individuals enrolled in formal education when it was administered. While there is no direct indicator for examining whether respondents exited any form of PSE prior to completion, data on years of formal education could be used to proxy for incomplete PSE.

Statistics Canada has already enhanced PIAAC data by linking the dataset with the 2011 National Household Survey, the 2011 Census, and IRCC landing records through deterministic hierarchical record linkage. The PIAAC could similarly be linked with historical tax data (1982-2015) from the T1FF. Even though the resultant study would only contain education information from respondents until 2012, the next survey cycle will take place in 2021, providing another linkable dataset for the ELMLP.

ADVANTAGES

The PIACC contains detailed education profiles of the target population. Anyone who attended PSE but did not complete the program could be indirectly identified. Finally, it is possible to specify whether highest level of education was obtained in Canada or elsewhere.

DISADVANTAGES

The PIAAC dataset includes self-reported education information, which is only a snapshot at a point in time (prior to 2012). Currently, there are no linkages to T1FF available. Even though Canada's sample of PIACC is the largest of all OECD countries, with 25,267 respondents, the control group could still be smaller than those of other available datasets.

RECOMMENDATIONS

We recommend four methods for obtaining control groups when calculating the returns to PSE:

1) Using the datasets already linked through the ELMLP to define control groups of individuals who pursued but did not complete PSE.

2) Defining and computing aggregated cohorts of individuals defined by age, sex, region, and other observable characteristics for specific education levels drawn from various unlinked datasets. The resultant earnings profiles can then be used as comparison groups in a synthetic cohort analysis following the approach of Green and Worswick (2012).

3) Isolating individuals without evidence of higher education in administrative tax records to obtain control groups of individuals whose highest level of education is secondary school or less.

4) Expanding the information currently accessible through the ELMLP with additional datasets, some of which have already been linked with tax records; some that are accessible but have not yet been linked; and some that Statistics Canada may eventually collect and link.

In particular, we suggest analyzing additional datasets to generate estimates on the earnings of high school graduates, high school non-completers, and PSE graduates who earned their credentials abroad—all cohorts which are currently not covered by the ELMLP, but which could be used in the development of earnings comparison groups.

While the steps already taken to link administrative data on education and the labour market have facilitated a more comprehensive approach to estimating the return to PSE among those with at least some postsecondary Canadian education, the ELMLP is limited by a lack of data on topics such as (but not limited to)

- PSE abroad
- ability measures
- occupation, age, and workplace experience
- and non-monetary benefits

Incorporating data from the Canadian Census, LAD, LISA, GSS, PIAAC, IMDB, and other databases would provide researchers access to information missing from the ELMLP, enabling the formation of more applicable control groups and thus producing more accurate estimations of the return to PSE in Canada.

IMMEDIATE RECOMMENDATIONS

Our immediate recommendation includes the use of data that is readily available, particularly the datasets linked through the ELMLP (PSIS, RAIS, T1FF).

In the ELMLP, T1FF data is currently linked to PSIS/RAIS data, which limits the existing sample to only individuals who have experienced some form of PSE. While this makes it difficult to obtain control groups of high school graduates without PSE experience, or high school non-completers, the available data can nonetheless be used to develop comparison groups of individuals with varying levels of PSE.

We also propose using information from unlinked datasets in the immediate formation of control groups. Following the method outlined by Green and Worswick (2012), we recommend using a synthetic cohort approach: combining aggregate pieces of background data (such as age, sex, region, level of education, and so forth) from individual datasets to obtain earnings profiles that can be grouped into cohorts for comparison in a regression analysis.

SHORT-TERM RECOMMENDATIONS

Our short-term recommendations involve obtaining control groups from administrative tax records, as well as from other datasets that have already been linked with tax records and could also be linked to ELMLP (Census, GSS, LAD, LISA, PSIS, RAIS).

We recommend a reverse linkage of the PSIS/RAIS/T1FF datasets already linked by the ELMLP. Linking PSIS/RAIS data to T1FF would allow researchers to develop control groups with high school credentials or less, based on the assumption that individuals in the T1FF dataset but excluded from PSIS/RAIS do not have any PSE credentials. However, this approach is limited to education attained in Canada, preventing researchers from distinguishing between immigrants with international PSE credentials and residents possessing only high school credentials or less.

Because the ELMLP provides access to the tax records for individuals whose PSE credentials are not reported, researchers may underestimate the actual return to PSE. When more highly educated persons (and their earnings) are placed in control groups of individuals without PSE, the estimated returns to PSE are thus more likely to be skewed.

We also recommend using LAD data on education tax credits and deductions to develop control groups without any PSE. By identifying individuals with consistent earnings from ages 18 or 19 onward, and whose tax filings lack evidence of higher education, one can obtain control groups based on the assumption that these individuals did not pursue any form of PSE and obtained only secondary school credentials or less. However, these administrative tax records – like the PSIS/RAIS/T1FF linkage – do not reflect PSE attained abroad and therefore do not distinguish between international higher education and high school diplomas or less, nor do they account for individuals who may have undertaken PSE but did not claim education expenses or tuition tax credits.

To address the limitations facing the PSIS/RAIS/T1FF linkage and education data from the LAD, we recommend an additional linkage with IMDB, which contains the education profiles of immigrants before their arrival in Canada. Combined with LAD data on immigration records and

citizenship status, the IMDB linkage could help develop a control group of immigrants who studied abroad. Moreover, the available (though self-reported) data on the highest education credentials at point of landing in Canada as permanent residents can be used to develop control groups for returns based on level *and* location of education. Nonetheless, these datasets cannot be used to identify Canadian-born residents who received PSE credential abroad.

We suggest incorporating the LISA dataset into the ELMLP. Because the 2012 and 2014 LISA cycles already have a high linkage rate with T1FF through Statistics Canada, the available linkage can be used to create a panel dataset in which data-merging brings cohorts of individuals with no postsecondary education into the sample universe, thereby adding a control group. LISA collects data on participants' highest level of education, current enrolment status, years of formal education, field of specialization, and location of educational completion, and this dataset can be used to form multiple control groups with various levels of education.

Like the LAD, the LISA dataset can be used to obtain control groups with high school or less, in addition to control groups mirroring those available through PSIS/RAIS. However, while the LAD control groups rely on researcher assumptions, LISA explicitly indicates which individuals do not have higher education. Moreover, LISA data can be used to obtain cohorts of PSE graduates who received their credentials outside Canada, as well as cohorts organized by field of specialization, both of which could be implemented by future studies to estimate varying returns to PSE based on location of education and on field of study.

Finally, collected data on the jobs, health, and families of LISA participants can serve as variables to control for factors such as individual ability and propensity to pursue PSE, while accounting for some non-monetary returns to PSE. It should be noted, however, that the sample sizes of the control group(s) will be substantially smaller than those already in the ELMLP linkage owing to the fact that LISA is a survey.

We recommend the integration of Census data into the ELMLP (LMIC, 2018), as the former contains the largest sample size from which to obtain control groups. Moreover, both short- and long-form censuses have already been linked to T1 and CRA records. Census information on the status and level of individuals' (self-reported) education can be used to develop various cohorts according to educational level, as well as completion or non-completion of high school or PSE. In addition, data on fields of study can help determine rates of return according to specialization, and the location of highest credentials received can be used to distinguish between individuals with no PSE and individuals who received PSE abroad.

Similarly, we propose linking the 2015 and 2016 cycles of the GSS to the ELMLP. While a linkage of previous GSS cycles and tax records would need to be requested, Statistics Canada has already linked the 2015 and 2016 GSS cycles with personal tax records (T1, T1FF, or T4), and this information is readily available. Control groups of high school graduates, individuals without a high school diploma, and PSE non-completers could be obtained from GSS data on the highest level and years of education, as well as current student status.

Furthermore, the survey's socioeconomic data on participants' families, income, social identity, and so forth could be used to control for ability and propensity to pursue PSE, and information

on the age of participants can be used as a proxy for workplace experience, thus allowing researchers to distinguish between true return to PSE and natural workplace progression.

INTERMEDIATE-TERM RECOMMENDATIONS

Our intermediate-term recommendations include the incorporation the PIAAC: a dataset that is currently accessible, but for which a linkage to tax records *and* the ELMLP must be requested. The PIAAC dataset has already been linked with NHS, Census, and IRCC data, and we suggest a further linkage with historical tax data through the T1FF. Similar to Census data, PIAAC data can be used to create comparison groups organized by specialization, location of highest level of education, and completion or non-completion of educational programs.

LONG-TERM RECOMMENDATIONS

Our long-term recommendations address variables that should be considered in future studies on the return to PSE, many of which are not in Statistic Canada's data holdings but which could be gathered through targeted questions in future surveys or by accessing more administrative datasets.

COST AND YEARS OF EDUCATION

For cost-benefit analyses of PSE, we would recommend obtaining data on the varying costs of tuition (by level, region, etc.), as well as the length of time spent obtaining PSE credentials. These variables are often imputed to average amounts, but more accurate data on each could have a large impact on the estimated returns, as well as the methodologies used to obtain them. In particular, degrees that took longer than average to complete would not only result in higher tuition costs but could also lead to a shift in the baseline age at which most studies begin measuring returns to PSE. If many students are taking more than four years to complete a university degree, for instance, it may be more beneficial to calculate the returns to 24-year-old PSE graduates than to calculate the returns to 21-year-old graduates.

ABILITY MEASURES

The ELMLP provides access to data on parental income through T1FF. Following the work of Corak (2013), researchers can use this information to control for some sociodemographic measures in estimating returns through the inclusion of controls for the family income of individuals when they were students. However, while family income can affect the propensity of students to attend a postsecondary institution, it does not necessarily reflect the abilities of the students who pursue further education. This could be addressed by analyzing data on high school grades and test scores—for those who attended a postsecondary institution and for those who entered the labour force immediately after completing high school. Resultant propensity score matching could enable researchers to develop comparison groups of high school graduates and PSE graduates who had similar grades and therefore, assumedly, similar abilities upon high school completion. This information could be used to distinguish between returns to PSE and returns to individual ability.

OCCUPATION, AGE, AND WORKPLACE EXPERIENCE

The ELMLP includes industry information provided through the T1FF's 3-digit North American Industry Classification System (NAICS) codes. However, the LMIC (2018) acknowledges that the lack of National Occupational Category (NOC) codes prevents researchers from determining whether PSE graduates are working in jobs closely related to their fields of study. Although the Census includes data on occupation, this information is self-reported and only refers to occupation as of the census year in question. A dataset containing further information on field of employment could be linked to PSIS, which includes data on field of study. This linkage would allow for a more comprehensive analysis of the varying returns to PSE graduates, which could be higher for graduates working in jobs more closely related to their specialization.

Incorporating data on the age and workplace experience of PSE graduates would enable researchers to determine whether higher relative wages reflect the return to PSE or the natural progression of the labour market. Researchers could look at tax records, and in particular the eligible EI claims present in tax records, to estimate workplace experience. This information would also enable a distinction between individuals who delayed or did not delay PSE enrolment after high school, which could impact earnings post-graduation. Because older PSE graduates/workers are more likely to have spent more time gaining experience in the labour force, they are more likely to see higher earnings than new graduates/workers in the same field, and accounting for these differences could have a large impact on estimated outcomes.

NON-MONETARY BENEFITS

Finally, we recommend that future studies consider the potential non-monetary benefits of PSE. While it may be difficult to measure non-material benefits such as the prolonged enjoyment of learning, certain non-monetary elements – such as health, longevity, civic participation, etc. – reflect a combination of both private and public returns. Excluding this data could lead to an underestimation of returns to PSE, especially when a traditional cost-benefit analysis is employed.

SUMMARY AND DISCUSSION

It is a widely accepted fact that investments in PSE yield positive returns. Existing literature and government publications back up this assumption, though there is no official consensus on the most reliable methods for calculating the returns. Yet, while the majority of existing reports implement different sources of data, control groups, and systems of measurement, researchers are moving toward more aggregate approaches. Echoing the work of Finnie and Usher (2005) and Frenette et al. (2014), the ELMLP effectively promotes the combination of multiple datasets as the most comprehensive way to compile information on Canadians' transitions from higher education into the labour market. However, the platform faces certain limitations, including a lack of sufficient earnings comparison groups.

Analyzing the existing literature and currently available datasets, we compile a list of recommendations we identify as the most accurate and effective methods for addressing this limitation and calculating returns to PSE. In particular, we suggest analyzing additional datasets to generate estimates on the earnings of high school graduates, high school non-completers, and PSE graduates who earned their credentials abroad—all cohorts which are currently not covered by the ELMLP, but which could be used in the development of earnings comparison groups.

Overall, our immediate recommendations include obtaining control groups from the datasets already linked through the ELMLP and creating aggregated cohorts of earnings profiles based on observable characteristics across multiple datasets. We also suggest using administrative tax records to estimate individuals' participation (or non-participation) in PSE. This step could be accomplished over a relatively short period of time. Finally, we propose expanding the information currently accessible through the ELMLP with additional datasets, including the Census, LAD, LISA, GSS, PIAAC, IMDB, and others. Over time, linking these datasets to the ELMLP would provide researchers access to missing information, enabling the formation of more applicable control groups and thus producing more accurate estimations of the return to PSE in Canada.

REFERENCES

Autor, D. H. (2014, May). Skills, education and the rise of earnings inequality among the "other 99 percent." Mass: American Association for the Advancement of Science. Retrieved from https://dspace.mit.edu/handle/1721.1/96768#files-area

Bar-Or, Y., Burbidge, J., Magee, L., & Robb, A. L. (1995, October). The wage premium to a university education in Canada, 1971-1991. *Journal of Labor Economics 13*(4), 762-794. Retrieved from http://web.a.ebscohost.com.proxy.hil.unb.ca/ehost/pdfviewer/pdfviewer?vid=4&sid=e86266ab-d5d8-4e4b-b9d5-d01940ddf1d1%40sessionmgr4008

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999, March). Human capital investment: The returns from education and training to the individual, the firm and the economy. *Fiscal Studies 20*(1), 1-23. doi:10.1111/j.1475-5890.1999.tb00001.x

Boothby, D., & Drewes, T. (2006). Postsecondary education in Canada: Returns to university, college and trades education. *Canadian Public Policy 32*(1), 1-21. doi: 10.2307 / 3552240

Boothby, D., & Drewes, T. (2010, August 24). The payoff: Returns to university, college and trades education in Canada, 1980 to 2005. Retrieved from *C. D. Howe Institute*: https://ideas.repec.org/p/cdh/ebrief/104.html

Boothby, D., & Rowe, G. (2002, June 26). Rate of return to education: A distributional analysis using the LifePaths model. Working Paper W-02-8E. Retrieved from Applied Research Branch Strategic Policy Human Resources Development Canada: https://files.eric.ed.gov/fulltext/ED505310.pdf

Boudarbat, B., Lemieux, T., & Riddell, C. (2010). The evolution of the returns to human capital in Canada, 1980-2005. *Canadian Public Policy 36*(1), 63-89. Retrieved from http://www.jstor.org.proxy.hil.unb.ca/stable/20695157

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review 75*(2), 273-302. doi: 10.1177/0003122410363567

Caponi, V., & Plesca, M. (2009, August). Post-secondary education in Canada: Can ability bias explain the earnings gap between college and university graduates? *Canadian Journal of Economics 42*(3), 1100-1131. doi: 10.1111/j.1540-5982.2009.01540.x

Cohen, E., & Addison, J. T. (1998). The economic returns to lifelong learning in OECD countries. *Education Economics 6*(3), 253-307. https://doi.org/10.1080/09645299800000021

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives 27*(3), 79-102. http://dx.doi.org/10.1257/jep.27.3.79

Dee, T. S. (2004, August). Are there civic returns to education? *Journal of Public Economics 88*(9-10), 1697-1720. https://doi.org/10.1016/j.jpubeco.2003.11.002

Dodge, D. A., & Stager, D. A. (1972, May). Economic returns to graduate study in science, engineering and business. *The Canadian Journal of Economics 5*(2), 182-198. doi: 10.2307/133885

Ferrer, A. M., & Menendez, A. (2009, October 31). The returns to flexible postsecondary education: The effect of delaying school. Retrieved from *Semantic Scholar*: https://pdfs.semanticscholar.org/6ff2/6503dbe8ad36f143e3fa525b801a6686fce5.pdf

Ferrer, A. M., & Riddell, C. R. (2002, November). The role of credentials in the Canadian labour market. *Canadian Journal of Economics 35*(4), 879-905. Retrieved from http://www.jstor.org.proxy.hil.unb.ca/stable/3131840

Finnie, R., & Usher, A. (2005, April). Measuring the quality of post-secondary education: Concepts, current practices and a strategic plan. Ottawa, Ontario: Canadian Policy Research Networks, Inc. Retrieved from https://immagic.com/eLibrary/ARCHIVES/GENERAL/CPRN_CA/C050407F.pdf

Frenette, M., Grekou, D., & Wannell, T. (2014, April). Enhancing research opportunities on the returns to adult learning with National Survey and administrative data sources. Social Research and Demonstration Corporation. Retrieved from https://www.deslibris.ca/ID/242872

Green, D. A., & Worswick, C. (2012). Immigrant earnings profiles in the presence of human capital investment: Measuring cohort and macro effects. *Labour Economics 19*, 241-259. doi: 10.1016/j.labeco.2011.11.002

Grubb, W. N. (1993). The varied economic returns to postsecondary education: New evidence from the class of 1972. *The Journal of Human Resources 28*(2), 365-382. doi: 10.2307/146208

Handa, M. L., & Skolnik, M. L. (1975, February). Unemployment, expected returns, and the demand for university education in Ontario: Some empirical results. *Higher Education 4*(1), 27-43. Retrieved from https://www-jstor-org.proxy.hil.unb.ca/stable/344601

Kane, T. J., Rouse, C. E., & Staiger, D. (1999, July). Estimating returns to schooling when schooling is misreported. (*NBER* Working Paper No. 7235). National Bureau of Economic Research, Inc. Retrieved from https://www.nber.org/papers/w7235

Kirby, D. (2007, November 3). Reviewing Canadian post-secondary education: Post-secondary education policy in post-industrial Canada. *Canadian Journal of Educational Administration and Policy 65*(3), 1-24. Retrieved from https://files.eric.ed.gov/fulltext/EJ806980.pdf

Lemieux, T. (2006). Postsecondary education and increasing wage inequality. *American Economic Review 96*(2), 195-199. doi: 10.1257/000282806777211667

Lemieux, T. (2014, November). Occupations, fields of study and returns to education. *The Canadian Journal of Economics 47*(4), 1047-1077. Retrieved from https://www.jstor.org/stable/43818764

Labour Market Information Council. (2018, December). Exploring data to assess labour market outcomes of post-secondary students and apprentices. *LMI Insights 4*. Retrieved from https://lmic-cimt.ca/wp-content/uploads/2018/12/LMI-Insights-No.-4.pdf

Reano-Alvarez, G. (2012, February). The economic returns to receiving higher education outside Canada. (*RDC* Research Paper No. 37). McMaster Research Data Centre. Retrieved from https://www.deslibris.ca/ID/233625

Stager, D. A. A. (1996). Returns to investment in Ontario university education, 1960-1990, and implications for tuition fee policy. *The Canadian Journal of Higher Education 26*(2), 1-22. Retrieved from http://journals.sfu.ca/cjhe/index.php/cjhe/article/view/183235/183198

Stark, A. (2007, February). Which fields pay, which fields don't? An examination of the returns to university education in Canada by detailed field of study. (Working Paper No. 2007-03). Government of Canada Department of Finances. Retrieved from https://www.fin.gc.ca/pub/pdfs/wp2007-03e.pdf

Statistics Canada. (2006, September 1). University tuition fees. Retrieved from https://www150.statcan.gc.ca/n1/daily-quotidien/060901/dq060901a-eng.htm

Statistics Canada. (2017, June 16). Use of administrative data. Retrieved from https://www150.statcan.gc.ca/n1/pub/12-539-x/2009001/administrative-administratives-eng.htm

Statistics Canada. (2018a, September 5). Tuition fees for degree programs, 2018/2019. Retrieved from https://www150.statcan.gc.ca/n1/daily-quotidien/180905/dq180905b-eng.htm

Statistics Canada. (2018b, December 17). Table 18-10-0005-01 Consumer Price Index, annual average, not seasonally adjusted. Retrieved from https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810000501#timeframe

Vaillancourt, F. (1986, September). The returns to university schooling in Canada. *Canadian Public Policy 12*(3), 449-458. doi: 10.2307/3550608

Vaillancourt, F. (1995, August). The private and total returns to education in Canada, 1985. *Canadian Journal of Economics 28*(3), 532-554. doi: 10.2307/136047.

Walters, D., White, J., & Maxim, P. (2004, September). Does postsecondary education benefit Aboriginal Canadians? An examination of earnings and employment outcomes for recent Aboriginal graduates. *Canadian Public Policy 30*(3), 283-301. doi: 10.2307/3552303

APPENDIX

Table 1a: Summary of ELMLP Dataset

Latest Year	2016
Cycle	Annual PSIS (2009/10 - 2015/16) RAIS (2008 - 2016 T1FF (2004 - 2016)
Target	Individuals in PSE in Canada
Education	Postsecondary education in Canada
Linking	Linked to T1FF
Timeline	Immediate
Advantages	Administrative Database Detailed information on postsecondary education Duration of degree/diploma Information on incomplete postsecondary education Current Enrollment Immigration Status of students
Disadvantages	Missing control group with high school education or less Missing Canadian Residents and Immigrants with overseas postsecondary education
Notes	

Table 1b: Summary of PSIS / RAIS Dataset

Latest Year	2015/16 (PSIS) 2016 (RAIS)
Cycle	Annual PSIS (1999/2000 - 2017/2018) RAIS (1998 - 2017)
Target	Individuals in PSE in Canada
Education	Postsecondary education Incomplete postsecondary education
Linking	Not reversely linked with T1FF
Timeline	Short-Term
Advantages	Administrative Database Duration of degree/diploma completion Information on incomplete postsecondary education Current Enrollment Immigration Status of students
Disadvantages	Missing control group with high school education or less Missing Canadian Residents and Immigrants with overseas postsecondary education
Notes	Reverse linking, PSIS/RAIS to T1FF to identify control group Linking to IMDB would allow identifying immigrants and their education profile

Table 1c: Summary of LAD / IMDB Dataset

Latest Year	2016
Cycle	Annual LAD (1982 - 2016) IMDB (1980 - 2015)
Target	All individuals who filed federal tax return
Education	Education of immigrants before landing
Linking	LAD is linked with IMDB
Timeline	Short-Term
Advantages	Administrative data Distinguishing between Canadian-born residents and immigrants with overseas education
Disadvantages	Not possible to distinguish between immigrants and Canadian-born prior to 1980 Not possible to identify Canadian born residents with overseas education Canadian residents with overseas education below bachelor's degree or who did not file for tax credits would be unobserved
Notes	Linking LAD to ELMLP would enable to observe control group based on education and deduction tax returns

Table 1d: Summary of Census Dataset

Latest Year	2016
Cycle	5 Years
Target	All residents of Canada as of the Census date
Education	Highest education level attained Major field of study available Location of highest level of education
Linking	2016 Census is linked with T1 and CRA records
Timeline	Short-Term
Advantages	Large sample size Indicator for education location Immigration status available
Disadvantages	Snapshot of earnings for Census year Self-reported education is as of the Census year 2016 Years of education is not reported Incomplete degrees/diplomas cannot be observed
Notes	Earlier Census years including National Household Survey 2011 are also available for linkage.

Table 1e: Summary of LISA Dataset

Latest Year	2016
Cycle	2 years (2012, 2014, 2016)
Target	Individuals aged 15 years and over
Education	Highest education level attained Current Enrollment for reference year Years of formal education
Linking	2016 linked with T1FF 1982 – 2015 Previous two waves (2012, 2014) are also linked with T1FF
Timeline	Short-Term
Advantages	Overall linkage rate is high. Balanced panel dataset can be created Relatively easy to identify control groups Indicator for education location Scope to identify multiple degrees/certification Immigration status available
Disadvantages	Sample size is 30,000, control group could be smaller than other datasets like Census. Linkage rate declines with previous years, increasing linkage rate by excluding certain groups could result in loss of losing some data No direct indicator to identify incomplete PSE
Notes	Balanced Panel Data is high linkage rate with less years 30-year panel - 74.3% linkage rate and over 6,500 respondents, 5-year panel - 88% linkage rate and over 16,500 respondents

Table 1f: Summary of GSS Dataset

Latest Year	2016
Cycle	Annually (1989 – 2016)
Target	Individuals aged 15 years and over
Education	Highest education level attained Current Enrollment for reference year Indicator for incomplete PSE, without specifying number of years
Linking	2015 and 2016 cycles are linked to historical T1FF GSS 2011 linked to T1FF 2013 and Immigrant Landing File 2014
Timeline	Short-Term for GSS 2016, 2015, and 2011 Intermediate-Term for other GSS cycles
Advantages	Detailed education information, including incomplete PSE Relatively easy to identify control groups Education profiles of parents and spouse available in some cycles Immigration status available
Disadvantages	Sample size is approximately 25,000, hence size of control groups could be smaller GSS cycle 2002 and 2007 only targets respondents aged 45 years and over Indicator for education location not available in all cycles
Notes	For previous cycles, linking using probabilistic matching can be requested.

Table 1g: Summary of PIAAC Dataset

Latest Year	2012
Cycle	Next cycle in 2021
Target	Individuals aged 16-65 years
Education	Highest education level attained Years of formal education
Linking	Not linked with T1FF or any tax files yet.
Timeline	Intermediate-Term
Advantages	Detailed education information. Relatively easy to identify control groups Current enrollment. Indicator for education location
Disadvantages	Snapshot of self-reported education information available until 2012 No direct indicator for incomplete PSE.
Notes	Deterministic hierarchical Linkage can be requested, as it was done with Census, NHS, and IRCC data.

Table 2: Summary of Literature Review

Study	Dataset(s)	Method(s)	Response
Dodge & Stager (1972)	1967 Highly Qualified Manpower Survey	Cost-benefit analysis of net present value of investment in PSE according to costs of graduate education and PSE graduates' annual earnings differentials until retirement	Earnings differentials are not a good proxy for output differentials. Estimates are only accurate insofar as bachelor's and graduate degree-holders are alike in all respects other than education level
Vaillancourt (1986)	Statistics Canada's 1981 Survey of Consumer Finance and Financial Statistics of Education	Uses cross-sectional data and OLS estimates to calculate age-earnings profiles for men who were 18 years old in 1981. Compares estimated future earnings of males without university schooling and those with at least some PSE	The 1981 datasets are limited, as they do not provide standardization of schooling characteristics, such as field and length of studies and degrees held
Bar-Or et al. (1995)	Canadian Survey of Consumer Finances (SCF)	Compares the full-time earnings premiums of a university education to that of 11-13 years of education	Data from 1970s relates to families, whereas later data relates to individuals. SCF data makes it difficult to identify high school graduates "cleanly"
Walters et al. (2004)	1995 National Graduates Survey (NGS)	Uses regression and compares the labour market outcomes of Aboriginal, visible minority, and other Canadian graduates from university, college, and trades school	Database lacks distinction between status and non-status First Nations & does not address health, culture, and ability characteristics
Caponi & Plesca (2009)	1994 Canadian General Social Survey (GSS)	Uses data on parental education to control for selection bias	Propensity score matching is more reliable than the IV estimator
Ferrer & Menendez (2009)	1997 Canadian National Survey of Graduates & 2000 Follow-up Survey	Compares 1997 log wages between single-degree continuing graduates, single-degree delayed graduates, multiple-degree continuing graduates, and multiple-degree delayed graduates	The assumption that schooling proceeds in a linear, uninterrupted fashion from primary school onward is convenient but not accurate
Lemieux (2014)	2005 NGS & 2006 Canadian Census	Measures "relatedness" between job and field of	NGS measures "relatedness" between jobs

		study of PSE and high school graduates. Constructs high school control group by using a dummy variable of 1 (for high relatedness) or 0 (for low relatedness) or by adopting the low level of relatedness seen in the humanities (51.4%)	and education but does not include information on individuals without PSE. Census contains more information about earnings of population but does not consider "relatedness"
**Grubb (1993)	National Longitudinal Survey for class of 1972 (NLS72)	Uses a linear functional form to distinguish between returns to credential completers and non-completers	NLS72 limited by only containing information on one cohort of students. Incomplete transcript data results in self-reported data, possibility of bias
**Kane et al. (1999)	Census Bureau & NLS72 & Post-secondary Education Transcript Survey (PETS)	Implements "method of moments" technique to identify measurement error in self-reported and transcript-reported schooling, as well as the implications these errors have for OLS and IV estimates of the return to PSE	Census questions regarding educational level are preferable to those of the NLS72
**Dee (2004)	High School and Beyond (HS&B) longitudinal survey & 1972-2000 GSS	Uses local proximity of 2-year colleges and measurements of newspaper readership as proxies for degrees of civic awareness	HS&B study faces limitations by only identifying civic returns at the PSE level and providing no measures of the degrees of civic engagement
**Lemieux (2006)	1972-74 May Current Population Survey (CPS) & 2003-2005 Outgoing Rotation Group (ORG) CPS	Uses data on male hourly wages to correlate wage inequality with increasing returns to PSE	Standard (log) wage equation does not sufficiently reflect that returns to PSE vary across all workers, so prefers human capital model with heterogeneous returns
Brand & Xie (2010)	National Longitudinal Survey of Youth & Wisconsin Longitudinal Survey	Judges between positive and negative selection hypotheses by invoking ignorability assumption and uses propensity score strata to analyze earnings effects of college completion. Estimates logistic regressions predicting propensity for PSE completion	It is implausible to assume that different members of a population respond identically to a treatment condition—in this case, college education

Vaillancourt (1995)	1986 Census individual microdata file	Calculates public/private rates of return by level and field of study by projecting lifetime earnings profiles against cost of schooling	Methodology overlooks other benefits and bases lifetime projections on single-year results. Data does not identify currently enrolled students
Stager (1996)	1991 Population Census of Canada & 1987 Profile of Post-Secondary Students in Canada	Cost-benefit analyses of mean earnings of university and high school graduates by 5-year age groups	Methodology overlooks innate ability; excludes non-monetary benefits; and does not control for different labour market structures
Boothby & Rowe (2002)	1991 Census data & Community College Student Information System/University Student Information System (CCSIS/USUS)	Uses Statistics Canada's LifePaths model to simulate lifetime earnings of pairs of high school and PSE graduates. Measures return by level and field of study	Individual characteristics, decisions about time use, or sheer luck might affect earnings potential. Methodology excludes tax/transfer system and allocation for time unemployed or out of labour force
Ferrer & Riddell (2002)	1996 Canadian Census	Uses Mincer's 1974 HCEF model, a "credentialist" model, and a hybrid model to compare high school, non-university PSE, and university graduates and dropouts	Rejects the separation of the HCEF and credentialist models, argues both dimensions should be included in empirical studies
Boothby & Drewes (2006)	1986-2000 Census data	Uses OLS estimates of a HCEF model to develop weekly earnings differentiations for the various levels of education to measure evolving earnings premiums	Notes importance of distinguishing between earnings premiums and rates of returns to investment in education, given the different costs associated education levels
Stark (2007)	1996 Census of Canada & Statistics Canada's Tuition Fees and Living Accommodation Costs at Canadian Universities Survey, 1995-96	Computes private returns to university graduates by level and field of study by considering impact of in/direct costs of PSE and income taxes	Dataset lacks information on earnings of high school graduates
Boothby & Drewes (2010)	2006 Census of Population	Measures financial returns and accounts for foregone earnings and time invested in education by assuming individuals	Notes the need for more comprehensive understanding of the length of studies and

		undertake full-time studies at 18 and graduate without interruptions	variable enrolment costs across jurisdictions
Boudarbat et al. (2010)	1981, 1986, 1991, 1996, 2001, and 2006 Census	Measures wage premiums to attending university vs. entering labour force post-high school. Reconciles differences from earlier studies	Census has limited information on annual hours of work, 2006 Census changed from previous versions, & the available data cannot control for motivation, ability, and perseverance
Reano-Alvarez (2012)	2006 Canadian Population Census long-form questionnaire (2B)	Uses weighted data & analyzes a subset (aged 25-59 years) whose PSE field of study is in government's list of targeted occupation	Data uses random sampling and lacks information on location of study for individuals who did not complete beyond high school
**Autor (2014)	Census Bureau P-60 (1979-1991) and P-25 (1992-2012)	Converts data on median earnings of full-time workers by education level into 2012 dollars by using the CPI-U-RS price series and into household income by comparing husband-wife earnings from each educational group	

***American studies*