Essays in Earnings, Academic Productivity, and School Competition

by

Hang Gao

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Applied Economics

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Christine Neill Associate Professor
Supervisor(s)	Anindya Sen Professor
Internal Member	Ana Ferrer Professor
Internal-external Member	Martin Cooke Associate Professor
Other Member(s)	Emmanuelle Pierard Associate Professor

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contribution

While the main ideas in Chapter 1 and 2 were developed jointly by my supervisor, Anindya Sen, and myself, I have made the major contribution to the work involved in producing the results of Chapter 1 and 2. A subset of data used in Chapter 2 was provided by Aniruddh Sachdev.

Abstract

This thesis consists of three self-contained essays evaluating current issues in earnings, academic productivity, and school competition.

The first chapter, coauthored with Anindya Sen, looks at returns to post-secondary education and the gender gap in Ontario. We construct a unique individual level panel dataset consisting of earnings of public sector employees of the Government of Ontario, facilitated by the Ontario Salary Disclosure Act which reveals earnings of $100,000 or more. Individual earnings from 2005-2013 were merged with publicly available profiles on www.linkedin.com, which contains details on educational attainment, field of study, job experience, and specific occupation. There are significant field specific differences in returns to post-secondary education. In terms of graduate education, on average, while Ph.D.'s earn a premium relative to undergraduates, there is a modest gender gap in earnings of doctoral degree holders, which is not present among undergraduates. The sample period also experienced significant salary increases for female undergraduates. However, there are significant gender differences in the proportion of individuals who are managers and also in earnings of senior managers belonging to early cohorts.

By creating and utilizing a unique panel data from several different sources including the Ontario Ministry of Finance, EconLit, Web of Science, Online CVs, and so forth on all tenured and tenure track professors in 16 Ontario economic departments over 1996 to 2012, the second chapter intends to analyze the pay and position of those professors to see how co-authorship affect an economist's research productivity and how research productivity impacts pay and promotion. The study demonstrates that there is a significant return to co-authored publications relative to solo-authored publications in Ontario universities. The investigation of the relationship between co-authorship and productivity reveals that co-authored publications are associated with higher citation counts. Our research has also demonstrated that higher quality publications have a greater effect on salary, and the likelihood of promotion is positively associated with past performance. The estimates also suggest that some gender differences exist concerning the impact of co-authored publications on the likelihood of promotion. Finally, we find that in Ontario, economists are more likely to co-author with their colleagues,who have the similar ability, experience, and research interest. We found no gender-sorting effect among Ontario economists.

In the last chapter, I use a data set obtained from the Ontario Ministry of Education and the Educational Quality and Accountability office (EQAO) to estimate whether average school performance is affected by competition from other nearby schools. The availability of data on a panel of schools allows me to control for the potentially confounding effects of unobserved school specific attributes. I employ fixed effects, random trend and Instrumental Variables estimation to eliminate the potential simultaneity bias associated with competition between schools. Following Gibbons, Machin and Silva (2008), I use proximity to school board boundaries as an instrumental variable for local school competition. IV estimates suggest a statistically insignificant association between school competition and school performance. Another important finding is that the estimated coefficient is stronger when the sample is restricted to the Toronto District School Board, which may suggest that competition may improve school performance where students are given more freedom to choose their school. This finding may lend support to the current policy which is designed to improve public school performance in Ontario.

Acknowledgements

I wish to express my gratitude to all the people who helped me in making this dissertation possible.

Firstly, I would like to express my deepest gratitude to my supervisor, Professor Anindya Sen, for his excellent guidance, continuous support and immense knowledge. It was an honor for me to work with him and he broadened my knowledge base, provided me with many insightful suggestions and helped sharpen my skill set. Without his guidance and persistent help this dissertation would not have been possible.

I would also like to give special thanks to my dissertation committee, Professor Ana Ferrer and Professor Emma Pierard, who have provided me with invaluable advice, untiring support and guidance. They were always on-hand to answer my questions and provide useful feedback.

Thank you to the entire department at University of Waterloo. All of the faculty and staff that I met throughout my graduate studies were wonderful people. I am grateful to the following professors: Matthew Doyle, Jean Guillaume Forand, DingHai Xu, Pierre Chausse, Tao Chen, Philip Curry, Francisco M. Gonzalez, Jean-Paul Lam, Thomas Parker and Mikal Skuterud, for their unfailing support and assistance in my graduate studies. Warm thanks are also due to Pat Shaw for her constant help. I am also grateful to the other staff of our department, including Pat Gruber, Kayla McKinnon, and Becky Moore. In addition, I would like to thank my friends and colleagues here for making my time in graduate school memorable. I am particularly grateful to Hongxiu Li, Yichun Huang, Sara Aghakazemjourabbaf, Allison Mascella, Ghazal Memartoluie, and many more.

And finally, last but by no means least, I would like to thank my parents for supporting me spiritually and financially. Thank you so much for listening, offering me advice, and supporting me through my life. I am also grateful to my little brother, who has been with me through all of life's ups and downs.

Thanks.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Returns to post-secondary education and the gender gap: panel data evidence from LinkedIn

1.1 Introduction

Understanding the long term effects of educational attainment on earnings and career advancement is of key policy relevance. A strong correlation between schooling and earnings provides a sound justification for government funding of and investment in education institutions as a mechanism for reducing income inequality. Educated individuals eventually pay governments back through taxes and are more likely to be engaged in society through civic participation, volunteering, and charitable activities [1]. In this respect, while there is a broad consensus that increased educational attainment is associated with progressively higher income, there is less agreement on the long run labor market consequences of more specific post-secondary choices such as the field of study, the university a student attends, and the pursuit of graduate studies. One reason is the difficulty of obtaining sufficiently

[1] Oreopoulos and Salvanes (2011) & Milligan et al. (2004)

detailed panel data, which are able to match such post-secondary educational choices with corresponding occupation, experience, and income over a relatively long time period.

We address these issues by creating a unique unbalanced panel data set of income and career advancement, which matches the salaries of public sector employees working for the Provincial Government of Ontario (Canada) with their profiles on the website linkedin.com. Data on income were compiled from salary disclosure reports facilitated by the 1996 Ontario Public Salary Disclosure Act, which reveals the earnings of all publicly employed individuals earning $100,000 or more. We employ available data from 2005 to 2013, which allows us to study income and promotion dynamics over time for a reasonable sample of individuals. The resulting data-set enables us to evaluate the differential effects of field specific educational attainment across cohorts who graduated during the nineteen-eighties, nineteen-nineties, and early two thousands. Pooling data across individuals allows us to obtain estimates of the average returns to different post-secondary educational choices in terms of specific fields and to the pursuit of graduate studies, while controlling for occupation and experience. On the other hand, the time-series variation available from the panel component of our data enables us to assess the marginal return of further educations. These features of our data are important given recent findings by Lemieux (2014), which suggests that a significant portion of the earnings premium from education can be attributed to an individual's occupation, field of study and the match between these two factors. The impacts of post-secondary choices are estimated with respect to not only income but promotions as well. The ability to construct a panel of individuals over time allows us to control for the potentially confounding effects of time-invariant characteristics that would otherwise result in biased estimates.

Our data also permit us to explore other issues of contemporary policy relevance. There has been much discussion on the gender wage gap in both the U.S. and Canada, with recent studies suggesting that on average, women earn between 70 to 80 cents for each dollar earned by men. There have also been claims that the gap has recently been widening [2] [3]. However, there remains much ambiguity on the source of these gaps, in terms of

[2] http://www.cbc.ca/news/business/wage-gap-oxfam-1.3478938.

[3] http://money.cnn.com/2016/03/23/pf/gender-pay-gap/index.html

the key occupations and fields of study that might be contributing to male-female salary differentials. The panel nature of our data allows us to study the gender wage gap by pooling together males and females in similar occupations within the public sector and with comparable educational backgrounds and years of experience, in an effort to understand when such a gap occurs and how it might change over time. The ability to control for occupations as well as educational attainment is critical as recent studies suggest that the gender wage gap diminishes once these factors are controlled for.

Our contribution is also premised on a careful analysis of the returns to education and career advancement of graduates from different fields. There has been much recent debate on the economic returns to pursuing post-secondary education in the humanities, social sciences, and fine arts. For example, a recent CIBC report [4] suggests that while, on average, commerce, engineering, math, computer and physical science graduates have a roughly 75%-117% wage premium to high school graduates (after taking into account degree costs paying for their degrees), the returns to a fine arts degree is less than that of a high school diploma, with a humanities degree yielding a 23% premium. Further, the difference in employment rates between undergraduates and high school diploma holders has considerably diminished over time. Certainly, there has been a significant negative perception on the economic value of a humanities and/or fine arts education. On the other hand, a recent study released by The Association of American Colleges and Universities and the National Center for Higher Education Management Systems based on more than three million observations find that, while science, math and engineering majors typically had higher salaries throughout their careers than those majoring in humanities and social sciences graduates in such disciplines, on average, had higher earnings during their mid-fifties.

There are, of course, limitations to our data that we acknowledge. Our data only consists of relatively high income earners in the public sector. However, there are very few studies that have focused on top earners, and from a government perspective, the ability to identify the determinants of high incomes is important in terms of developing policies, which facilitate and encourage movement to such occupations. There has certainly been

[4] Tal and Enenajor (2013)

some recent emphasis on the behavior of top income earners in the U.S. and Canada. We also acknowledge other concerns on whether our data accurately reflects broad population characteristics and whether the sample of individuals posting their resumes on LinkedIn is in any way, significantly different from peers who choose not to use LinkedIn. However, we think that the possibility of sample selection bias is relatively limited as a variety of sensitivity tests suggest our data to be broadly representative.

Our empirical estimates offer several findings, which shed more light on the returns from specific fields of study, gender, and graduate studies. The remainder of our paper is structured as follows. The next section contains a brief review of the literature. Section 3 discusses the creation of our unique dataset. Our empirical strategy is detailed in section 4. Section 5 discusses our main findings. Section 6 concludes the paper with a summary of key points.

1.2 Literature

1.2.1 Canadian Literature

There is an extensive literature on the average returns to higher education. In general, these studies exploit cross-sections of individuals for a specific year and estimate the wage premium associated with a post-secondary education relative to completing or dropping out of high school. An individual's education is defined as the highest level of educational attainment at the time of the survey. Therefore, the returns to a specific degree or program are estimated by pooling individuals with varying levels of education and estimating the average earnings associated with different degrees, relative to a chosen benchmark.

In this respect, there is an abundance of Canadian research that have estimated the returns to education from the nineteen-seventies onwards. Freeman and Needels (1991) and Burbidge et al. (2002) suggest that the wage gap between more and less educated workers remained stable during the nineteen-eighties and nineties. Specifically, Burbidge et al. (2002) find stable trends in university wage premiums for males from 1981-2000 but obtain decreasing premiums for females [5]. However, as noted by Boudarbat et al. (2010), the above studies do not take into account experience, which can result in a significant downwards bias in wage premium estimates.

Other studies suggest an increasing premium for more educated workers. While Bar-Or et al. (1995) do find that the university wage premium (relative to those with 11-13 years of education) declined during the nineteen-seventies for most workers, it actually increased for young males over the nineteen-eighties and early nineteen-nineties. Beaudry and Green (1998) and Lemieux and Card (2001) also obtain evidence of increases in the university wage premium during the nineteen-eighties and nineties. Boothby and Drewes (2006) investigate the magnitude of university, college, and post-secondary certificate wage premiums from nineteen-eighty to two thousand. Their results suggest a larger premium for university relative to college degrees for both men and women. However, they also emphasize the lower costs of a college education.

Boudarbat et al. (2006) pool Canadian Census data from 1980-2000, and find that the university wage premium increased for both males and females, but more modestly for females. These results are robust to the inclusion of controls for years of experience. In an update to this study, Boudarbat et al. (2010) conclude that while the university wage premium increased considerably for both Canadian men and women over the 1980-2005 time period, the increase was more significant for men[6]. In general, much of the recent Canadian literature seems to be in agreement that there was an increase in the university wage premium from the nineteen-eighties to the early two thousands, despite the considerable increase in labor supply over the same time period. The increase corresponds with a higher demand for skilled individuals stemming from globalization and technological

[5]Using Survey of Canadian Finance (SCF) data, Bar-Or et al. (1995) look at the wage premium to a university education in Canada from 1971 to 1991, they show that the university premium decreased for males during the 1970s and remained stable during the 1980s, but it is hard for them to observe any reliable trends for females due to the noisiness in the data; Burbidge et al. (2002) also find the similar stable trends in university education for males during the 1980s and 1990s based on both Labor Force Survey (LFS) and The Survey of Labor Income Dynamics (SLID) evidence, in addition, they demonstrate the wage premium has declined for females.

[6]However, it is important to acknowledge that consistent with other studies, Boudarbat et al. (2010) find higher returns to education for women.

progress.

Ferrer and Riddell focus on identifying “sheepskin effects” by employing waves of the Census from 1981 to 2000 and separately estimating the effects of years of schooling from degrees and diplomas received. Sheepskin effects are defined as the gain in earnings associated with receipt of a degree, controlling for years of schooling. Therefore, corresponding coefficient estimates yield the on average difference in earnings between degree holders and non-completers with the same years of schooling. Ferrer and Riddell (2002) conclude that sheepskin effects are also evident in the Canadian labor market, with magnitudes similar in size to those estimated for the U.S. For native-born Canadians degree receipt is associated with higher earnings even after controlling for years of schooling and other influences. Ferrer and Riddell (2008) find significant sheepskin effects for natives and immigrants to Canada.

However, studies based on more recent data demonstrate that the university wage premium has not continued on an upward trend. Fortin and Lemieux (2015) and Frenette and Morissette (2014) use Labor Force Survey (LFS) data from the late nineteen-nineties and through the two thousands. Their estimates suggest that returns to university education started to decline sometime during the late two thousands. Foley and Green (2016) employ Census data and Labor Force Surveys from 1980-2013 and obtain similar results. Specifically, while they find that the wages of men in college and trades category and university categories from 1980-2000 increased relative to those of high school graduates, after 2000, the returns to a university degree began to fall. In contrast, the returns to a college or trade education continued to increase relative to returns to a high school diploma. However, the returns to education of prime aged women did not vary greatly between 1980 and 2013. Fortin and Lemieux (2015) and Foley and Green (2016) explain these findings by noting the rise in wages of male high school graduates in western Canada and the Atlantic provinces attributable to the resource boom in these provinces. As a result, the traditional penalty for dropping out of high school greatly diminished, with spillover effects on wages in non-resource sectors, as employees in these sectors were able to use higher wages in resource sectors in their own bargaining. On the other hand, wages in Quebec and Ontario mirrored corresponding trends in the United States with more subdued movements in wage differentials. Finally, Schirle (2015) studies the gender wage gap across provinces using the Labor Force Surveys and finds that a large portion of the wage gap in each province is explained by gender differences in industry and occupation. However, her study is not based on panel data.

In summary, the literature on the returns to education in Canada and based on the Census and Labor Force Surveys is quite comprehensive, with a focus on understanding gender specific differences in university wage premiums relative to high school and less than high school attainment. However, there are some gaps, particularly in our understanding of the returns to specific fields of university education and graduate degrees [7]. Our data allows us to match university degree choices to earnings and the likelihood of promotion. This is important as the returns to education go beyond estimating incremental earnings from higher levels of schooling. An alternative perspective is that individuals trained in specific fields may earn more rapid increases in income and are more likely to be promoted. Constructing a panel of individuals with salary information culled from the Ontario Salary Disclosure Act allows us to credibly identify such trends by exploiting time-series across cross-sections of individuals. Most Canadian studies on the returns to university education and specific fields have been unable to employ panels of individuals. We are unaware of other research that has defined and estimated the returns to further education through individual specific increases in income and the likelihood of promotion to management [8]. Finally, in some cases we are able to observe individual earnings before and after the completion of post-secondary diplomas, which allows us to identify the marginal gains to

[7] To the best of our knowledge, only Boothby and Drewes (2006) and Lemieux (2014) extensively study returns to specific fields.

[8] Among university graduates, there are also large differences in the earnings distribution across fields of study, as in the NGS data. Lemieux (2014) shows that arts graduates earn barely more than high school graduates and much less than engineering graduates. For example, 65% of engineering graduates are in the top earnings quantile, which is four times as large as the earnings quartile for high school graduates (17%) and three times as large as that for arts graduates (22%). Lemieux points out that most of this difference is explained, however, by occupational upgrading and match effects, and the pure return in the humanities is only 5 percentage points lower than it is in business and health. This suggests that the general skills provided by a degree in the humanities are as valuable as they are in health or in business. The difference is that humanities graduates are less likely to end up in high-paying occupations or in occupations that make good use of the specific skills they learned in university.

further educational attainment [9].

1.2.2 US Literature

There are numerous U.S. based studies that have investigated the returns to post-secondary education by focusing on the effects of years of schooling, awarded degrees, major field of study, and the quality of institutions. Card (1999) summarizes many of these studies. In general, there is an absence of papers that have relied on panels of individuals over several years. Most studies based on nineteen-eighties and nineties data use cross-sections of individuals and years of completed education as a measure of schooling based on U.S. Census data and the Current Population Surveys. The general consensus is that additional years of education are correlated with higher incomes [10].

However, a problem with exclusively relying on years of education as a measure of educational attainment, is that the associated coefficient estimate yields an on average effect of an additional year of schooling. As a result, coefficient estimates of additional years of education that are a result of a masters or doctoral degree would be biased downwards if graduate studies lead to a much higher marginal impact on earnings, relative to earlier years. In the late nineteen-eighties, the U.S. Census Bureau switched to measuring educational attainment through the highest degree achieved and no longer collected information on total years of education. As pointed out by Ferrer and Riddell (2002) this lack of comparable data on credentials and years of education made it difficult for researchers to evaluate the existence of sheepskin effects, which acknowledge that specific degrees have corresponding impacts that are independent of a person's years of education. Jaeger and Page (1996) and Park (1999) do investigate sheepskin effects but not through panels of

[9] It is also important to acknowledge Canadian studies (Gunderson (1979), Shapiro and Stelcner (1989), Prescott and Wandschneider (1999), and Mueller (2000) and Tiagi (2010) on public-private sector wage differentials. In general, most of these studies find that controlling for all else, public sector wages tend to be higher. However, most papers are based on limited time-series variation. As a result, coefficient estimates of the public sector wage premium may simply reflect other unobserved individual specific characteristics than the true monetary benefit of being a public sector employee.

[10] For example, Angrist and Keueger (1991) and Card (1993). Both studies rely on instrumental variables and imply that the return to an additional year of education to be roughly 10%.

individuals over a long period of time.

Other studies focused on the effects of specific degrees and the quality of post-secondary institution on earnings [11]. However, these papers are either based on graduates from a single university or predominantly cross-sectional surveys such as the Survey of Recent College Graduates and National Survey of College Graduates that follow graduates over a limited period of time [12]. Webber (2014) does use the 1979 and 1997 waves of the National Longitudinal Survey of Youth (NLSY79 and NLSY97), but these panel data sets are restricted to single cohorts [13]. Similarly, Arcidiacono (2004) uses data from a single cohort belonging to the National Longitudinal Study of the Class of 1972 (NLS72). Jepsen et al. (2014) estimate labor-market returns to community college diplomas and certificates by exploiting detailed administrative data from Kentucky, which allows them to match degrees to individual earnings from 2000-2008. However, the focus of the study is on community college. In general, there is an absence of U.S. based empirical research on the effects of field of study and graduate education based on panel data over a reasonable period of time.

1.3 Data

We created our data set by combining information downloaded from the Ontario Ministry of Finance [14] and LinkedIn website [15]. The harmonized data includes salary, education, experience and other individual-specific information on 6,406 government employees representing 6 public sectors and 418 sub-sectors from the academic year 2005/2006 to

[11] Fewer studies have attempted to investigate the effect of education on promotion in the United States. Spilerman and Lunde (1991) find mixed evidence for educational credentials on promotion.

[12] Examples of such studies are Rumberger and Thomas (1993), Del Rossi and Hersch (2008), and Graham and Smith (2005).

[13] Webber also uses data from the the 2014 American Community Survey (ACS); select March Current Population Surveys (CPS) and the 1993 and 2003 waves of the National Survey of College Graduates (NSCG).

[14] http://www.fin.gov.on.ca/en/publications/salarydisclosure/pssd/

[15] https://www.linkedin.com/

2013/2014. This section details the construction of each of our key variables and sensitivity tests designed to evaluate the presence of sample selection bias [16].

1.3.1 Variable Construction

Salary

Annual data on salaries of individuals earning more than $100,000 in Ontario's public sectors for the period 2005 to 2013 were downloaded on a year-by-year basis from the Ontario Ministry of Finance website. These data are publicly available through the 1996 Public Sector Salary Disclosure Act.

Table 1.1 documents the distribution of individuals for the beginning and end years of the sample. We possess salary data for 1,097 individuals in 2005 and 6,406 individuals for 2013. In nominal terms a majority (Approximately 82%) of individuals earned a salary of between $100,000-$120,000. By 2013, there is a perceptible shift towards higher incomes as there is a sizable increase in the proportion of individuals earning from $120,000-$180,000.

Table 1.2 documents the distribution of individuals by the number of years they are present in our data and nominal salaries averaged across years. A majority of our panel (almost 62%) is based on individuals for whom we possess five to nine years of data, with around 15% coming from individuals who are present for the entire duration of our sample. Average salaries follow an increasing trends with number of years with average salary being (approximately) $112,961 for employees on the list for two years and (approximately) $129,391 for those who are present in the entire sample.

Education

Each individual was then matched to her/his corresponding publicly available LinkedIn website. The match process was conducted by name and the specific government depart-

[16] Table A.2 in the Appendix documents the specific government agencies covered by the Ontario Public Sector Salary Disclosure Act while Table A.1 contains the distribution of employees by degree and government organization. Broadly speaking, public sectors in Ontario consist of the various Ministries, Legislative Assembly and Offices, Judiciary, Crown Agencies, Hydro One and Ontario Power Generation, Crown Agencies, Municipalities, Hospitals and Boards of Public Health, School Boards, Colleges and Universities. We drop individuals in the Legislative Assembly and Judiciary because of small sample sizes.

ment/agency/ministry, both of which are available from the Ontario Ministry of Finance and LinkedIn websites. The resulting data set consists of 6,406 individuals over time resulting in 35,070 observations. These individuals specifically report the time period and the field of study of each awarded degree above high school, and the institution name awarding each degree. We then categorize the educational background into three main different dimensions of educational achievement to examine the effect of education on earnings and career advancement: years of schooling, different educational credentials as well as field of study.

a. Years of schooling

Consistent with other studies, individual specific years of schooling S_i is calculated based on the following formula:

$$S_i = 12 + \sum_{j=1}^{n} s_j$$

where s_j is the number of years of each degree took. The above formula is based on the assumption that all individuals spent twelve years to complete primary and secondary school [17].

b. Credentials

All individuals in our sample possess education beyond high school. Hence, we construct five credential dummy variables (College, BA, BA and diploma or Double BA, Master, and PhD) to represent the highest level of educational attainment by each individual [18]. Roughly 85% of our sample have a university bachelor degree or an undergraduate degree with an additional college or university diploma or degree. Fifteen percent possess only a college degree with no other credentials. As documented in Table 1.3, a very large

[17]Since a majority of individuals do not post their education information below high school on their LinkedIn profile, we assume everyone spent twelve years to complete primary and secondary school.

[18]It is important to emphasize that college education in Canada is different from the United States, where “college” also usually implies attendance at a university. College education in Canada refers to post-secondary degrees or diplomas that are in most cases shorter than a four year university degree, and offer a more specialized professional or vocational education relevant to specific employment fields.

proportion of observations stem from individuals who have either a Masters (33.9%) or Ph.D. (6.88%) as their highest level of educational attainment.

c. Field of study

Each individual's field of study is defined by his/her highest completed post-secondary degree, diploma or certificate. Employing the 2011 Canada Classification of Instructional Programs (CIP), we group field of study into the following eight categories: Humanities, Health, Business, Engineering, Science, Law, Education and Other. As can be seen from the last row of Table 1.4, most observations in our sample are in Humanities (32.93%), followed by Engineering (17.13%), Business (14.42%) and Science (12.3%).

Experience

We construct actual labor market experience from each persons public LinkedIn web profile, which contains the name of employer, position, and the beginning and end dates of each position. Career experience is specifically calculated by taking into account the starting and ending point of each individuals different jobs posted on her/his LinkedIn profile. Most studies estimate experience by using the Mincer formula which is: experience = age - year of education - 6 or defined as the number of years since an individual obtained her/his highest degree. In this respect, our data allows a more precise measure of experience as not all people begin their career immediately after graduation. Lemieux also notes the importance of measuring experience accurately as opposed to relying on cruder proxies (such as age). Figure 1.1 graphs the distribution of individuals by years of experience averaged across the sample. A significant majority of individuals possess between thirteen to twenty four years of experience, with very few observations belonging to individuals with less (more) than five (thirty-one) years of experience.

Job Rank

Controlling for all else, individuals in a management position should earn more relative to other peers with similar education and years of experience. An inability to control for job rank might otherwise result in an upwards bias in coefficient estimates of educational attainment or years of experience [19]. We are able to determine whether an individual is a

[19] A recent study analyzing the relationship between pay and research productivity by Sen et al. (2014)

manager as salary data from the Ontario Ministry of Finance website also contains individual specific job titles. The description allows us to construct indicator variables to denote if an employee is either a senior or middle manager. This was accomplished by matching the employee's job title to senior and middle manager category descriptions available from the 2011 National Occupational Classification (NOC) established by Employment and Social Development Canada and Statistics Canada [20]. Thirteen percent of our sample are senior managers, while 32% are in middle management.

Gender

In our analysis, gender is primarily identified from an individual's LinkedIn profile image [21]. We employed free online tools available from GenderChecker and GenderGuess as a sensitivity analysis and in order to assign gender to individuals who do not post a LinkedIn profile picture. These websites have a large database of names that are derived from national census data for some countries, and use the data to suggest gender based on name. As a robustness test, we ran our primary regressions using only those who do have a LinkedIn profile pictures and did not obtain any significant differences with respect to our empirical estimates.

Visible minority

Studies based on Canadian data suggest that after controlling for other factors, being a visible minority is significantly correlated with lower earnings. However, there are differences across ethnicity, and the amount of pay difference relative to Caucasians, seems to be lower than results from U.S. based research [22]. Further, recent studies suggest a

showed the importance of taking into account job rank dummy variables in any analysis of earnings. They found that the only other controls which are consistently significant among all the specifications are the academic rank dummies.

[20] As detailed in its website (`http://www5.hrsdc.gc.ca/NOC/English/NOC/2011/AboutNOC.aspx`): "The National Occupational Classification (NOC) is the nationally accepted reference on occupations in Canada. It organizes over 40,000 job titles into 500 occupational group descriptions. It is used daily by thousands of people to compile, analyze and communicate information about occupations, and to understand the jobs found throughout Canada's labor market."

[21] approximately 45% of our sample post their pictures

[22] The data base derived from Census 2000 is available at `http://www.census.gov/topics/population/genealogy/data/2000_surnames.html`

downward trend in earnings differences between Caucasians and visible minorities. Strict regulation in terms of pay equity considerably lessens the likelihood that being a visible minority results in lower pay in public sector positions. Nonetheless, we attempt to capture visible minority status by using a database derived from the 2000 U.S. Census, which assigns likely ethnicity to individual names. The database enables us to categorize names belonging to one of the following six ethnicity categories: Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, non-Hispanic American Indian and Alaska Native and non-Hispanic Multiracial. Our indicator variable takes a value of 1 if a name is determined as non-Hispanic Black, non-Hispanic Asian, non-Hispanic American Indian and Alaska Native, and non-Hispanic Multiracial, and is 0 for individuals classified as non-Hispanic White. Care should be exerted in interpreting coefficient estimates of this variable given that it is derived from U.S. data.

1.3.2 Sensitivity Test

It is important to determine whether our data yields similar sample characteristics relative to other well established data sets employed by researchers addressing similar questions. As discussed in the literature review, the Canadian Census is the most common data set employed by researchers. The public use files of the 2006 Census do not specifically identify individuals in the public sector, and instead allow employees to specify their occupation as public administration (NAICS code = 20). We downloaded data for 476 individuals with reported earnings equal to or greater than wage $100,000 and who confirm their occupation as public administration and plotted the earnings distribution [23]. Figures 1.2 and 1.3 show the distribution compared against our 2005 data and 2005, 2006, and 2007 data. In both cases, the different distributions generated from the Census and LinkedIn data are quite similar and offer some reassurance that our LinkedIn data offers an accurate representation of earnings for high income public sector employees in Ontario [24].

[23] We employ data from the 2006 Census given the widely documented issues with the 2011 National Household Survey, which was not mandatory.

[24] It is important to note that a t test for equality of mean salaries from the two different data sets was rejected , which is unsurprising given the difference in peaks.

However, it is important to acknowledge the possibility of sample selection bias stemming from unobserved heterogeneity as our data consists of individuals who are willing to post their resumes on LinkedIn, and there is a significant proportion of public sector employees who have not shared their resumes[25]. For example, individuals who post their resume online might be more aggressive in searching for other employment opportunities and/or make networking connections, relative to persons with similar educational qualifications but who choose LinkedIn or other resume based websites. If this is true and such individuals experience enhanced job mobility, then coefficient estimates of the returns to education generated by our data will be biased upwards.

In order to explore this further we plotted the earnings distributions with and without LinkedIn resumes across different ministries and agencies. As can be seen from 1.4, earnings distributions are quite similar. Hence, we think that the potential for sample selection bias is relatively limited.

1.3.3 Salary differences by educational attainment, gender, and experience

Table 1.6 contains some exploratory analysis through a comparison of male-female wage differentials conditioned on experience and educational attainment [26]. We divide our sample by year of graduation of final post-secondary degree and categorize individuals on the basis of final post-secondary educational attainment before 1980, from 1981-1989, 1990-99, and from 2000 onwards. We also defined subsets by gender and whether the individual obtained a graduate degree.

The first observation is that there are statistically significant differences in salaries by gender for individuals graduating before 1980, with men earning more than women, irrespective of educational attainment. On average, the differentials range from approximately

[25] The total number of unique individuals with a salary equal to or exceeding \$100,000 from 2005-2013 is 154,045.

[26] The results in this table are based on a smaller sample of individuals whose gender identity were based on Genderchecker and LinkedIn pictures.

8%-10.5%. While, gender-specific differences persist for individuals who graduated from 1981-89, the magnitudes are much smaller. This trend continues for the 1990-99 cohort, with no statistically significant gender differences for individuals with an undergraduate degree and a very small premium (1.46%) for male graduate degree holders. There is a bit of a gender difference (approximately 2%-4%) for individuals who graduated with any post-secondary degree during the 2000s. The second observation is that the number of observations during the 2000s associated with exclusively an undergraduate degree for both males and females decreased dramatically, relative to corresponding sample sizes for the 1990s. These trends are consistent with the observed growth in enrollment in graduate degrees and the fact that employees with lower educational attainment have to possess more experience in order to earn higher salaries.

Table 1.7 yields some further insight by focusing on gender differences through sample averages by the number of years an individual is present in the sample. For example, the first row consists of the average starting and ending salaries of all individuals in our data set who are present for at least two years, and the corresponding growth in nominal salary. We also make separate calculations by gender. The sample means are conditioned on individuals who have five or more years of experience. We do not think that the choice of this threshold in experience is arbitrary, given that it captures a majority of our observations and allows us to define relatively large samples [27]. This sample also enables us to evaluate the salary growth of individuals who have been working for a few years.

There are some gender differences in average salaries, with a maximum of roughly $3,000 (approximately a 3% difference) for individuals present for eight years in the data. On the other hand, average salaries for both genders are almost equal for individuals who are in our data for six years. It is also important to note that the percentage increases in nominal salaries over time, as defined by the number of years they are in the data, are very similar for men and women across the different cohorts. Hence, while simple sample means does suggest some gender gaps in terms of salary levels, such gaps are not strongly present in terms of salary growth. The amount of average salary growth is quite significant and rises with each year we observe an individual in our data. For employees whom we

[27] Number of observations is 34,792, number of individuals is 6,334

possess data for 5-9 years, we calculate average salary growth ranging from 13% to 33%. This increase in nominal salaries is consistent with the trends evident in Table 1.6.

In summary, our sample consists of highly educated individuals, the vast majority of whose salaries we are able to observe for 5-9 years. On average, there seems to be significant time-series variation in nominal salaries. Most of these public sector employees possess more than a decade of work experience. Our data are quite refined as we are able to match individuals to educational attainment, occupations, and field of study. Some preliminary sample averages suggest the presence of a statistically significant gender gap in salaries for employees who achieved their highest level of education sometime during the nineteen-eighties. However, this gap has shrunk considerably for more recent graduating cohorts. Further, salary growth has been comparable for both males and females. The next section outlines a more formal empirical framework in order to disentangle the effects of educational attainment, field of study, experience, and gender on salaries and salary growth.

1.4 Empirical Specification

Following Ferrer and Riddell (2002) and other recent studies, we estimate the following extended Mincerian semi-log earnings function:

$$\ln W_{ist} = \beta_0 + \beta_1 * S_{is} + \beta_2 * EXP_{ist} + \beta_3 * EXP_{ist}^2 + \beta_4 * CD + \beta_6 * M + \alpha + T + \epsilon_{ist} \quad (1.1)$$

Where W_{ist} is the annual salary of individual i, at time t, in sector s; S_{is} is the years of schooling of the individual; EXP_{ist} is experience and EXP_{ist}^2 is experience squared; CD is a vector of credential dummies representing individuals with college, a double undergraduate degree or an undergraduate degree with a diploma, a Master's degree, or with a Ph.D., with possessing an undergraduate degree as the omitted category; M is a vector of indicator variables for major field of study; α is a sector specific fixed effect[28], it accounts for factors

[28]Science is the omitted category for field and crown agencies is the omitted category for sectors.

that differ across sectors but are time invariant; T is a vector of year fixed effect. We also include a gender dummy (1 = female, 0 = male), a visible minority dummy (1 = visible minority, 0 = white), and dummy variables for whether individual i obtained his/her highest degree from a Top 10 ranked Ontario university, other Canadian university or from a US university, with the omitted category being individuals with other Canadian university. ϵ_{ist} is an idiosyncratic error term. We use province-specific consumer price indices to convert the salary data into constant 1992 dollars. Unless otherwise stated, standard errors of all OLS estimates are clustered at individual level to control for within individual attributes which might be correlated over time.

While our empirical specification is similar to models used in previous studies, it is important to emphasize that the identification is quite different. Most other Canadian research rely primarily on cross-sectional data at a single point in time. Our panel data allows us to identify the returns to higher education not only through salaries earned by individuals who remain within specific educational categories throughout our sample, but also by pooling such individuals with others who choose to pursue more education through an additional diploma or a graduates degree.

1.5 Empirical Results

Table 1.8 contains baseline OLS estimates of the effects of education on individual earnings. Column (1) contains estimates from a parsimonious specification, which only includes measures for educational attainment. Years of experience and its squared are added in column (2). Dummy variables for field of study are included in column (3). Estimates in column (4) are conditioned on the above covariates as well as controls for management, visible minority, gender, and university. Column (5) contains estimates from a similar specification to column (4) but with the addition of years of education as a control for possible sheepskin effects. Column (6) adds separate interactions of the female dummy variable with the Master and Ph.D. dummy variables. Finally, column (7) contains estimates of a log-log specification, where all continuous variables are in natural logarithms. Standard errors of coefficient estimates are clustered at the individual level.

Empirical estimates in column (1) reveal the coefficient estimate of the Ph.D. dummy to be positive and statistically significant (at the 1% level). The specific estimate suggests that having a doctorate is associated with a 6.5% premium relative to an undergraduate degree. In contrast, the Master's dummy variable is statistically insignificant. Unsurprisingly, having a college degree results in a lower return (5.8% and statistically significant at the 1% level) relative to being a university undergraduate. However, having a double undergraduate degree or an undergraduate degree with a diploma is significantly correlated with a lower return. Adding years of experience in column (2) does not significantly alter the coefficient estimates of most of the covariates. Results contained in column (2) reveal that final educational attainment at the doctoral, college, and double undergraduate/diploma levels are all significantly associated (at the 1% level) with 9.2%, -6.2%, and -5% returns, respectively, relative to an undergraduate degree. The one difference is that the coefficient estimate of the Master's degree is now statistically significant (at the 5% level) with an implied return of 1.5%. Experience is also significantly correlated with earnings (at the 1% level), with each year associated with roughly a 1% return. However, the negative and statistically significant coefficient estimate of the squared experience covariate implies diminishing returns to these increases.

Results in column (3) confirm that the field of a final degree also significantly impacts earnings. Humanities graduates earn approximately 1.6% more than science graduates. However, this estimate should be viewed with caution given that it is only statistically significant at the 10% level. On average, business, health, law and engineering graduates also earn more than science graduates, with premiums of roughly 7.8%, 10.45%, 10.5%, and 2.9%, respectively. In contrast, the coefficient estimate of education is negative and implies a roughly 4% return lower relative to science degree holders. All these coefficient estimates are statistically significant at the 1% level.

Adding other covariates and sector and year fixed effects results in some changes. However, coefficient estimates are comparable to those in column (3). Years of experience remains positive and statistically significant (at the 1% level). Coefficient estimates of educational attainment dummy variables are also statistically significant (at the 1% level), with the exception of possessing a double degree/undergraduate. This covariate remains statistically significant across the remaining columns. Most of the field dummies are similar to corresponding estimates in the previous column. However, the engineering dummy is now statistically insignificant and remains so, across other columns. Having an educational background is now significantly correlated (at the 1% level) with higher earnings. In terms of the new covariates, being a senior manager is significantly associated (at the 1% level) with an approximately a 26.5% increase in earnings. The female and visible minority dummies are significantly (at the 1% level) correlated with lower earnings, as is graduating from a U.S. university (relative to a lower ranked Canadian/Ontario university).

The years of education variable in column (5) is positive statistically significant at the 1% level. Its inclusion does impact the Ph.D. and Master's dummies, which become reduced in magnitude and remain statistically precise. However, no other covariates are significantly impacted and we choose to focus on specifications without a separate covariate for the years of education, which in most cases, is time invariant across individuals and therefore, possibly capturing the effects of other unobserved time-invariant characteristics.

Relative to other columns, the difference in column (6) is the inclusion of the interaction of the female gender variable with Ph.D. and Master's dummy variables. The interactions are statistically significant (at either the 1% or 5% levels) and imply that females with a Master's (Ph.D.) earn roughly 3% (6%) less than male counterparts, controlling for all else. The female dummy variable itself is statistically significant at 1% and suggests that women with educational attainment lower than the graduate level experience a 1.75% penalty in earnings, relative to men. In terms of other key estimates, having a Master's (Ph.D.) is significantly associated (at the 1% level) with a 2.9% (10.5%) increase in earnings relative to an undergraduate degree. In terms of fields, having a degree in the Humanities, Business, Health, Law, and Education are all significantly correlated (at the 1% level) with a 2%, 4.6%, 7.8%, and 3.6% increase in earnings relative to individuals with a Science degree. As in other columns, senior managers earn roughly 26% more than employees who are not in management. Finally, estimates in column (7) are quite comparable to those in column (6), with no pronounced differences.

The above results point to the importance of educational attainment as well as specific fields of study. Their effects on salary growth are also of interest in terms of assessing the overall effects of education level and degree choice on labor market success. Therefore, we run similar regressions to those contained in Table 1.9, but with the dependent variable being defined as the percentage change in salary for each person over a three year period. This time period is chosen to generate some variation in salary changes and minimize any anomalous differences that might occur from changes calculated over consecutive years. The dependent variable is specifically generated by taking the first observation available for each person and then calculating a three year difference. The next observation is then the three year difference generated from the second year of available individual specific data and so on [29].

The empirical estimates are organized quite similarly to Table 1.8, with different columns reflecting the addition or omission of different covariates. In contrast to results in the previous table, years of experience possesses a negative sign across all columns. Coefficient estimates are statistically significant at either the 1% or 5% levels, and suggest that on average, individuals have actually experienced reduced real salary growth rates over the sample period. Coefficient estimates of different education levels are not consistently significant. The coefficient estimates of possessing a Ph.D. and college degree are negative and statistically significant (at the 5% or 1% levels) in some columns, implying lower growth rates for such degree holders. However, we place limited emphasis on this finding given that the coefficient estimates are not statistically significant across columns. Similarly, coefficient estimates of field covariates are not consistently significant across columns. In contrast, being manager is associated with significant growth rates with statistically significant coefficient estimates (at the 1% level) of roughly 0.54 and 5 for middle and senior management, respectively.

Perhaps the most striking results are the statistically significant coefficient estimates (at the 5% or 1% levels) of the female dummy variable, which suggest a growth rate of 0.43-0.79 percentage points each year. This finding is in contrast to the negative and statistically significant coefficient estimates of the female dummy variable in the previous

[29] In terms of a simple example, if we possess data from 2000-06 for a person, the first observation is the percentage change in salary from 2000-03 and the second observation is the percentage change in salary from 2001-04.

table. However, these results are confined to women with less than graduate level education. Specifically, results contained in column (6) reveal that the interaction between the Master's and Ph.D. dummy variables with the female dummies to be negative and statistically significant, while the corresponding interaction for the Ph.D. dummy variable is only weakly significant (at either the 5% or 10% levels). Nonetheless, the results do imply that salaries for certain women have risen at a faster rate than men with similar education levels.

Table 1.8 contains estimates of the effects of credentials and field of study on reported salary (converted into real terms). An alternative definition of the dependent variables would be to create categories based on salary levels that would allow an evaluation of possible factors, which might explain the presence of individuals in very high, high, medium, and lower earnings brackets. In order to explore this further, we created a variable that is 1 if the real salary is less than $100,000, 2 for earnings between $100,000 and $150,000, 3 for earnings between $150,000 and $200,000 and 4 if earnings are more than $200,000. The resulting specification is an ordered probit model. The marginal effects of coefficient estimates with respect to each category are reported in Table 1.10. The vast majority of observations are in the first category, with 20% in the second category, and with a much smaller proportion for individuals earning $150,000 or more in real terms.

Experience is clearly associated with higher earnings. In terms of credentials, final educational attainment with a college degree is significantly and negatively correlated with higher earnings categories. On the other hand, a Masters or a Ph.D. is significantly correlated with higher earnings levels. These coefficient estimates are statistically significant at the 1% level. Individuals with humanities, business, health, and law degrees are also more likely to be in higher categories, relative to science graduates. In each category, the magnitude of marginal effects are larger for health and law graduates, in comparison to persons with humanities and business degrees. Unsurprisingly, the senior management dummy is positive and statistically significant (at the 1% level) in higher categories. However, in the same specification, the middle management variables is sometimes negative and significant, but at the 10% level. Comparable to results in Table 1.8 the female and visible minority dummy variables are negative and statistically significant across most columns.

Returns to Specific Fields by Education Level & Gender

A relevant question is whether there are significant gender specific differences in earnings across fields and by educational attainment. In order to explore this possibility we ran regressions with interactions between the Ph.D. and Master's dummy variables and the field specific dummies. As the undergraduate degree and science dummies are omitted, the coefficient estimates of the degree covariates reflects earnings from science degrees. Table 1.11 contains the estimates. Column (1) contains results based on all data while columns (2) and (3) consist of findings with respect to men and women.

Coefficient estimates of experience are quite similar across columns, with both men and women receiving a little over a 1% increase in annual real earnings. Recall that the omitted variables are the undergraduate and science dummy variables. Therefore, coefficient estimates of the educational attainment covariates must be interpreted as the return to a science degree relative to an undergraduate degree in science. The coefficient of the Ph.D dummy in column (1) suggests that Ph.D. science degree holders, on average, enjoy a 12.75% return relative to undergraduates in science (statistically significant at the 1% level). However, there is a statistically significant gender difference (at the 1% level) as results in columns (2) and (3) reveal the return to be almost 14% for men but 8.64% for women. The coefficient estimate for Master's degrees in science is significant at the 10% level for men but insignificant for women. Coefficient estimates of the college dummy are negative and statistically significant (at the 1% level) across columns without significant gender differences.

Given the use of interactions of Ph.D. and Master's dummy variables with field dummies, the coefficient returns of the field dummies represent the return to a non-graduate level education in that field, relative to an undergraduate degree in science. On average, such graduates with a humanities degree experience a roughly 3.8% (significant at the 1% level) return relative to science graduates, and the corresponding premium to a business degree is 5.3% (significant at the 1% level). What is interesting is that the return to women is slightly higher for a humanities degree and about twice as much for women (relative to men) for a business degree. This trend continues for engineering degree holders as the coefficient dummy is statistically insignificant for men (in column (2)) but statistically significant for women (column (3)) at 5% with an estimate of 5.58. Similarly, the coefficient dummies for a law or education degree are statistically significant (at the 1% level) for women and larger in magnitude than corresponding estimates for men (in column (2)). The one difference is the coefficient estimate for a health degree for men, which is more than double that of women, but weakly significant at the 10% level. The above results demonstrate that there are clear differences in returns to education across fields. In terms of contemporary relevance, humanities undergraduates do not necessarily have significantly lower earnings than individuals in other fields. On average, they earn more than science and engineering undergraduates [30] Unsurprisingly, coefficient estimates of the business and law dummy variables (in column (1)) are the largest relative to the other fields.

However, empirical estimates are quite different when taking into account earnings by Ph.D. holders. The coefficient estimates of the interactions between the Ph.D. dummy and field dummies for humanities, business, and engineering are negative and generally significant for both men and women, implying that Ph.D. holders in science earn more, in comparison to Ph.D.'s in these specific fields. With respect to these fields, there are no significant statistical differences across gender, which can be verified by taking the coefficient estimates of the Ph.D. dummy from columns (2) and (3) and adding them to the corresponding Ph.D. interaction term of interest (from either column (2) or (3)). The one field, in which a Ph.D. can earn more than a Ph.D. in science, is health. And in this respect, there is a pronounced difference in coefficient estimates across gender, with male Ph.D.'s earning roughly 21% than science counterparts, while the corresponding interaction term between the Ph.D. and health dummies is statistically insignificant. Hence, the earlier observed negative and statistically significant interaction between the Ph.D. and gender dummies in Table 1.8, seems to be an artifact of specific fields, such as science and health. Finally, most of the interactions between the Master's and field dummies are statistically insignificant.

Credentials, Field of Study, and Management

Our data allows us a unique opportunity in terms of identifying significant determinants of promotion. Table 1.12 contains a summary of observations and individuals classified by

[30] Of course, the coefficient estimate of the engineering field dummy with respect to females is slightly larger than the corresponding coefficient estimate for female students in humanities.

promotion to management. Of the 3,055 individuals who start in non-management positions, 423, or roughly 14% are promoted to middle management while 150 (approximately 5%) obtain a senior management position. In contrast, 2,693 individuals begin at middle management, of which 501 (almost 19%) become senior managers over the sample period. Therefore, our estimates are not only identified by variation across individuals who remain in a certain category throughout the sample period, but through some time-series variation as well. Table 1.13 contains marginal effects estimates from an ordered probit model where the different categories are: non-management (1); middle management (2); and senior management (3).

Years of job experience is positively and significantly correlated (at the 1% level) with an increased chance of being in either middle or senior management. In terms of educational credentials, having a double undergraduate or undergraduate with diploma or a Master's degree are positively and significantly correlated (at the 1% level) with a middle and senior management position. However, the magnitude of marginal effects is larger for the Master's covariate. In contrast, possessing a Ph.D. is significantly associated (at the 1% level) with a lower chance of being either a middle or senior manager. Consistent with previous estimates, there are some specific fields, which are associated with career advancement. Controlling for other factors, graduates with either a humanities, business, or engineering degree are more likely than science graduates to hold a middle or senior management position (these marginal effects are all significant at the 1% level). On the other hand, education graduates are less likely than science degree holders to be in management.

The female dummy variable is statistically insignificant across all specifications, implying that being a woman does not impact the probability of holding a management position. The visible minority dummy is negative and statistically significant (at the 1% level) with respect to middle and senior management positions. However, we do not emphasize these results given the manner in which the variable was constructed. Having a degree from a Top 10 Ontario university is positively and significantly (1% level) associated with being in management.

We summarize our findings on the determinants of labor market accomplishment as follows. First, conforming to basic intuition, an increase in the number of years of work experience is significantly correlated with higher earnings, and we find no evidence of gender differences. Second, even after controlling for years of work experience, higher levels of educational attainment clearly impact earnings with Master's and Ph.D. degree holders experiencing up to roughly 3% and 10% premiums relative to undergraduate degree holders, and college graduates earning less than undergraduates. Third, we do not find strong evidence of sheepskin effects. Fourth, there are significant differences in returns by field and gender. On average, females earn less than men (roughly 3%), with a more pronounced spread at the doctoral level (roughly 6%). The penalty to females at the doctoral level seem to be driven by science and health degree holders. In contrast, for undergraduates, double undergraduate degree or undergraduate and diploma, and college graduates, female graduates in business, engineering, law, and education, earn more than men. There is an absence of statistically significant gender differences for Master's degree holders. Finally, while females earn less than males, they have experienced higher salary growth rates in comparison to males.

Academics and Earnings

The estimates in the previous table are based on government employees who are not affiliated with post-secondary institutions. However, there are a significant number of employees who also work in publicly funded colleges and universities, consisting of academics and administrators [31]. We were able to create a panel data set of 20,722 observations consisting of 5,427 observations related to administrative staff, 14,137 observations of professors without significant administrative responsibilities, and 1,158 observations of professors engaged in full-time administration. Table 1.14 consists of regression results from these data. While the empirical specifications are similar to previous tables, our focus is on years of experience and gender, and an evaluation of whether the coefficient estimates are similar to results obtained from data on mainstream government employees.

[31] Examples of administrative staff include the Vice President, Dean, Academic Director, Provost and Vice President Academic,Assistant Director, Assistant Dean, Secretary, Chief Librarian, Advisor, Chair, Manager, Chief Communications Officer, Chief Development Officer,Chief Financial Officer, Consultant, Course Director, Director, and so forth. Academics are full professors, Associate professors, Assistant professors, while professors in administration consist of Vice Dean/Professor, Professor/Director, Professor/Department Chair,Professor/Endowed Chair/Associate Dean, Professor/Program Coordinator/Associate Dean,Professor/Associate Dean, Professor/Coordinator, Professor/Dean, and so forth.

Consistent with previous findings, each year is significantly correlated (at the 1% level) with slightly more than a 1% increase in real earnings and results in columns (2) and (3) reveal no statistically significant differences between men and women. The gender dummy is negative and statistically significant (at the 1% level) in column (1) and suggests that controlling for all other factors, women earn roughly 3% less than men. This result is robust and comparable to our results with respect to mainstream government employees. However, the gender dummy becomes statistically insignificant in column (2) with the inclusion of the interaction dummies. The gender dummy and its interactions with respect to earnings growth regressions. The final result of interest are the negative and statistically significant (at the 1% level) coefficient estimates of the professor dummy variable across all columns, implying that academics, on average, earn less and experience lower earnings growth relative to non-academics who are on the sunshine list.

Time-series variation in earnings across cohorts

The above empirical estimates are premised on regression models whose coefficient estimates yield "on average" effects after controlling for time-specific unobservable shocks. However, the specifications do not offer an idea on the evolution of earnings and earnings growth rates over time. In order to better understand sources in earnings by gender and cohort over time, we calculate median earnings for each year of available data (2005-2013) for cohorts defined by the year of graduation of their final degree. Cohorts are defined by the following years of graduation: 1990-1994; 1995-1999; and 2000-2005. Further, we calculate median earnings separately for undergraduates and undergraduates with diplomas from individuals with Master's and Ph.D. degrees, given the significant differences in earnings that arise from our empirical specifications. Allocating individuals in this manner allows us to condition our estimates on differences in years of experience that are important to control for. Median earnings are relied upon in order to minimize skewness from a small number of outliers.

Table 1.15 contains calculations of median earnings by gender, cohort, for the beginning and end years of the sample, and the percentage change over time. The results demonstrate clear differences by cohort and educational attainment. First, while there are no statistically significant gender differences in 2005 for undergraduates in the 1990-1994 and

1995-1999 cohorts, there is a roughly 9% difference in favor of males for the 2000-2005 cohort. However, this difference becomes statistically insignificant by 2013. On the other hand, in 2013 there is a roughly 3% difference in median earnings between males and females with undergraduate degrees belonging to the 1990-94 cohort.

In contrast, there seems to be more pronounced gender differences for more educated individuals with graduate degrees. In 2005, with respect to the 1990-1994 cohort, male median earnings are 5.6% higher than corresponding earnings for females. On the other hand, for the 1995-1999 cohort, female median earnings are approximately 3% higher than those for males, while earning differences are not statistically significant for the 2000-2005 cohort. By 2013, differences in median earnings follow a different path. While there is no statistically significant gender difference for graduate degree holders in the 1990-1994 cohort, there are statistically significant gaps of 2.7% and 4.8% in median earnings in favor of males (graduate degree holders) for the 1995-1999 and 2000-2005 cohorts. In summary, these results are consistent with previous coefficient estimates that imply statistically significant differences in gender earnings are present among graduate degree holders and shed some light on which cohorts such trends are clearly visible.

Given these differences by cohort, Table 1.16 contains estimates based on separate samples constructed for all observations in 2005-2007 and for 2011-2013. Columns (1) and (2) consists of results for the 2005-2007 cohort with column (1) specifically containing estimates with respect to a base specification and column (2) adding gender dummy interactions with graduate degrees. Columns (3) and (4) are organized similarly as columns (1)-(2) but are based on the 2011-2013 cohort.

The first result is that an additional year of experience is significantly associated with higher earnings for men and women, ranging from 1%-1.4%. Second, the return to a Ph.D. relative to an undergraduate degree, increased over time, with an implied return of 5.7% in 2005-2007 (column (1)) and 9% in 2011-2013 (column (3)). In contrast, the returns to a Master's degree, on average, is comparable at roughly 2%, across samples. The negative return to a college degree increases across cohorts, from roughly 3% to 5.7%. Third, the incremental returns to Ph.D. and Master's graduates accrue to males, as the coefficient estimates of the interaction between the gender dummy and these degree dummies are almost as large as the corresponding estimates for the Ph.D. and Master's dummy variables. More specifically, as the coefficient estimates of graduate degrees in columns (2) and (4) reflect the return to males, the interaction with the gender dummy yields the marginal return to females. Adding the coefficient estimate of these interaction terms to coefficient estimates of either the Ph.D. or Master's dummy variables results in an estimate that is close to zero, implying that males with graduate degrees earn a significantly higher return than females. Therefore, the main finding from these estimates is that there exists a discrepancy in earnings by gender for graduate degree holders that persist over time.

Finally, Table 1.17 consists of estimates based on a restricted sample spanning 2011-2013. The first column contains estimates based on our standard specification. Column (2) includes the gender interactions with the Ph.D. and Master's dummy variables. Column (3) contains estimates based on a similar specification to the one employed in column (2) with additional interactions of the gender dummy with the senior and middle management indicators. Columns (4), (5), and (6) contain results based on the 1990-94, 1995-99, and 2000-2005 cohorts, respectively. Finally, column (7) reports estimates of a sub-sample consisting of only senior managers.

Estimates in the first two columns are comparable to previous findings, with the exception of the statistical insignificance of the gender interactions with the Ph.D. and Master's dummies. The gender dummy interactions with the Ph.D. and Master's dummies are statistically significant, but only at the 10% level. The key findings from columns (4)-(6) are the negative and statistically significant coefficient estimates (at the 1% or 5% levels) of the interactions between the gender indicator and the senior and middle manager dummies in column (4) implying that controlling for other factors, female managers in the 1990-94 cohort earn less than other counterparts. In contrast, the corresponding gender interactions with the manager dummies are statistically insignificant with respect to the 1995-99 and 2000-2005 cohorts. The results in column (7) support these findings. Specifically, while the coefficient estimate of the 1990-94 dummy is positive and statistically significant, the interaction between this dummy and the gender indicator is negative and statistically significant. In other words, controlling for other factors, the more senior females in our cohort who are probably in their mid and late forties and are in management, earn less than male counterparts.

1.6 Conclusion

We attempt to contribute to the literature by creating a unique panel data set of individual specific earnings by collating information available from the Ontario Salary Disclosure Act along with corresponding details from publicly available Linkedin profiles. There are very few studies that used individual specific data over a relatively long period of time to assess the comparative importance of educational attainment along with specific field of study, attendance at a major university, and conditioned on gender and years of experience. The use of panel data allows us to evaluate differences across cohorts and also control for the potentially confounding effects of unobserved year specific shocks. Most U.S. and Canadian studies rely on cross-sections of individuals or panel data with limited time-series.

Our data does have limitations. It is a truncated data set which only consists of high income earners in the provincial government. However, the provincial government is a significant employer and understanding differences in earnings among high income individuals as well as the underlying reasons for such discrepancies, is still of policy importance. In this respect, the data offers some novel insights. Graduate degrees and especially Ph.D's do result in higher earnings relative to undergraduates. There are distinctions across fields, with Ph.D's in science and health earning more than doctoral degree holders in other fields. However, excluding these fields, returns to a Ph.D. are generally comparable, including the much maligned field of humanities. Further, humanities undergraduates do not necessarily have significantly lower earnings than individuals in other fields. On average, they earn more than science and engineering undergraduates.

On average, females do earn less than male counterparts, which seems to be driven by graduate degree holders and to some extent, by gender differences in management. Specifically, as Figure 1.5 demonstrates that the proportion of men who are managers is far higher than the proportion of women in management positions. This result is in a sense, counter to the common belief that such discrepancies should not exist after higher levels of education [32]. However, the magnitude of these differences is smaller than conventional estimates of the average gender gap. And our data indicates that salary increases for women with undergraduate degrees have been higher than for men with similar educational attainment. Further, our results indicate that controlling for other factors, older women in our sample in senior management positions earn less than male counterparts. Future research will be focused on offering explanations for these observed trends. At the minimum our research offers some insight on research, which can be conducted through scraping data that are publicly available data on the web.

1.7 Tables

Table 1.1: Distribution of nominal salary, 2005 and 2013

	2005		2013	
Salary bin	Number of obs	%	Number of obs	%
100000-120000	604	55.06%	2,820	44.02%
120000-150000	292	26.62%	2,122	33.13%
150000-180000	122	11.12%	754	11.77%
180000-200000	31	2.83%	288	4.50%
> 200000	48	4.38%	422	6.59%
total	1,097	100.00%	6,406	100.00%

Source: authors' own calculation.

[32] It has been suggested that the gender wage gap gets smaller with higher level of education. Using the 1997 data from the Survey of Labor and Income Dynamics (SLID), Drolet (2001) find that the gender wage gap is lowest among university graduates. Based on 1986 and 1991 censuses evidence, Christie and Shannon (2001) show that the improving female's educational attainment compared to men has helped narrowing the gender wage gap. Frenette and Coulombe (2007) also finds that the educational trends have contributed towards a decrease in the gender wage gap in the 1990s.

Table 1.2: Distribution of individuals by the number of years they are observed

Number of years in our data	Average salary (nominal)	Number of individuals	Percentage of observations
2	112,961	519	8.10%
3	111,222	971	15.16%
4	112,761	978	15.26%
5	116,116	938	14.64%
6	111,797	868	13.55%
7	114,109	674	10.52%
8	115,931	537	8.38%
9	129,391	922	14.39%

Source: authors' own calculation.

Table 1.3: Distribution of observations, by credentials

Credentials	Number of obs	Percent
college	4,653	15.30%
BA	15,168	49.87%
BA + Diploma or Double BA	2,836	9.32%
Master	10,320	33.93%
PhD	2,093	6.88%
all	30,417	100.00%

Source: Author's own calculation.

Table 1.4: Descriptive statistics of fields of study, by credentials

	Fields of study						
Degree	Business	Education	Engineering	Health	Humanities	Law	Science
BA	2395	289	3015	632	6100	852	1885
	15.79%	1.91%	19.88%	4.17%	40.22%	5.62%	12.43%
BA + Diploma or Double BA	514	210	316	137	1153	136	370
	18.12%	7.40%	11.14%	4.83%	40.66%	4.80%	13.05%
college	537	67	1261	249	770	435	295
	11.54%	1.44%	27.10%	5.35%	16.55%	9.35%	6.34%
Master	1567	1380	1184	1567	2899	421	1302
	15.18%	13.37%	11.47%	15.18%	28.09%	4.08%	12.62%
PhD	45	64	233	433	626	231	461
	2.15%	3.06%	11.13%	20.69%	29.91%	11.04%	22.03%
all	5058	2010	6009	3018	11548	2075	4313
	14.42%	5.73%	17.13%	8.61%	32.93%	5.92%	12.30%

Source: Author's own calculation. Other: 1039

Table 1.5: Two sample t test: Census VS our sample

	All	College	BA	Graduate degree
Census	139,420.5	118,449.8	143,070.1	143,543.2
	(3,278.28)	(2,833.83)	(4,097.92)	(6,899.28)
LinkedIn	129,146.2	121,615.1	128,533.1	133,007
	(1,074.78)	(1,725.756)	(1,559.84)	(1,891.61)
diff	10, 274.28***	-3,165.32	14, 537.04***	10, 536.13*
t	3.79	-1.00	4.06	1.91
p-value	0.0002	0.3160	0.0001	0.0560

Source: authors' own calculation. Standard errors are in parenthesis. diff = mean(Census) - mean(LinkedIn).

Table 1.6: summary statistics, by year graduated

Graduated year	Gender	Measure	BA	Grad
before 1980	Female	Average real salary	94,696	104,308
		Number of observations	212	109
	Male	Average real salary	106,054	113,212
		Number of observations	541	252
		Difference	$11,357^{***}$	$8,903^{***}$
		Percentage difference	10.41%	7.86%
1980-1989	Female	Average real salary	94,413	99,190
		Number of observations	764	647
	Male	Average real salary	95,808	107,636
		Number of observations	1,879	1,081
		Difference	1,395	8,445
		Percentage difference	1.46%	7.85%
1990-1999	Female	Average real salary	91,882	96,120
		Number of observations	662	947
	Male	Average real salary	91,677	97,547
		Number of observations	1,265	1,099
		Difference	-205	$1,426^{*}$
		Percentage difference	-0.22%	1.46%
After 2000	Female	Average real salary	82,328	88,156
		Number of observations	231	493
	Male	Average real salary	65,155	89,865
		Number of observations	424	494
		Difference	$3,827^{***}$	$1,708^{**}$
		Percentage difference	4.44%	1.90%

Sample is limited to individuals with LinkedIn profile pictures; Total number of observations is 11,000. t value and p value is calculated using two sample t test. $H_0 : mean(Realsalary_{male}) - mean(Realsalary_{female}) = 0$. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. difference = $mean(Realsalary_{male}) - mean(Realsalary_{female})$. percentage difference = $\frac{mean(Realsalary_{male}) - mean(Realsalary_{female})}{mean(Realsalary_{male})} * 100$

Table 1.7: Average starting and ending nominal salary by number of years in our sample

	All			Female			Male		
# of years in data	Ave starting salary	Ave ending salary	△	Ave starting salary	Ave ending salary	△	Ave starting salary	Ave ending salary	△
2	113,174	118,657	4.85%	110,711	116,193	4.95%	114,746	120,230	4.78%
3	111,295	120,209	8.01%	110,808	118,456	6.90%	111,646	121,467	8.80%
4	112,882	125,149	10.87%	111,162	124,258	11.78%	113,878	125,665	10.35%
5	116,068	131,728	13.49%	114,829	130,035	13.24%	116,822	132,756	13.64%
6	111,671	133,324	19.39%	111,199	135,065	21.46%	111,931	132,366	18.26%
7	114,143	146,536	28.38%	113,037	145,677	28.88%	114,568	146,867	28.19%
8	115,928	152,001	31.12%	114,776	154,023	34.19%	116,473	151,043	29.68%
9	129,391	173,007	33.71%	126,998	172,119	35.53%	130,253	173,328	33.07%
Total	115,764	137,471	18.75%	113,840	134,361	18.03%	116,793	139,133	19.13%

Source: authors' own calculation. Sample is limited to people with five years of experience on-wards.

Table 1.8: OLS estimates of earnings, 2005 - 2013

	semi-log model						log-log model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of schooling (S_{is})					0.0098^{***}		
					(0.0019)		
Experience(EXP_{ist})		0.0087^{***}	0.0099^{***}	0.0121^{***}	0.0122^{***}	0.0119^{***}	0.0759^{***}
		(0.00140)	(0.00139)	(0.00128)	(0.00128)	(0.00128)	(0.0057)
Exp Squared (EXP^2_{ist})		-0.00009^{***}	-0.00011^{***}	-0.00019^{***}	-0.00018^{***}	-0.00018^{***}	—
		(0.000034)	(0.000033)	(0.000030)	(0.000030)	(0.000030)	—
PhD	0.0651^{***}	0.0927^{***}	0.0863^{***}	0.0826^{***}	0.0381^{**}	0.1051^{***}	0.1051^{***}
	(0.0182)	(0.0182)	(0.0171)	(0.0167)	(0.0194)	(0.0227)	(0.0227)
Master	0.00177	0.0159^{**}	0.0163^{**}	0.0181^{***}	0.00061	0.0288^{***}	0.0289^{***}
	(0.0077)	(0.0077)	(0.0081)	(0.0072)	(0.0077)	(0.0089)	(0.0089)
College	-0.0577^{***}	-0.0623^{***}	-0.0646^{***}	-0.0514^{***}	-0.0355^{***}	-0.0500^{***}	-0.0500^{***}
	(0.0074)	(0.0075)	(0.0087)	(0.0082)	(0.0088)	(0.0082)	(0.0082)
BA + Diploma/Double BA	-0.0637^{***}	-0.0501^{***}	-0.0460^{***}	-0.0093	-0.0132	-0.0104	-0.0104
	(0.0099)	(0.0100)	(0.0099)	(0.0085)	(0.0085)	(0.0085)	(0.0085)
Field of study							
Humanities			0.0163^{*}	0.0205^{***}	0.0208^{***}	0.0201^{***}	0.0200^{***}
			(0.0089)	(0.0082)	(0.0081)	(0.0082)	(0.0081)
Business			0.0718^{***}	0.0471^{***}	0.0468^{***}	0.0463^{***}	0.0464^{***}
			(0.0121)	(0.0104)	(0.0104)	(0.0104)	(0.0104)
Engineering			0.0292^{***}	-0.00413	-0.00286	-0.00387	-0.00369
			(0.0101)	(0.0089)	(0.0088)	(0.0089)	(0.0089)
Health			0.1045^{***}	0.0779^{***}	0.0771^{***}	0.0783^{***}	0.0778^{***}
			(0.0185)	(0.0171)	(0.0171)	(0.0169)	(0.0169)
Law			0.1050^{***}	0.0801^{***}	0.0819^{***}	0.0804^{***}	0.0800^{***}
			(0.0144)	(0.0137)	(0.0137)	(0.0137)	(0.0137)
Education			-0.0396^{***}	0.0353^{***}	0.0341^{***}	0.0364^{***}	0.0359^{***}
			(0.0119)	(0.0112)	(0.0113)	(0.0113)	(0.0113)
Others			0.00691	0.0202	0.0212	0.0205	0.0193
			(0.0144)	(0.0129)	(0.0127)	(0.0129)	(0.0129)
Job rank							
Senior manager				0.2647^{***}	0.2633^{***}	0.2638^{***}	0.2644^{***}
				(0.0115)	(0.0115)	(0.0115)	(0.0115)
Middle manager				0.00622	0.00684	0.0059	0.0063
				(0.0058)	(0.0058)	(0.0058)	(0.0059)
Female				-0.0319^{***}	-0.0309^{***}	-0.0175^{***}	-0.0171^{***}
				(0.0063)	(0.0063)	(0.0072)	(0.0072)
Visible Minority				-0.0217^{***}	-0.0211^{**}	-0.0224^{**}	-0.0230^{**}
				(0.0095)	(0.0095)	(0.0094)	(0.0094)
Top 10 ON university				0.00130	0.00161	0.0012	0.0012
				(0.0066)	(0.0066)	(0.0066)	(0.0066)
US university				-0.0312^{***}	-0.0349^{***}	-0.0315^{***}	-0.0314^{***}
				(0.009)	(0.0091)	(0.0089)	(0.0090)
Master * Female						-0.0301^{***}	-0.0305^{***}
						(0.0124)	(0.0125)
PhD * Female						-0.0639^{**}	-0.0655^{**}
						(0.0307)	(0.0308)
Year fixed effect	No	No	No	Yes	Yes	Yes	Yes
Sector fixed effect	No	No	No	Yes	Yes	Yes	Yes
Observations	35,070	35,070	35,070	35,070	35,070	35,070	35,070
R^2	0.0160	0.0399	0.0649	0.2603	0.2640	0.2616	0.2607

Note: The omitted category for the highest degree rewarded is "BA" and the omitted category for field of study is "Science". Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 1.9: OLS estimates of salary growth, 2005 - 2013

	(1)	(2)	(3)	(4)	(5)	(6)
Years of schooling (S_{is})					0.00035	
					(0.00057)	
Experience(EXP_{ist})		-0.0014^{**}	-0.0013^{**}	-0.0024^{***}	-0.0023^{***}	-0.0024^{***}
		(0.0006)	(0.0007)	(0.0006)	(0.0006)	(0.0006)
Exp Squared (EXP_{ist}^2)		-0.000000	-0.000000	0.000015	0.000015	0.000016
		(0.000013)	(0.000013)	(0.000013)	(0.000013)	(0.000013)
PhD	-0.0020	-0.0107^{**}	-0.0103^{***}	-0.0080^{**}	-0.0096^{***}	-0.0032
	(0.0045)	(0.0045)	(0.0044)	(0.0043)	(0.0049)	(0.0055)
Master	0.0027	-0.0027	-0.0016	-0.0015	-0.0022	0.0011
	(0.0021)	(0.0022)	(0.0023)	(0.0022)	(0.0024)	(0.0028)
College	-0.0092^{***}	-0.0075^{***}	-0.0077^{***}	-0.0052^{*}	-0.0047	-0.0049
	(0.0026)	(0.0026)	(0.0030)	(0.0032)	(0.0033)	(0.0032)
Diploma/Double BA	-0.0024	-0.0067^{**}	-0.0064^{**}	-0.0029	-0.0031	-0.0032
	(0.0032)	(0.0032)	(0.0033)	(0.0034)	(0.0034)	(0.0034)
Humanities			0.00036	-0.0017	-0.0017	-0.0018
			(0.0028)	(0.0028)	(0.0028)	(0.0028)
Business			0.0036	-0.0041	-0.0041	-0.0043
			(0.0035)	(0.0034)	(0.0034)	(0.0034)
Engineering			-0.00046	-0.0055^{*}	-0.0055^{*}	-0.0055^{*}
			(0.0033)	(0.0033)	(0.0033)	(0.0033)
Health			-0.0010	-0.0047	-0.0045	-0.0044
			(0.0047)	(0.0047)	(0.0047)	(0.0047)
Law			0.0095^{**}	0.0045	0.0046	0.0046
			(0.0048)	(0.0049)	(0.0049)	(0.0049)
Education			-0.0036	0.0027	0.0027	0.0030
			(0.0037)	(0.0040)	(0.0040)	(0.0040)
Others			0.00039	-0.0033	-0.0032	-0.0032
			(0.0055)	(0.0055)	(0.0056)	(0.0056)
Senior manager				0.0500^{***}	0.0500^{***}	0.0498^{***}
				(0.0037)	(0.0037)	(0.0037)
Middle manager				0.0054^{***}	0.0054^{***}	0.0053^{***}
				(0.0019)	(0.0019)	(0.0019)
Female				0.0043^{**}	0.0044^{**}	0.0080^{***}
				(0.0020)	(0.0020)	(0.0024)
Visible Minority				-0.00054	-0.00051	-0.00073
				(0.0028)	(0.0028)	(0.0028)
Top 10 ON university				0.000000	0.00003	-0.00004
				(0.0021)	(0.0021)	(0.0021)
US university				-0.0006	-0.00018	-0.00006
				(0.0032)	(0.0031)	(0.0031)
Master * Female						-0.0078^{**}
						(0.0040)
PhD * Female						-0.0144^{*}
						(0.0085)
Year fixed effect	No	No	No	Yes	Yes	Yes
Sector fixed effect	No	No	No	Yes	Yes	Yes
Observations	22,258	22,258	22,258	22,258	22,258	22,258
R^2	0.0009	0.0122	0.0127	0.0821	0.0821	0.0824

Note: The omitted category for the highest degree rewarded is "BA" and the omitted category for field of study is "Science". Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 1.10: Marginal effects from ordered probit estimates of different salary categories

	$W_{ist}=1$	$W_{ist}=2$	$W_{ist}=3$	$W_{ist}=4$
Experience(EXP_{ist})	-0.0191^{***}	0.0136^{***}	0.0030^{***}	0.0019^{***}
	(0.0024)	(0.0017)	(0.00041)	(0.00024)
Exp Squared (EXP^2_{ist})	0.00032^{***}	-0.00022^{***}	-0.000050^{***}	-0.000025^{***}
	(0.00005)	(0.00004)	(0.0000)	(0.0000)
College	0.0722^{***}	-0.0516^{***}	-0.0114^{***}	-0.0092^{***}
	(0.0165)	(0.0118)	(0.0030)	(0.0022)
Master	-0.0307^{**}	0.0219^{***}	0.0048^{***}	0.0039^{***}
	(0.0114)	(0.0082)	(0.0018)	(0.0014)
PhD	-0.0921^{***}	0.0658^{***}	0.0146^{***}	0.0117^{***}
	(0.0213)	(0.0153)	(0.0034)	(0.0028)
Diploma/Double BA	0.0279	-0.0199	-0.0044	-0.0035
	(0.0180)	(0.0128)	(0.0028)	(0.0023)
Humanities	-0.0246^{*}	0.0175^{*}	0.0038^{***}	0.0031^{***}
	(0.0148)	(0.0105)	(0.0023)	(0.0018)
Business	-0.0657^{***}	0.0469^{***}	0.0104^{***}	0.0083^{***}
	(0.0164)	(0.0126)	(0.0028)	(0.0023)
Engineering	0.0227^{***}	-0.0162	-0.0036	-0.0028
	(0.0164)	(0.0117)	(0.0026)	(0.0021)
Health	-0.1004^{***}	0.0717^{***}	0.0159^{***}	0.0128^{***}
	(0.0231)	(0.0164)	(0.0037)	(0.0032)
Law	-0.0888^{***}	0.0634^{***}	0.0140^{***}	0.0113^{***}
	(0.0206)	(0.0148)	(0.0035)	(0.0027)
Education	-0.0399	0.0285	0.0063	0.0050
	(0.0260)	(0.0185)	(0.0041)	(0.0033)
Other	0.0088	-0.0062	-0.0014	-0.0011
	(0.0293)	(0.0209)	(0.0046)	(0.0037)
Senior manager	-0.2879^{***}	0.2057^{***}	0.0456^{***}	0.0367^{***}
	(0.0124)	(0.0098)	(0.0031)	(0.0028)
Middle manager	0.0196^{*}	-0.0140^{*}	-0.0025^{*}	-0.0025^{*}
	(0.0106)	(0.0075)	(0.0017)	(0.0014)
Female	0.0349^{***}	-0.0249^{***}	-0.0045^{***}	-0.0044^{***}
	(0.0105)	(0.0075)	(0.0017)	(0.0014)
Visible Minority	0.0428^{***}	-0.0306^{***}	-0.0055^{***}	-0.0054^{***}
	(0.0165)	(0.0119)	(0.0026)	(0.0021)
Top 10 ON university	-0.0071	0.0050	0.0011	0.00091
	(0.0107)	(0.0076)	(0.0017)	(0.0014)
US university	0.0273	-0.0196	-0.0043	-0.0034
	(0.015)	(0.017)	(0.0024)	(0.0019)
Year fixed effect	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes
obs	26,250	7,239	964	617

Note: Regression include experience, experience squared, gender dummy, rank dummy, year dummy,sector dummy, fields of study, credentials, ethic,etc. The omitted category for the highest degree rewarded is "BA" and the omitted category for field of study is "Science". Standard errors are in parenthesis and clustered at the individual level.

Table 1.11: OLS estimates of the returns to specific fields by education level & gender

	All	Male	Female
Experience(EXP_{ist})	0.0120***	0.0119***	0.0134***
	(0.00128)	(0.00154)	(0.0023)
Exp Squared (EXP^2_{ist})	−0.00018***	−0.00018***	−0.00023***
	(0.000030)	(0.000036)	(0.000056)
Credentials			
PhD	0.1275***	0.1391***	0.0864***
	(0.0328)	(0.0434)	(0.0468)
Master	0.0353**	0.0370*	0.0322
	(0.0166)	(0.0220)	(0.0253)
College	−0.0515***	−0.0544***	−0.0433***
	(0.0082)	(0.0096)	(0.0171)
BA + Diploma/Double BA	-0.0064	0.0022	-0.0140
	(0.0085)	(0.0110)	(0.0133)
Field of study			
Humanities	0.0378***	0.0338***	0.0441***
	(0.0092)	(0.0112)	(0.0156)
Business	0.0529***	0.0350***	0.0805***
	(0.0115)	(0.0143)	(0.0195)
Engineering	0.0156*	0.0078	0.0558**
	(0.0252)	(0.0115)	(0.0238)
Health	0.0389*	0.0689*	0.0233
	(0.0252)	(0.0448)	(0.0283)
Law	0.0849***	0.0653***	0.1493***
	(0.0145)	(0.0158)	(0.0339)
Education	0.0358***	0.0211	0.0578***
	(0.0145)	(0.0148)	(0.0322)
Interaction of Fields with PhD/Master			
PhD * Humanities	−0.1495***	−0.1574***	−0.1188***
	(0.0371)	(0.0514)	(0.0490)
PhD * Business	−0.1430***	−0.1998***	-0.0383
	(0.0479)	(0.0505)	(0.0698)
PhD * Engineering	−0.1212***	−0.1361***	-0.0852
	(0.0381)	(0.0477)	(0.0622)
PhD * Health	0.1621***	0.2105**	0.0748
	(0.0656)	(0.0946)	(0.0798)
PhD * Law	-0.0348	-0.0402	-0.0393
	(0.0544)	(0.0678)	(0.0921)
PhD * Education	-0.0946	−0.1572***	-0.0433
	(0.0680)	(0.0540)	(0.0957)
MA * Humanities	-0.0234	-0.0210	-0.0278
	(0.0188)	(0.0252)	(0.0279)
MA * Business	0.00047	0.0221	-0.0233
	(0.0235)	(0.0309)	(0.0359)
MA * Engineering	−0.0504**	−0.0437*	−0.0969***
	(0.0216)	(0.0268)	(0.0368)
MA * Health	0.0197	0.0452	0.0251
	(0.0341)	(0.0640)	(0.0392)
MA * Law	0.0035	-0.0221	-0.0186
	(0.0373)	(0.0467)	(0.0611)
MA * Education	0.0019	0.0474*	-0.0442
	(0.0224)	(0.0281)	(0.0360)
Year fixed effect	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes
Observations	35,070	23,525	11,545
R^2	0.2717	0.2964	0.2718

Table 1.12: summary of observations and individuals classified by promotion to management

	original rank		# of individuals promoted to manger	
	obs	individual	middle manager	senior manager
non-management	15,407	3,055	423	150
middle manager	14,761	2,693	—	501
senior manager	4,902	658	—	—

Table 1.13: Marginal effects from Ordered Probit estimates of rank promotion

	non-management	middle manager	senior manager
Experience(EXP_{ist})	-0.0172^{***}	0.0073^{***}	0.0098^{***}
	(0.0112)	(0.0012)	(0.0015)
Exp Squared (EXP_{ist}^2)	0.00002^{***}	-0.000095^{***}	-0.00013^{***}
	(0.00006)	(0.00002)	(0.000035)
College	0.0663^{***}	-0.0283^{***}	0.0379^{***}
	(0.0193)	(0.0093)	(0.0111)
Master	-0.0757^{***}	0.0323^{***}	0.0435^{***}
	(0.0123)	(0.0053)	(0.0071)
PhD	0.0998^{***}	-0.0427^{***}	-0.0572^{***}
	(0.0282)	(0.0122)	(0.0160)
Diploma/Double BA	-0.0582^{***}	0.0249^{***}	0.0333^{***}
	(0.0200)	(0.0086)	(0.0115)
Field of study			
Humanities	-0.0452^{***}	0.0193^{***}	0.0258^{***}
	(0.0163)	(0.0069)	(0.0093)
Business	-0.1624^{***}	0.0694^{***}	0.0929^{***}
	(0.0185)	(0.0093)	(0.0107)
Engineering	-0.0659^{***}	0.0282^{***}	0.0377^{***}
	(0.0186)	(0.0080)	(0.0106)
Health	-0.0257	0.0110	0.0148
	(0.0240)	(0.0102)	(0.0137)
Law	0.0333	-0.0142	-0.0191
	(0.0274)	(0.0117)	(0.0157)
Education	0.0916^{***}	-0.0392^{***}	-0.0524^{***}
	(0.0288)	(0.0123)	(0.0166)
Other	-0.0255	0.0109	0.0145
	(0.0358)	(0.0153)	(0.0205)
Female	-0.0145	0.0062	0.0083
	(0.0112)	(0.0048)	(0.0063)
Visible Minority	0.0613^{***}	-0.0262^{***}	-0.0351^{***}
	(0.0168)	(0.0072)	(0.0097)
Top 10 ON university	-0.0319^{***}	0.0137^{***}	0.0182^{***}
	(0.0116)	(0.0050)	(0.0067)
US university	0.0353^{***}	-0.0151^{**}	-0.0203^{**}
	(0.0178)	(0.0076)	(0.0102)
Year fixed effect	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes

Note: The omitted category for the highest degree rewarded is "BA" and the omitted category for field of study is "Science". Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 1.14: OLS estimates of the returns to education in Ontario universities and colleges

	salary				salary growth rate			
	All	All	Male	Female	All	All	Male	Female
Experience(EXP_{ist})	0.0108***	0.0108***	0.0112***	0.0104***	-0.0051	-0.0051	-0.0390	0.0683
	(0.0014)	(0.0015)	(0.0018)	(0.0024)	(0.0575)	(0.0574)	(0.0651)	(0.1259)
Exp Squared (EXP^2_{ist})	−0.000082***	−0.00008***	−0.000086***	−0.000099*	−0.00303***	−0.0031*	−0.0026*	-0.0037
	(0.000034)	(0.00004)	(0.00004)	(0.00006)	(0.00128)	(0.0013)	(0.0014)	(0.0031)
Credentials								
PhD	0.1026***	0.1028***	0.0979***	0.1049***	1.5070***	1.5114***	1.5071***	1.816***
	(0.01236)	(0.0123)	(0.0156)	(0.0202)	(0.3859)	(0.3848)	(0.4472)	(0.8055)
Master	0.00402	0.0038	-0.0050	0.0105	0.34040	0.3279	1.0012*	-0.9276
	(0.0123)	(0.0123)	(0.0158)	(0.4175)	(0.0190)	(0.4203)	(0.5430)	(0.6779)
College	−0.0783***	−0.0782***	−0.1028***	-0.0394	-0.3311	-0.3073	-0.3740	-0.4412
	(0.0215)	(0.0215)	(0.0260)	(0.0352)	(0.9125)	(0.9142)	((1.2680)	(1.1835)
BA + Diploma	0.07451***	0.0743***	0.0878**	0.0463	1.17223	1.1658	0.8817	1.6168
(Double BA)	(0.0314)	(0.0316)	(0.0446)	(0.0386)	(1.2120)	(1.2156)	(1.6974)	(1.5789)
Field of study								
Humanities	−0.0431***	−0.0433***	−0.0498***	0.0362*	-0.6430	0.6496	-0.6301	-0.6464
	(0.01254)	(0.0126)	(0..0159)	(0.0203)	(0.4415)	(0.4420)	(0.5371)	(0.7837)
Business	0.00239	0.0024	0.0170	-0.0230	-0.9500)	-0.9467	-0.6884	-1.4677
	(0.0233)	(0.0232)	(0.0329)	(0.0281)	(0.6861	(0.6858)	(0.8344)	(1.1844)
Engineering	−0.04207***	−0.0422***	−0.0450***	-0.0416**	-0.6489	-0.6543	-0.6323	-0.8226
	(0.0120)	(0.0121)	(0.0148)	(0.0203)	(0.4291)	(0.4293)	(0.5081)	(0.8208)
Health	−0.0833***	−0.0835***	−0.0934***	−0.0597**	−1.3314***	−1.3352***	−1.5918***	-0.6935
	(0.01690)	(0.0168)	(0.0199)	(0.0314)	(0.5323)	(0.5307)	(0.6213)	(1.0823)
Law	-0.01089	-0.0107	-0.0154	-0.0042	-0.8050	-0.7974	-0.9947	-0.1372
	(0.01733)	(0.0173)	(0.0215)	(0.0281)	(0.5897)	(0.5911)	(0.6790)	(1.2022)
Education	−0.0472*	−0.0474*	−0.1004*	-0.0085	-0.7214	-0.7315	-2.5581	1.3212
	(0.02924)	(0.0291)	(0.0217)	(0.0453)	(1.4935)	(1.4970)	(1.8952)	(2.1312)
Other	−0.0254*	−0.0255*	−0.0339*	-0.0084	0.2149	0.2196	0.1246	0.5330
	(0.01585)	(0.0158)	(0.0186)	(0.0303)	(0.5159)	(0.5164)	(0.5958)	(1.0519)
Job rank								
Prof	−0.0691***	−0.0636***	−0.0635***	−0.0793***	−2.0877***	−1.8594***	−1.7974***	−2.8008***
	(0.0095)	(0.0115)	(0.0119)	(0.0160)	(0.3387)	(0.3827)	(0.3895)	(0.6863)
Prof in admin	0.03563**	0.0344*	0.0337*	0.09012	0.0467	0.0352	0.1039	0.0181
	(0.0180)	(0.0209)	(0.0213)	(0.0334)	(0.5407)	(0.6160)	(0.6171)	(1.1104)
Female	−0.0308***	-0.0208	—	—	0.3540	0.7997	—	—
	(0.0072)	(0.0150)	—	—	(0.2595)	(0.5683)	—	—
Visible Minority	0.00592	0.0058	0.0023	0.0171	0.0801	0.07614	0.0294	0.2950
	(0.0103)	(0.0102)	(0.0121)	(0.0189)	(0.3030)	(0.3022)	(0.3462)	(0.5980)
Top 10 ON university	0.00966	0.0099	0.0085	0.0176	−0.5872***	−0.5732**	-0.5508	-0.4993
	(0.01085)	(0.0108)	(0.0129)	(0.0194)	(0.2841)	(0.2839)	(0.3326)	(0.5527)
US university	-0.01073	-0.0102	0.0089	-0.0302	-0.4817	-0.4553	0.0103	−1.1436*
	(0.01842)	(0.0185)	(0.0232)	(0.0295)	(0.5911)	(0.5915)	(0.8775)	(0.6957)
Prof * Female	—	-0.0162	—	—	—	-0.7289	—	—
	—	(0.0172)	—	—	—	(0.6282)	—	—
Prof in admin * Female	—	0.0099	—	—	—	0.4356	—	—
	—	(0.0209)	—	—	—	(1.1564)	—	—
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,722	20,722	14,615	6,107	13,166	13,166	9,515	3,651
R^2	0.2092	0.2095	0.2180	0.1703	0.0560	0.0562	0.0579	0.0575

Table 1.15: median real salary, 2005 and 2013

Year	measure	1990-1994	1995-1999	2000-2005
undergrads				
2005	female	92,684	87,774	81,844
	male	91,589	87,513	90,268
	difference	-1,095	-261	8,424
	percentage difference	-1.20%	-0.30%	9.33%
2013	female	80,950	81,033	79,840
	male	83,402	81,426	80,307
	difference	2,452	393	466
	percentage difference	2.94%	0.48%	0.58%
graduates				
2005	female	91,968	93,470	84,957
	male	97,428	90,590	86,269
	difference	5,460	-2,880	1,312
	percentage difference	5.60%	-3.18%	1.52%
2013	female	83,720	80,278	79,367
	male	83,987	82,491	83,339
	difference	267	2,213	3,972
	percentage difference	0.32%	2.68%	4.77%

difference = $median(salary_{male}) - median(salary_{female})$.

percentage difference = $\frac{median(salary_{male}) - median(salary_{female})}{median(salary_{male})} * 100$.

Table 1.16: OLS estimates based on separate samples constructed for observations in 2005-2007 and 2011-2013

	2005-2007		2011-2013	
	(1)	(2)	(3)	(4)
Experience	0.0112***	0.0108***	0.0139***	0.0138***
	(0.00247)	(0.0024)	(0.0013)	(0.0013)
Exp Squared	−0.00020***	−0.00019***	−0.00022***	−0.00021***
	(0.00006)	(0.00006)	(0.000029)	(0.000030)
PhD	0.0571**	0.0752***	0.0899***	0.1101***
	(0.0260)	(0.0328)	(0.0155)	(0.0210)
Master	0.0231*	0.0384***	0.0181***	0.0270***
	(0.0125)	(0.0145)	(0.0067)	(0.0084)
College	−0.0297**	−0.0277**	−0.0573***	−0.0562***
	(0.0137)	(0.0137)	(0.0084)	(0.0084)
Diploma/Double BA	-0.00636	-0.0083	-0.0111	-0.0118
	(0.0167)	(0.0166)	(0.0085)	(0.0085)
Field of study				
Humanities	0.00146	0.00062	0.0253***	0.0250***
	(0.0142)	(0.0141)	(0.0079)	(0.0079)
Business	0.0523***	0.0501***	0.0443***	0.0438***
	(0.0189)	(0.0189)	(0.0099)	(0.0099)
Engineering	-0.0199	-0.0206	0.0030	0.0034
	(0.0155)	(0.0155)	(0.0088)	(0.0088)
Health	0.0781***	0.0785***	0.0715***	0.0721***
	(0.0324)	(0.0323)	(0.0153)	(0.0152)
Law	0.0408**	0.0410***	0.0932***	0.0937***
	(0.0202)	(0.0203)	(0.0142)	(0.0142)
Education	0.0267	0.0280	0.0430***	0.0440***
	(0.0236)	(0.0238)	(0.0110)	(0.0110)
Visible Minority	-0.0067	-0.0081	−0.0232**	−0.0238***
	(0.0181)	(0.0180)	(0.0084)	(0.0084)
Female	−0.0334***	-0.0123	−0.0319***	−0.0204***
	(0.0112)	(0.0128)	(0.0059)	(0.0068)
Master * Female	—	−0.0497**	—	−0.0240**
	—	(0.0226)	—	(0.0119)
PhD * Female	—	-0.0674	—	−0.0534*
	—	(0.0473)	—	(0.0298)
Year fixed effect	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes
Observations	4,837	4,837	18,658	18,658
R^2	0.2066	0.2091	0.2776	0.2784

Note: The omitted category for the highest degree rewarded is "BA" and the omitted category for field of study is "Science". Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 1.17: OLS estimates of the effects of education on earnings, 2011 - 2013

	All			1990-1994	1995-1999	2000-2005	senior managers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experience(EXP_{ist})	0.0138***	0.0136***	0.0135***	0.0103	0.0191**	0.0172***	—
	(0.0022)	(0.0021)	(0.0021)	(0.0115)	(0.0092)	(0.0052)	—
Exp Squared (EXP_{ist}^2)	−0.00027***	−0.00027***	−0.00027***	-0.00018	-0.00027	−0.00038***	—
	(0.000067)	(0.000067)	(0.000067)	(0.00024)	(0.00023)	(0.00013)	—
PhD	0.0433***	0.0638***	0.0653***	0.0539	0.0322	0.1113***	0.0737
	(0.0171)	(0.0248)	(0.0249)	(0.0530)	(0.0385)	(0.0441)	(0.0556)
Master	0.0191**	0.02757∗∗∗	0.0254***	0.0268	0.0382*	0.0116	0.0855***
	(0.0095)	(0.0115)	(0.0115)	(0.0194)	(0.0184)	(0.0200)	(0.0339)
College	−0.0600***	−0.0587***	−0.0582***	−0.0725***	−0.0555***	-0.0349	-0.0790
	(0.0122)	(0.0122)	(0.0123)	(0.0177)	(0.0246)	(0.0289)	(0.0731)
BA + Diploma/Double BA	-0.0057	-0.0063	-0.0062	-0.0170	0.0081	-0.0146	0.0434
	(0.0115)	(0.0113)	(0.0114)	(0.0187)	(0.0211)	(0.0211)	(0.0675)
Humanities	0.0030	0.0025	0.0030	0.0097	0.0015	-0.0232	0.0227
	(0.0115)	(0.0115)	(0.0115)	(0.0195)	(0.0205)	(0.0243)	(0.0651)
Business	0.0187	0.0182	0.0187	0.0248	0.0237	0.0011	0.0901
	(0.0145)	(0.0145)	(0.0146)	(0.0245)	(0.0267)	(0.0265)	(0.0680)
Engineering	−0.0345**	−0.0345**	−0.0350***	−0.0409**	-0.0355	−0.0543**	-0.0503
	(0.0126)	(0.0126)	(0.0125)	(0.0203)	(0.0238)	(0.0244)	(0.0691)
Health	0.0354*	0.0351*	0.0352*	0.0337	0.0511	0.0591*	0.1686***
	(0.0213)	(0.0210)	(0.0211)	(0.0392)	(0.0423)	(0.0382)	(0.0776)
Law	0.0698***	0.0690***	0.0681***	0.0816**	0.0948***	0.0121	-0.0244
	(0.0183)	(0.0182)	(0.0183)	(0.0401)	(0.0300)	(0.0274)	(0.0762)
Education	0.0103	0.0103	0.0103	0.0436*	0.0181	−0.0506*	-0.0348
	(0.0139)	(0.0139)	(0.0139)	(0.0243)	(0.0251)	(0.0266)	(0.0902)
Others	0.0324	0.0320	0.0320	0.0254	0.0324	-0.0314	0.0887
	(0.0198)	(0.0197)	(0.0197)	(0.0312)	(0.0325)	(0.0431)	(0.1078)
Senior manager	0.2307***	0.2294***	0.2348***	0.2994***	0.1972***	0.1825***	—
	(0.0169)	(0.0168)	(0.0212)	(0.0243)	(0.0334)	(0.0418)	—
Middle manager	-0.0073	-0.0075	0.0015	0.0228	-0.0140	-0.0139	—
	(0.0077)	(0.0076)	(0.0090)	(0.0163)	(0.0158)	(0.0170)	—
Female	−0.0301***	−0.0183*	-0.0060	0.0169	0.00009	−0.0587***	-0.0517
	(0.0083)	(0.0099)	(0.0121)	(0.0220)	(0.0211)	(0.0207)	(0.0607)
Top 10 ON university	-0.0067	-0.0067	-0.0069	0.0092	-0.0112	-0.0234	-0.0377
	(0.0089)	(0.0088)	(0.0089)	(0.0162)	(0.0154)	(0.0148)	(0.0319)
US university	-0.0150	-0.0152	-0.0145	-0.0075	-0.0143	-0.0186	-0.0350
	(0.0122)	(0.0122)	(0.0122)	(0.0193)	(0.0233)	(0.0243)	(0.0564)
Master * Female		-0.0180	-0.0167	-0.0191	-0.0383	0.0228	—
		(0.0155)	(0.0155)	(0.0265)	(0.0265)	(0.0299)	—
PhD * Female		-0.0487	−0.0529*	-0.0701	-0.0627	-0.0240	—
		(0.0314)	(0.0314)	(0.0819)	(0.0475)	(0.0598)	—
Senior Manager * Female			-0.0165	−0.1394***	0.0661	0.0581	—
			(0.0327)	(0.0494)	(0.0555)	(0.0674)	—
Middle Manager * Female			−0.0249*	−0.0506**	-0.0163	0.0106	—
			(0.0134)	(0.0245)	(0.0233)	(0.0263)	—
cohort 1990-1994							0.2198***
							(0.0459)
cohort 1995-1999							0.1141***
							(0.0489)
female * cohort 1990-1994							−0.1303*
							(0.0746)
female * cohort 1995-1999							0.0322
							(0.0820)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,472	16,472	16,472	6462	5,177	3,403	1,873
R^2	0.2162	0.2171	0.2177	0.2522	0.2448	0.2027	0.2483

1.8 Figures

Figure 1.1: Distribution of individuals by years of experience averaged across the sample

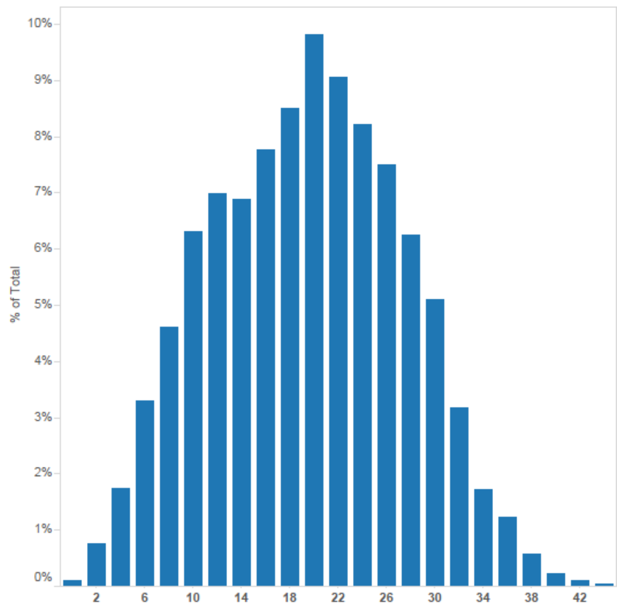
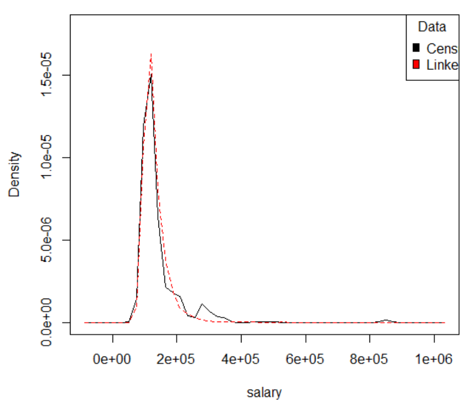
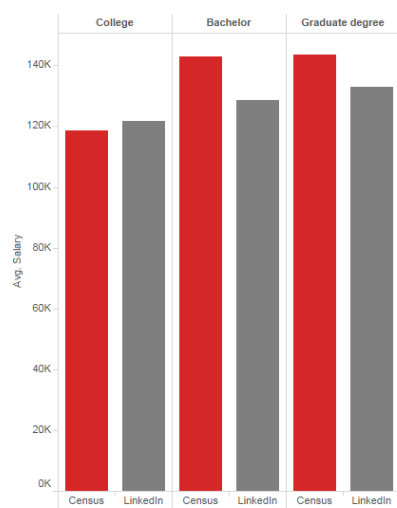


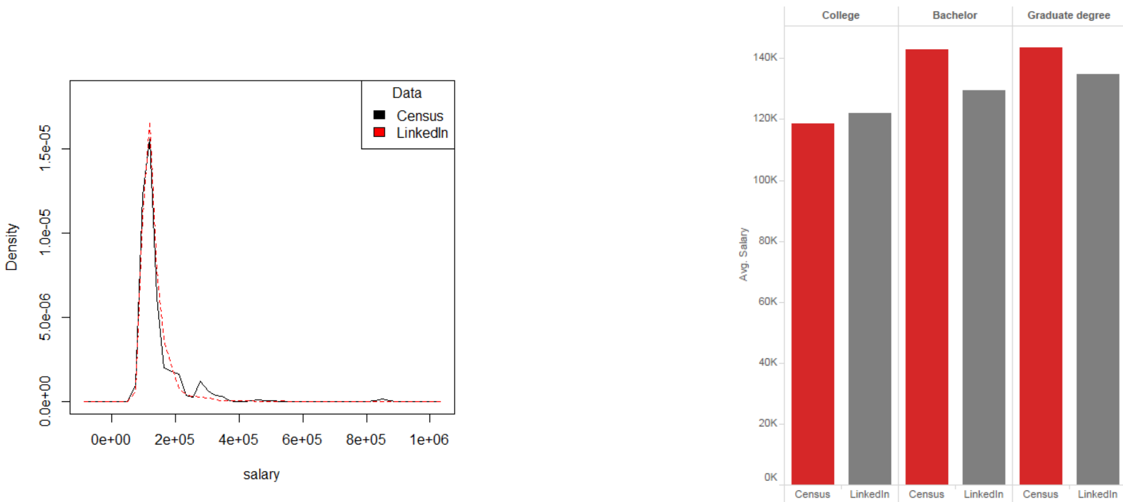
Figure 1.2: Nominal salary comparison, Census 2006 and our sample





Note: LinkedIn salary is limited to year 2005 and number of observations is 1097.

Figure 1.3: Nominal salary comparison, Census 2006 and our sample



Note: LinkedIn salary used here is the average salary of year 2005, 2006 and 2007.

Figure 1.4: Density of average salary of individuals on Sunshine List with and without LinkedIn profile, by different sectors

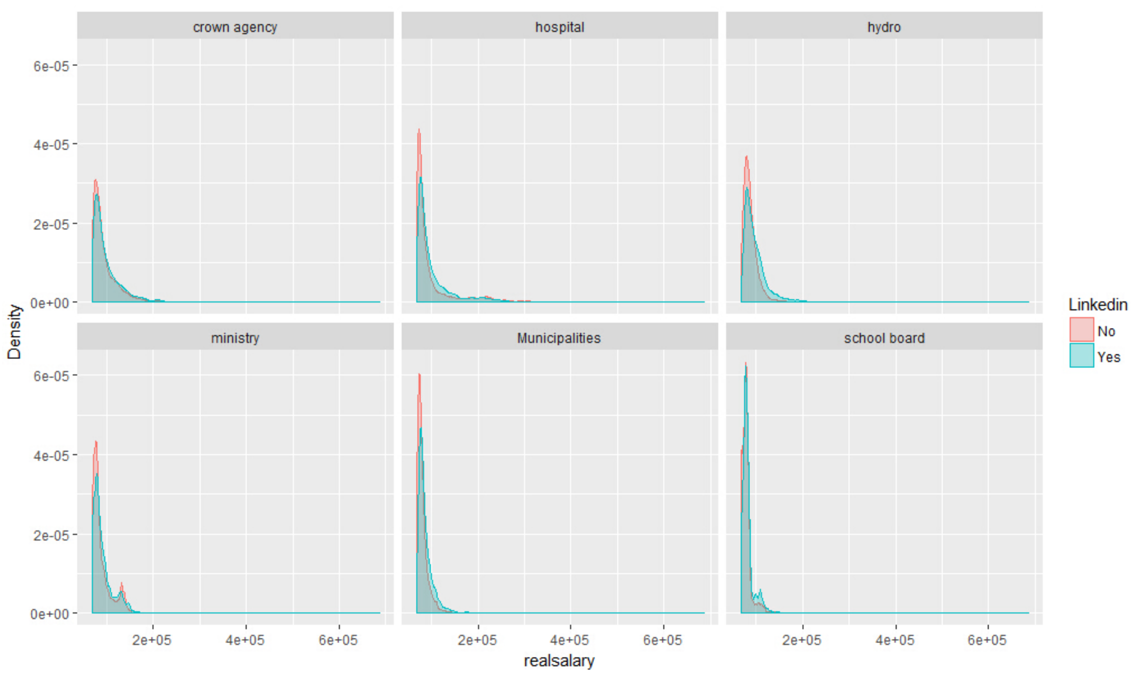
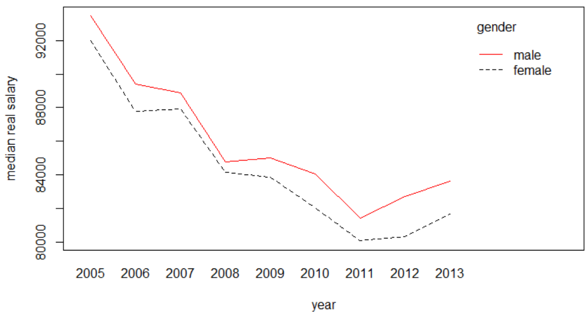
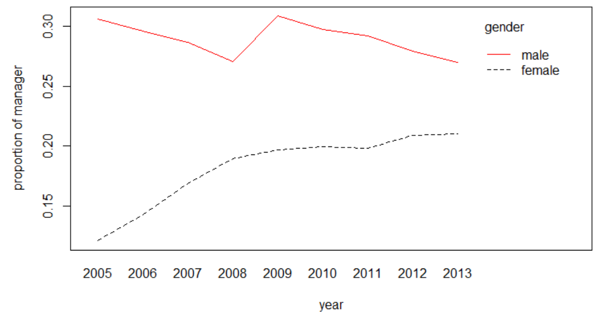


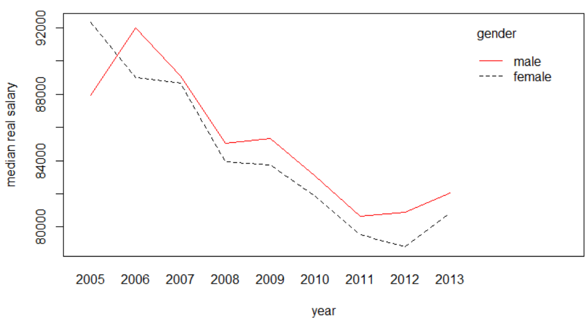
Figure 1.5: Median real salary and proportion of managers by gender



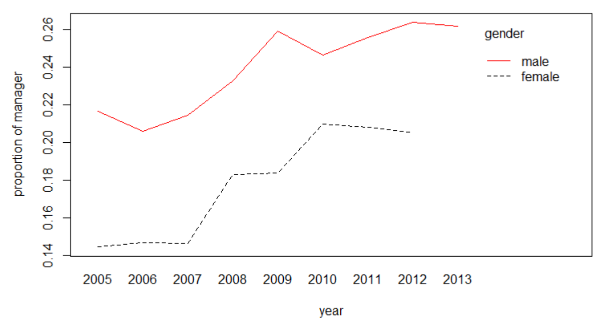
(a) Median real salary,1990-1994



(b) proportion of managers,1990-1994



(c) Median real salary,1995-1999



(d) proportion of managers,1995-1999

(e) Median real salary,2000-2005

(f) proportion of managers,2000-2005

Notes: Sample is restricted to those who graduated between 1990 and 2005.
Manages include both middle managers and senior managers.

Chapter 2

How the Market Rewards Academic Economists: Panel data evidence from Ontario

2.1 Introduction

Using unique panel data on all tenured and tenure-track professors from 16 Ontario economics departments from 1996 to 2012, this paper analyzes the pay and position of professors to understand how co-authorship affects research productivity, pay, and promotion among economists in academia. We also investigate the effect of productivity on pay and the effect of research and researcher's characteristics on the likelihood of collaboration. The ability to construct longitudinal data that tracks a sample of individuals over a reasonable period allows us to control for the potential confounding effects of some unobserved time-invariant factors.

Given the significant private and social value of academic research and the increasing amount of collaborative work [1] among researchers, understanding the influence of co-authorship on research and researchers is of importance and, in this respect, insufficient attention has been paid to this subject in Canada. Is co-authorship associated with higher productivity? Do different types of collaboration have different effects on productivity? These questions ought to be answered because, if co-authored publications are less productive than single-authored publications, any reward scheme that promotes co-authorship

[1] According to Katz and Martin (1997), co-authorship of a publication is the most commonly used indicator of research collaboration in the literature.

will be more costly [2]. By matching salary information, which is available from the 1996 Ontario Salary Disclosure Act, with publication characteristics from EconLit and citation counts from Google Scholar and Web of Science (WOS), our research provides a comprehensive picture of the impact of co-authorship on productivity, pay, and promotions among economists in Canada.

Gender inequality has long been a question of great interest in a wide range of fields, and economics is no exception. Some US studies have shown that some gender differences exist concerning pay, the number of publications, the propensity to co-author, and the likelihood of promotion. A recent working paper by Sarsons (2015) suggests that co-authored publications are detrimental to women but benefit men. Our research provides additional insights into male and female gender gaps by providing a unique analysis of the impact of the number of co-authored publications on the likelihood of promotion for males and females in Ontario.

Despite a proliferation of studies analyzing the impacts of co-authorship, there is a paucity of empirical evidence concerning the attributes that facilitate and impede co-authorship [3]. In this analysis, the effect of research and researcher characteristics on the likelihood of collaboration is investigated by matching pairs of economists in Ontario.

The emphasis of this analysis is on evaluating the impact of co-authorship. Overall, our results [4] suggest that there is a significant return to co-authored publications over solo-authored publications in Ontario universities. And the estimates associated with publications in the top-ranked journals are higher than the estimates at of non-top-ranked journals. In addition, our analysis has shown that co-authored publications are significantly associated with higher citation counts for economists; however, this impact varies between top-ranked and non-top-ranked journals. We also distinguish co-authorship types by directly counting the number of different affiliations associated with each publication to evaluate how different formats of collaboration are related to research productivity. We find that US collaboration and international collaboration, on average, generate higher research productivity than other types of collaboration.

To preview the impacts of productivity on promotion, we find that the probability of promotion is positively related to an individual's past productivity, and an additional co-authored publication is associated with a higher likelihood of promotion compared to an

[2] Sauer (1988)

[3] Fafchamps et al. (2006)

[4] The main estimation method used in our analysis is fixed effect. To decide whether to use a fixed effects or random effects estimator, we applied the correlated random effect (CRE) model. Our results suggest that the fixed effects model is preferred over the random effects model.

additional solo-authored publication. Our results also suggest that some gender differences exist in terms of the impact of co-authored publications on the likelihood of promotion.

Our analysis of the effect of research and researcher's characteristics on the likelihood of collaboration among Ontario economists suggests that, in Ontario, economists are more likely to co-author with their colleagues who have similar ability and research interest. There is no gender-sorting effect among Ontario economists.

The paper is structured as follows. Section 2 presents the related literature. Section 3 displays data used in our analysis. Section 4 describes some demographics of co-authorship and co-authors. In section 5, we present a variety of empirical models and results. Section 6 concludes.

2.2 Literature review

2.2.1 Studies related to academic economists in the United States

In the US literature, there has been a considerable amount of studies involving academics with topics including the determinants of co-authorship, male/female publishing patterns, co-authorship and the output of academics, pay and research productivity, aging and research productivity, research productivity and ranking of department, determinants of research productivity at the individual, program and organization level, promotion and job mobility patterns of academics. Table 2.1 summarizes the main empirical studies related to academic economists in the US on the following topics: the determinants of co-authorship, the relationship between co-authorship and pay, co-authorship and research productivity, pay and research productivity as well as mobility and promotion patterns of academic economists.

Extensive research has shown that there is an increasing trend toward co-authorship [5] in the economics. Hudson (1996) shows that the proportion of multi-authored papers published in the Journal of Political Economy and the American Economic Review (AER) increased from 6 percent and 8 percent in 1950 to 39.6 percent and 54.9 percent in 1993, respectively. Laband and Tollison (2000) find that the percentage of co-authored papers published in three economics journals the American Economics Review, the Journal of Political Economy, and the Quarterly Review of Economics increased from 10 percent in the early 1950s to 70 percent in 1994. So why do economists cooperate with each other?

[5] Maske et al. (2003)

Several US studies, as presented in the first panel of Table 2.1, offer possible explanations. Using self-collected time series data on individual papers published in the AER from 1960 to 1985, Barnett et al. (1988) conclude that the increase in co-authorship in economics publications can be explained by specialization, the increasing opportunity cost of time, an growing incentive to avoid uncertainty in the editorial review process. Based on a probit analysis of the impact of JEL subject codes on the probability of co-authorship for every article published in the AER, the Journal of Political Economy, and the Quarterly Journal of Economics from 1886 to 1995, Laband and Tollison (2000) suggest that the increasing incidence of co-authorship may result from greater quantitative content of papers. The authors also find empirical evidence that co-authored papers are more likely to be accepted for publication than single-authored papers. On the other hand, it is also possible that co-authorship has an adverse impact on the research productivity of academics. According to Hudson (1996), collaboration has the following disadvantages: it involves compromise; it may increase organization and communication costs, which may lead to diseconomies of scale [6], and it may have the free rider problem in the sense that the more people working together, the higher the likelihood that an author will contribute less.

The second panel displays the studies on pay and co-authorship of academic economists in the United States. In a widely cited paper, Sauer (1988) reports that a 10-AEQ-page paper in the top journal is associated with a 4 percent increase in salary based on the non-linear least square method and a sample of 140 academic economists from seven economics departments.Additionally, the author finds that the return to a research paper co-authored by n people is $\frac{1}{n}$ times that of a single-authored paper [7]. Using a data set connecting annual salaries for 326 faculty members from top-ranked PhD-granting programs to their peer-reviewed publication histories, Hilmer and Hilmer (2005) study the labor market for academic agricultural economists. The authors find that the estimated return from an additional peer-reviewed article is higher for single-authored articles (approximately 0.8 percent) than the estimated return for an additional co-authored article (approximately 0.3 percent).

Several recent studies investigating academic economists on the subject of co-authorship and research productivity are presented in the third panel of Table 2.1. A much-debated question is how to appropriately measure research productivity and co-authorship. Two of the most popular attempts at quantifying an individual's research productivity are publication counts and citation counts. For co-authorship, co-author status and the number of co-authors are the two most used measures. To date, there has been little agreement on the effect of co-authorship on output. It is possible that collaboration among economists can

[6] Hudson (1996)
[7] Sauer (1988)

enhance their research productivity either because of knowledge spillovers or, as claimed by Fox and Faver (1984), better assessment of the paper because of the social context of the research [8]. It is also possible that co-authorship has an adverse impact on the research productivity of academics. Using panel data for 339 economists, Hollis (2001) shows that collaboration between academic economists tends to yield higher quality publications. However, after the publications are discounted by the number of authors, the relationship between research productivity and collaboration is negative. Similarly, Medoff (2003) believes that co-authorship does not enhance the research quality of economists. However, as the author notes, his estimates may be biased because his research was based on a core set of eight leading journals, which may not provide enough variation in the quality of publications. Based on a large international data set that has information on authorship and citations of all papers published in the field of mathematics over the last 70 years, Borjas and Doran (2012) surprisingly find that the collapse of the Soviet Union had an adverse effect on the productivity of American mathematicians whose research overlapped with the Soviets. However, the authors also conclude that co-authoring with a Soviet reduces this adverse effect.

The fourth panel shows several studies exploring the relation between pay and research productivity in the United States. In 1978, Hansen et al. (1978) performed a three-stage least square analysis based on the 1966 survey data of economists undertaken by the National Register of Scientific and Technical Personnel. They reveal that an additional unit of research productivity is associated with an 8 percent increase in annual salary. Hamermesh et al. (1982) finds that an additional reference adds more to an individual's salary than an extra book or an article based on a sample of 148 full professors of economics in seven large public universities during the years 1979 to 1980. In a widely cited paper, Sauer (1988) reports that a 10-AEQ-page paper in a leading journal is associated with a 4 percent increase in salary based on the nonlinear least square method and a sample of 140 academic economists from seven economics departments during the 1982 to 1983 academic year.

Previous US literature also suggests that there are some differences in salary, promotion, publishing patterns, and the propensity to cooperate with other colleagues between males and females. Controlling for years of experience, institution type, teaching loads, and co-authorship rates, Maske et al. (2003) find that, on average, males publish approximately seven more papers than females. A significant salary difference between men and women is found by Broder (1993). Ferber and Teiman (1980) and McDowell and Smith (1992) report the differences in co-authorship rates between men and women. McDowell and Smith (1992) also find that economists tend to co-author with others of the same gender.

[8] Fox and Faver (1984)

2.2.2 Canadian studies of academic economists

In contrast to the large amount of literature in the United States, Table 2.2 shows that few studies focus on academic economists in Canada. To date, most studies on Canadian academic economists focus on the relationship between research productivity and the assessment of economics departments except for some recently published studies.

Investigating the research productivity of 733 economists holding tenure or probationary appointments at Canadian universities, Lucas (1995) shows that Canadian economists published one single-author-equivalent article every 2.5 years, on average, during the 1980s. By exploring all the publications between 1980 and 2000, Davies et al. (2008) conclude that Canadian economists contribute approximately 5 percent of publications in the leading 10 journals and approximately half of the publications in the Canadian Journal of Economics. These two studies also provide a ranking of Canadian Economics departments based on faculty research productivity.

However, the authors' research has only been carried out in a descriptive manner. In 2012, based on both descriptive evidence and a time-dependent panel regression model, Simpson and Emery (2012) find a declining interest in publishing papers with Canadian content among new faculty hired since 1990 in Canadian Economics departments.

Two recent studies have been conducted on other aspects of the research productivity of academic economists. Conley et al. (2013) investigate how an increase in publication delays affects the life cycle of publications based on a panel data set of 14,271 individuals who obtained PhDs between 1986 and 2000 in US and Canadian economics departments. The authors found a downward trend in publication records. Research productivity and pay is another interesting subject that has received minimal attention in Canada except for one longitudinal study recently published by Sen et al. (2014). Based on a unique panel data of 543 tenured and tenure-tracked professors in 16 economics departments in Canada between 1996 and 2006, a full range of measures for research productivity including the number of publications in peer-reviewed journals, the number of publications in top-10 and top-21 journals, the number of books, and the aggregated number of citations, Sen et al. indicate that a leading journal publication is associated with a 1 percent to 3 percent increase in annual salary.

Collectively, studies of academic economists in Canada are minimal, and a systematic understanding of academic economists is lacking. Do economists who collaborate with others produce higher quality publications than single authors? Is the monetary return of co-authored publications higher than that of single-authored publications? Does the quality and quantity of publications have equal impact on salary and promotion? Most of these questions have not been answered. Our research intends to answer these questions and fill the gap in the Canadian literature.

2.3 Data

Our data set was created by matching information assembled from several different sources including the Ministry of Finance, EconLit, Web of Science (WOS), and online curricula vitae. The finalized data set was composed of information on 571 economists and 3,414 publications from 16 Ontario universities over the period 1996 to 2012.

We first obtained a list of all tenured and tenure-track professors in 16 Ontario economics departments. We then extracted their salary information from the Ministry of Finance's website. Peer-reviewed publication information on all professors in Ontario economics departments during the period 1996 to 2012 was obtained from the EconLit database. The peer-reviewed publication information included the following: the title, the journal name, the number of pages, the number of co-authors, the EconLit subject code, the name of co-authors, and their corresponding affiliations. Based on that information, we constructed measures of pay, research productivity, and co-authorship. To thoroughly evaluate the effect of co-authorship, we also controlled for other individual, publication and university characteristics that may affect the quality of a publication. Our complete economist-year level data set consists of all tenured and tenure-tracked professors in 16 Ontario universities. Because of new hires and the retirement of senior professors, it is an unbalanced panel data. Table 2.3 shows the detailed data sources and variable definitions.

Salary

Annual data on salaries from 1996 to 2012 were obtained on a yearly basis from the Ministry of Finance website. The data were available because of the 1996 Ontario Public Sector Salary Disclosure Act. We only obtained salary information for professors who earned $100,000 or more in the previous year [9]. We used the province-specific consumer index to convert the salary data into constant 1992 dollars. Table 2.4 displays some descriptive statistics of the main variables used in our paper by different categories. Overall, the professors in our sample earned approximately $99,834 per year, and men earned slightly more than women. The average salaries of professors in medical/doctoral universities are higher than professors in primarily undergraduate and comprehensive universities.

Research productivity

[9] Salary information for professors who earned less than $100,000 is not available from the Ministry of Finance website.

In the literature, the accurate measure of research productivity is open to debate. The simplest and earliest attempt at quantifying an individual's research productivity is to count the number of publications in EconLit or the Journal of Citation Reports databases [10]. The total number of citations is another commonly used measure of research productivity. In accordance with the literature, research productivity in our paper is also classified into two broad categories: the number of articles published in peer-reviewed academic journals in the previous year and the total number of citations associated with a given publication.

1. Papers published in peer-reviewed academic journals

Data on papers published in peer-reviewed academic journals were abstracted from the EconLit website [11]. We first evaluated the effects of the total number of publications; then, we estimates the separate effects of the top-10 and top-21 journals [12]. The categorization of top-10 and top-21 journals is based on Kodrzycki and Yu (2006). Additionally, we convert each publication into a number of AEQ equivalent pages (AEQ-page) based on the method proposed by Conley and Önder (2014). Applying the authors' method, the following publication lists are all approximately equivalent to one AER paper: (a) one article in the AER or Econometrica; (b) one and one-half articles in the Journal of Political Economy or Quarterly Journal of Economics; (c) two articles in the Review of Economic Studies, Journal of Econometrics, Econometric Theory, or Journal of Economic Theory; (d) three articles in the Journal of Monetary Economics or Games and Economic Behavior; (e) four articles in the European Economic Review, Review of Economics and Statistics, International Economic Review, or Economic Theory; (f) five articles in the Economic Journal, Journal of Public Economics, or Economics Letters; (g) six to 10 papers in all other journals. We also adjusted for the number of co-authors on a given paper. In other words, if an economist in our sample publishes a paper with n co-authors in a journal with a quality index of Q relative to the AER, then that author is credited with $\frac{Q}{n}$ AER papers.

Table 2.5 provides the breakdown of the total amount of publications per year from 1996 to 2012. The total number of publications are relatively stable while the total number of co-authored publications has been increasing over time.

2. Total number of citations associated with a given publication

The total number of citations is another widely used measure for research productivity

[10] Hilmer et al. (2015)

[11] As I mentioned previously, we first obtained a list of all tenured and tenure-track professors. We then searched their name and affiliated university on EconLit to obtain their publication information.

[12] Check the appendix for details.

[13] in the literature. The two most powerful citation databases are the WOS and Google Scholar. We mainly use the citation counts collected from Google Scholar in our analysis. As noted by Hilmer et al. (2015), the advantage of Google Scholar is that it indexes more journals than WOS. The limitation is that "it is hard to know how it treats multiple versions of a working paper, some of which may have different titles [14]." Therefore, in our sensitivity analysis, we ran regressions using the citation counts collected from WOS to enhance the credibility of our analysis. We also ran regressions using the five-year journal impact factor [15] collected from WOS to check the robustness of our results. The journal impact factor is a proxy indicator of the importance of a journal.

Table 2.6 provides the summary statistics of citations and the journal impact factor. The mean of citation counts collected from WOS is lower than that of the mean of citation counts collected from Google Scholar. Table 2.7 shows the number of citations by the number of authors. The table shows that papers co-authored by two and three authors are cited more often than solo-authored papers.

Co-authorship

We define a co-authored publication as a publication with at least two authors. We also classify different types of co-authorship by strictly counting the number of different affiliations of coauthors [16]. In other words, we classify co-authorship into five groups to determine the impact of different types of co-authorship: inter-department collaboration, intra-department collaboration, domestic collaboration, US collaboration, and international collaboration. This measure of co-authorship is identified by the information collected from EconLit. For each publication, Econlit provides the name and the corresponding affiliation of each author. We first manually coded each author's affiliation excluding the focus author for a given paper into the following categories based on their affiliations: same department, econ-Ontario, non-econ-Ontario, Canada, United States, and others. We then strictly count the number of co-authors from each category. Table 2.8 shows the definition of each type.

Other individual characteristics

We also controlled for gender, experience, teaching rate, job rank, the university attended, and whether a Social Sciences and information Humanities Research (SSHRC) award was held for a given year in our analysis. Gender [17] and job rank were mainly

[13] Hamermesh et al. (1982)

[14] Hilmer et al. (2015)

[15] According to Thomson Reuters Institute for Scientific Information, the five-year journal impact factor is the average number of times journal-published papers have been cited in the journal citation report.

[16] It is a commonly used method in the literature.

[17] Gender is identified by his/her website photos.

identified by each economist's online curriculum vitae. Job rank was categorized as full professor, associate professor, and assistant professor. The university attended was identified from the PROQUEST Dissertation and Thesis database. Based on the location of the university attended, we created dummy variables for whether an economist obtained their highest degree from a US university, a Canadian university, or an other university with the other university as an omitted category. Experience in our analysis was defined as the number of years since an individual received their PhD. This information was also obtained from the PROQUEST Dissertation and Thesis database. We included experience in our analysis for two reasons. First, it is well documented that experience in any job is one of the main factors influencing productivity. Second, as Ductor (2015) shows, more experienced individuals have more contacts. Therefore, individuals with more years of experience are more likely to collaborate with others. Teaching ratings are collected from www.ratemyprofessor.com [18]. Information on the holding of an SSHRC award is abstracted from the Social Sciences and Humanities Research Council of Canada Award search engine [19].

Other paper characteristics

Piette and Ross (1992) and Laband and Tollison (2000) both find that co-authorship patterns often differ by subject and suggest that some fields are more amenable to scholarly interaction than others. To control for this factor, we added a vector of dummies for subjects. Subject dummies were created according to the Journal of Economic Literature (JEL) classification system [20].

Other university characteristics

We also controlled for other university characteristics such as the type of university and whether a university is unionized. Based on Maclean's Magazine categorization of universities, the type of university in our analysis is classified into the following three main groups: medical/doctoral universities, comprehensive universities, and primarily undergraduate universities. The categorization takes the form of a dummy variable in our

[18] www.ratemyprofessor.com is a popular evaluation site where students post their evaluations/-comments on instructors. A growing body of research analyzes the validity and usefulness of www.ratemyprofessor.com. Using a sample of 399 online ratings from ratemyprofessor.com, Otto et al. (2008) evaluate the usefulness and validity of students' online ratings of instructors. The authors conclude that the ratings reflect student learning, and the rating is a valid measurement of student learning. However, care should be taken when interpreting estimates of teaching ratings.

[19] Most Canadian economists' funding is from SSHRC.

[20] JEL subject codes are listed in the appendix. We first collect the Journal of Economic Literature (JEL) codes of each paper from EconLit and based on their JEL codes, we categorize all articles into 19 fields using the first digit of their JEL codes.

analysis with primarily undergraduate universities as omitted categories. According to Sen et al. (2014), salary may also be influenced by university-specific incentive schemes, salary caps, and unionization [21]. Following their procedure, if an individual is from University of Guelph, McMaster University, Queen's University, the University of Toronto, Western University, Laurentian University, or the University of Waterloo, we create a "merit pay and no salary cap" dummy variable equal to 1 and 0 otherwise. We also include a dummy variable to capture the cross-university and time-series variation in unionization among Ontario universities. The detailed information on university characteristics is presented in Table 2.9.

2.4 Some Demographics of Co-authorship and Co-authors

We explore some demographics of co-authorship and co-authors in this section. We first track the change in percentage of co-authored papers over time, we then look at the variation in the average number of authors per paper, and we visualize the research community of Ontario economists with others based on their co-authorship relations.

1. Percentage of co-authored papers

Figure 2.1 shows the share of co-authored papers for the period 1996 to 2012. The green line represents co-authored papers as a percent of the total number of publications. Overall, there is a slow but relatively stable increase in the amount of collaborative work among Ontario economists over time, from around 68 percent in 1996 to 80 percent in 2012. The orange and blue lines show co-authored papers published in the top-21 and top-10-ranked journals, respectively. Figure 2.1 shows that the percentage of co-authored papers published in the top-10 and top-21-ranked journals are higher than the percentage of single-authored papers.

Figure 2.2 plots the number of co-authored publications against the number of solo-authored publications by university. From the graph, we can see that all Ontario universities produce more co-authored publications than solo-authored papers while the ratio of co-authored publications is different across universities.

Figure 2.3 shows the breakdown of single and co-authored papers over time by gender. The left panel illustrates the percentage of single and co-authored papers for male

[21] According to Sen et al. (2014), the University of Guelph, McMaster University, Queen's University, the University of Toronto, Western University, Laurentian University, and the University of Waterloo are universities in Ontario that do not impose a salary cap and have merit pay.

economists in Ontario, and the right panel shows the patterns for females. Overall,the graph indicates that there is an increasing trend in the percentage of co-authored publications for both men and women. For the number of solo-authored publications, the trend is different. The graph shows that there has been a small, relatively stable, decrease in the proportion of solo-authored papers published by males. However, the number of solo-authored publications has fluctuated for females.

2. Number of authors

Table 2.10 shows the distribution of the number of authors per paper. The most common number of authors per economics paper is two. Table 2.11 lists the breakdown of the number of papers and the percentage of publications over time by different numbers of authors. As can be seen, the share of single-authored economics papers has been steadily decreasing over time. In 1996, approximately one-third of all publications had only one author, this proportion decreased to approximately 20 percent in 2012. On the other hand, there has been a gradual rise in the number of papers with at least two authors.

3. Research community based on co-authorship relations of Ontario economists

We are trying to understand the research community based on co-authorship relations of Ontario economists with others in this subsection.

Figure 2.4 shows the percentage of papers published by Ontario economists that were the product of collaboration with their colleagues and the percentage of papers co-authored with someone from a US university over time. It is apparent from the graph that the percentage of papers published by authors from the same university is higher than the percentage of papers co-authored with someone affiliated to a US university. However, the share of papers published with an author from a US university has been gradually increasing over time from approximately 15 percent in 1996 to approximately 25 percent in 2012.

Table 2.12 displays the number and the percentage of papers published in top-10 and top-21- ranked journals by two different types of collaboration: same university collaboration and US collaboration. The table shows that for publications in top-10-ranked journals, the share of same-university collaboration is 21 percent, which is smaller than the proportion of US collaboration at 35 percent. A similar pattern for publications in top-21-ranked journals can be observed from the second row of the table.

3.1. Co-authorship network among Ontario economists

Following the literature of visualizing networks of scientific research, by using authors as nodes and their co-authorship as edges, we studied and visualized the links between nodes and co-author networks. This visualization provides a sense of the structure of the research community among Ontario economists.

Figure 2.5 shows the co-authorship network among Ontario universities (1996 to 2012). The colors identify different universities, and the thickness of the links determines the number of collaborations. Two general conclusions can be drawn from this figure: for each university, within university collaboration is the most popular type of co-authorship. Overall, Ontario universities are connected by research collaboration although nearby universities are more connected. For instance, economists at the University of Toronto have greater research collaboration with economists at the University of York than other universities, and economists at the University of York are more likely to collaborate with economists at the University of Toronto and the University of Ryerson.

Figure 2.6 shows the co-authorship network among Ontario economists. In this figure, colors identify different universities, and each node represents an economics professor who co-authored a publication with another economics professor from Ontario. The co-authorship network among Ontario economists can be divided into a large sub-community and many small sub-communities, as indicated by the clusters of the nodes, and, approximately, these sub-communities correspond to locations of each university. The big sub-communities include the University of Toronto, University of Guelph, University of Waterloo, University of Ottawa, Queen's University, McMaster University, and one smaller cluster with Brock University, Carleton University, and the University of Ottawa. There are also several tiny clusters. Additionally, economists at the University of Toronto and the McMaster University co-author papers with their colleagues more often than economists in other universities since there are relatively more links within those two universities.

3.2. The global distribution of Ontario's economics co-authors

In addition to the figures showing the co-authorship network among Ontario economists, Figure 2.7 shows the global distribution of Ontario's economics co-authors. The colors identify different countries. The economists in Ontario universities collaborate globally, and the United States and the United Kingdom are the two most important countries for Ontario economists' partnerships.

4. Gender differences

Table 2.13 presents the percentage of different types of publications by gender. Column (1) shows the results based on all data while columns (2) and (3) contain the corresponding statistics for males and females, respectively. The table shows that in Ontario, female economists are just as likely as male economists to engage in long-distance scholarly team production. There are no significant gender differences in this respect.

The demographics of co-authorship and co-authors have been described in this subsection. The findings show slow but stable growth in the share of co-authored publications relative to single-authored publications. At the university level, all Ontario universities produce more co-authored publications than single-authored publications. The most common number of authors per publication is two. Additionally, co-authorship with researchers affiliated with the same university is an important type of co-authored publication, and the incidences of co-authorship with researchers from US universities are slowly increasing over time. The network visualization of Ontario economists shows that researchers from certain universities co-author with their colleagues more often than others. Finally, there are no significant gender differences in long-distance scholarly team production.

2.5 Empirical Model and Results

For our empirical model and results we first look at the relationship between co-authorship, pay, and productivity. We then analyze the effect of past productivity and co-authorship on promotion. The final objective of our paper is to evaluate the effect of research and researcher's characteristics on the likelihood of co-authorship.

2.5.1 Co-authorship and productivity

In this subsection, we focus on the impact of co-authorship on productivity. We first test whether co-authorship increases the productivity of Ontario economists. Then, we strict our data to all co-authored publications to consider the effects of different types of co-authorship.

Productivity is measured by the total number of publications

Research productivity in our analysis is classified into papers published in peer-reviewed academic journals and the total number of citations associated with a given publication. We evaluate the effect of co-authorship on productivity using the total number of publications published in corresponding journals as the measure of productivity first.

We start our analysis by estimating the following econometric model:

$$Y_{it} = \beta_0 + \beta_1 PROPCO_{it} + \beta_2 Z_{it} + \beta_3 X_i + \alpha_u + \gamma_t + \epsilon_{it} \tag{2.1}$$

where $PROPCO_{it}$ is defined as (1), the number of top-10 articles that have been co-authored as a proportion of all top-10 articles aggregated over a three-year period; (2) the number of top-21 articles that have been co-authored as a proportion of all top-21 articles over a three-year period; (3) the number of other articles co-authored as a proportion of other articles over a three-year period; Y_{it} is the total number of publications published in corresponding journals over a three-year period; α is university-specific fixed effect; γ is a vector of year fixed effect; Z_{it} is a vector of time-variant individual characteristics including teaching rate, possession of an SSHRC award, job rank, experience, and experience squared; X_i is other time-invariant individual characteristics, such as gender, and a dummy for whether individual i obtained the highest degree from a US or Canadian university. ϵ_{ist} is an idiosyncratic error term.

The OLS estimates of the effects of co-authored publications in the top-10, top-21, and non-top-ranked journals are presented in Table 2.14. The estimates of the proportion of co-authored publications in columns (1) and (2) are negative and statistically insignificant, and the estimates in other columns are all positive and statistically significant. The estimates in columns (3) and (4) suggest that a one percentage point increase in the proportion of co-authored top-21-ranked publications is associated with approximately 0.7 additional top-21-ranked publications. The corresponding estimates of other publications, which are listed in columns (5) and (6), are smaller.

We also add lagged dependent variables [22] on the right-hand side to control for the possibility that if someone has a history of top-ranked publications, they are more likely to publish in top-ranked journals currently. The estimated coefficients are shown in Table A.6 in the appendix.

Productivity is measured by the number of citations the authors receives for all publications published in that year

We use the number of citations the author receives for all publications published in a given year as the measurement of research productivity in this subsection. We start our analysis by estimating the following econometric model:

$$InC_{it} = \beta_0 + \beta_1 NUMs_{it} + \beta_2 NUMc_{it} + \beta_3 Z_{it} + +\beta_4 X_i + \alpha_u + \gamma_t + \epsilon_{it} \tag{2.2}$$

where C_{it} is the number of citations individual i received for all publications published at time t. Because the distribution of citation maybe skewed to the right, following the

[22] If first years are 1999 to 2003, then, in addition to variables that measure proportions of different articles, we add the number of corresponding articles in 1996 to 1998 in our specification.

literature, we use the log transformation of citations to minimize the impact of highly productive individuals on our estimates; $NUMs_{it}$ and $NUMc_{it}$ are defined as the total number of single and co-authored publications of individual i, at time t; the definition of all other variables are all the same as in equation 2.1; ϵ_{ist} is an idiosyncratic error term.

Table 2.15 contains the effects of co-authored publications relative to solo work. We begin our analysis by including all individual characteristics in column (1), then, we add year fixed effect in the second column. In column (3), we include other university characteristics. Finally, we add university fixed effects in the last column. Across all specifications, the estimated coefficients of single-authored publications and co-authored publications are positive and statistically significant. Additionally, the coefficient estimates show that the number of co-authored publications is higher than the coefficient estimates of solo-authored publications.

The separate effects of publications in the top-10 and top-21-ranked journals are presented in Table 2.16. The first two columns contain the impacts of top-10 publications, and the last two columns show the effects of top-21 publications. First, the impacts of top journal publications are higher than the impacts of non-top-ranked journals. For publications in top-21-ranked journals, the coefficient estimates of co-authored publications are greater than the coefficient estimates of solo-authored publications. However, the coefficient estimates are almost the same for co-authored publications and solo-authored publications for top-10-ranked journals.

The effects of different types of co-authorship on productivity

We restrict our data into all co-authored work to explore the effect of different types of co-authorship. We expect that different types of co-authorship have different impacts on productivity. Although the effect of geographical distance has been diminishing overtime because technological progress has made the exchange of ideas much easier and cheaper [23], factors such as academic tradition in different countries may still play a role in determining the effect of different types of co-authorship. It is possible that the relationships between international collaboration are greater in number than domestic collaboration relationships since problem-solving and holding discussions with researchers from another country might inspire different ideas and, thus, have more positive influence on publications [24]. On the other hand, the cost of communication with international researchers in terms of time and convenience are still higher than the cost of communication with domestic researchers. The

[23] Laband and Tollison (2000)

[24] Using a data set of 2.5 million US scientific publications for the years 1985 to 2008, Freeman and Huang (2014) find that the diversity of authors' ethnicity, location, and references is associated with greater scientific contributions (measured by citations).

high price of inputs may have an adverse effect on the quality of the publication.

The primary econometric model is summarized as follows:

$$InC_{itj} = \beta_0 + \beta_2 Inter_{ij} + \beta_3 Intra_{ij} + \beta_4 Domestic_{ij} + \beta_5 US_{ij} + \beta_6 International_{ij} + \alpha + \gamma + \epsilon_{itj} \tag{2.3}$$

where C_{itj} is the total number of citations of paper i at time t by individual j; $Inter_{ij}$, $Intra_{ij}$, $Domestic_{ij}$, US_{ij} and $International_{ij}$ are the number of different types of co-authors on paper j. We also include some other university and individual characteristics, such as a dummy for whether individual i obtained their highest degree from a US, Canadian or other university; α is the university specific fixed effect or individual fixed effect in some regressions; γ is a vector of year fixed effect; and ϵ_{itj} is an idiosyncratic error term. Standard errors are clustered at the individual level. The main parameter of interest is β, which captures the effect of different types of co-authorship on productivity.

Table 2.17 presents the impact of the different types of co-authorship. In column (1), the OLS regression includes five collaboration variables and all other control variables. We then add year dummies in column (2) but not university dummies and individual dummies. We add university dummies in column (3). In column (4), we replace university dummies with individual dummies, and we drop all the other time-invariant dummies to overcome multicollinearity. Most of the collaboration variables are positively related to a publication's quality as measured by the total number of citations adjusted by year. Both US collaboration and international collaboration are statistically significant across all specifications. The estimates of domestic collaboration are statistically significant in some of the specifications. However, the estimates of inter-department collaboration and intra-department collaboration are never statistically significant.

For the sensitivity test, the upper panel of Table 2.18 provides the OLS estimates using the total number of citations and the five-year journal impact factor collected from WOS as alternative measures of research productivity. We find strong evidence of the impacts of US collaboration and international collaboration. Overall, our results suggest that US collaboration and international collaboration are significantly associated with higher research productivity, which is measured by citation counts.

We limit our sample to publications in the top-ranked journals, and the estimates are displayed in the lower panel of Table 2.18. The estimates of US collaboration are notable in the table, and all are negative and statistically insignificant. International collaboration is still positive and statistically significant across all columns, and the estimates of international collaboration are much higher than those presented in Table 2.16.

To compare our findings with others, Table 2.19 presents selected empirical works related to ours. Using a self-constructed panel data set of 65 biomedical scientists at a New Zealand university, [25] find that both international collaboration and within-university collaboration are positively related to the quality of a publication regardless of the choice of research productivity and model specification. When the authors use the citation counts a paper received in a two-year window as their measure of research output, their coefficient estimates of international collaboration, domestic collaboration, and within university collaboration are approximately 15 percent, 1.6 percent, and 19 percent, respectively. Our estimates for international collaboration and domestic collaboration are comparable with He et al. (2009).

We also define types of co-authorship as: (1) proportion of co-authored articles in all journals with the same university colleagues; (2) proportion of all co-authored publications with other Canadian colleagues; (3) proportion of all co-authored publications with US authors; (4) proportion of co-authored publications where co-authors are from the United States and Canada; (5) proportion of other types of co-authored articles. The estimates are presented in Table A.7 in the appendix. The first two columns list the impacts of top-10 publications, the next two columns present the effects of top-21-ranked journals, and the last two columns show the effects of other journals. The estimates for the proportion of co-authored articles in all journals with the same university colleagues and proportion of co-authored publications with US authors are positive across all columns. However, the estimates for the proportion of all co-authored publications with other Canadian colleagues are not consistently positive. Additionally, the estimates of the proportion of co-authored publications, where co-authors were from the United States and Canada, are negative and statistically significant across all columns except column (3).

2.5.2 Pay and co-authorship

In this subsection, we evaluate the returns of co-authored publications relative to solo-authored publications. This is an important factor because there is a growing body of co-authored papers published over time, and the rational agencies intend to participate in activities that are associated with a higher relative return.

Following Sauer (1988) and other related studies, we estimate the summarized econometric model as follows:

$$InY_{iut} = \beta_0 + \beta_1 NUMs_{iut} + \beta_2 NUMc_{iut} + \beta_3 X_i + \alpha_u + \gamma_t + \epsilon_{iut} \tag{2.4}$$

[25] He et al. (2009)

where Y_{ist} is the annual salary of individual i at time t in university u; $NUMs_{iut}$ and $NUMc_{iut}$ are defined as the total number of single and co-authored publications of individual i, at time t in university u; α is university-specific fixed effect; γ is a vector of year fixed effect. We also control for other time-invariant individual characteristics, such as gender, and a dummy for whether individual i obtained their highest degree from a US or Canadian university. ϵ_{ist} is an idiosyncratic error term.

Table 2.20 summarizes the OLS results obtained from the preliminary analysis of the relationship between pay and co-authorship. The standard errors of all estimates are clustered at an individual level. To conserve space, we do not report all the coefficient estimates.

The empirical estimates of Table 2.20 are displayed as follows. We start our analysis with the coefficient estimates of equation 2.4 with all individual characteristics. The estimate for single-authored publications is not statistically significant. However, the estimate for co-authored publications is statistically significant at the 1 percent level. Specifically, the estimate of co-authored publications is approximately 0.0219, which implies that an additional co-authored paper can increase a salary by approximately 2 percent. We added year fixed effect in column (2); the coefficient estimates do not vary much. In column (3) [26], we add all other university characteristics. Both the estimates of single-authored and co-authored publications drop substantively. To account for university-specific differences in salary structure, we add university fixed effect in column (4). The estimates for co-authored publications are approximately 0.011, which is higher than the estimates for single-authored publications (0.005). In column (5), we added university-specific time trends to allow each university its own time trend. Despite some minor variation, the estimates are similar to those displayed in column (4). Columns (6) and (7) replicate columns (4) and (5) but with a smaller number of years: 1996, 1999, 2003, 2008, and 2012. The estimates for single-authored and co-authored publications are higher than those presented in columns (4) and (5).

Turning to other controls, experience and the dummy for holding an SSHRC award are positive and statistically significant across all columns. The estimates for gender dummy are never statistically significant. Obtaining a PhD from a Canadian university is negative and statistically significant in columns (1) and (2). After we add year fixed effect in column (3) and university fixed effect in column (4), obtaining a PhD from a Canadian university

[26] To decide whether to use a fixed effects or random effects estimator in our analysis, we applied the correlated random effect (CRE) [27] model. Our results suggest that a fixed effects model is preferred over a random effects model. We first compute the panel-level average of our time-varying covariates. Then, we use a random effects estimator to regress our covariates and the panel-level means are generated against our outcome. Finally, we test whether the panel-level means generated in the first step are jointly zero.

becomes statistically insignificant.

Returns to co-authored publications in the top ranked journals

We expect that the returns to co-authored and solo-authored publications vary with journal quality. Table 2.20 contains the estimates of the effect of the total number of single-authored publications and co-authored publications. In Table 2.21, we look at the separate effects of publications in the top-10 and top-21-ranked journals. The first two columns contain the impacts of top-10 publications, and the following two columns present the effects of top-21-ranked journals. To replicate more variation, we restrict our sample to a smaller number of years: 1996, 1999, 2003, 2008, and 2012 in the last four columns.

Table 2.21 shows that the estimates of co-authored publications in the top-10 and top-21-ranked journals are statistically significant across all columns. The results in columns 1 and 2 suggest that an additional co-authored top-10 paper can increase a salary by approximately 5 percent. For solo-authored publications in the top-10-ranked journals, the estimates are slightly higher than the estimates for co-authored publications. On the other hand, for publications in the top-21-ranked journals, the estimates for co-authored publications are higher than the estimates for solo-authored publications, and the estimates for solo-authored publications are statistically insignificant. The estimates in columns 5 to 8 are comparative to those presented in columns (1) to (4).

This subsection has evaluated the return to co-authorship in Ontario universities. Our results suggest that there is a statistically significant positive return for co-authored publications compared to solo-authored publications. We find that an additional co-authored publication can increase salary by approximately 1 to 2 percent. As expected, the estimates associated with top-ranked publications are higher.

2.5.3 Pay and productivity

Having analyzed the relationship between co-authorship and productivity, we evaluate the relationship between pay and research productivity. Given the positive effect of co-authorship on productivity, we adjust for the number of co-authors on a given paper in some regressions.

This section is an update of Sen et al. (2014)'s study. Following Sen et al. (2014), our primary econometric model is summarized as follows:

$$InY_{iut} = \beta_0 + \beta_1 X_{iut} + \beta_2 Z_{iut} + \alpha + \gamma + \epsilon_{iut} \tag{2.5}$$

where Y_{ist} is the annual salary of individual i, at time t, in university u; X_{iut} is research productivity of the individual; Z_{iut} is the measure of time-variant individual characteristics such as teaching rate; we also include some time-invariant individual characteristics such as gender and a dummy for whether individual i obtained their highest degree from a US or Canadian institution; α is university specific fixed effect; γ is a vector of year fixed effect; ϵ_{ist} is an idiosyncratic error term.

As mentioned previously, research productivity in our analysis is classified into papers published in peer-reviewed academic journals in the previous year and a total number of citations associated with a given publication aggregated over a year. We evaluate the effect of productivity on pay using those two measures.

Productivity is measured through publications in peer-reviewed academic journals in the previous year

We estimate the effects of the total number of publications and the separate effects of publications in the top-10 and top-21-ranked journals. Additionally, following the literature, we used the AER adjusted measure of research productivity in some regressions. In other words, we convert each publication into several AER-equivalent papers, and we also adjust for the number of co-authors.

Table 2.22 presents baseline OLS estimates of the effects of productivity on pay. The first two columns show the estimates of publications in the top-10-ranked journals. Columns (3) and (4) contains the results of papers in the top-21-ranked journals. Column (5) and (6) display the results of all publications. The last two columns present the estimates where we convert each publication into a number of AER papers. All specifications in Table 2.22 include all controls and year fixed effects. The difference between the first and second column of each category is that we add university fixed effect in the second column. Standard errors of coefficient estimates are clustered at the individual level.

Most estimates of publications are significant at the 1 percent level. The estimate for publications in the top-10-ranked journals (column (1)) suggests that an additional top-10 publication is associated with a 3 percent increase in an individual's annual salary, which is higher than that of non-top publications (0.006). To control for the university-specific effect, we add the university dummy in the second column. Interestingly, the estimate for top-10 publications shrinks. The estimates in column (3) and (4) are similar to the estimates in columns (1) and (2). For all publications, our results imply that a publication is associated with a 0.8 percent rise in annual salary. The estimates in columns (7) and (8) suggest that an additional AER publication is associated with an approximate 0.2

percent to 0.3 percent increase in annual salary. Using quality adjusted AEQ-pages for each individual as their publication measure, Sauer (1988) finds that the estimates of the return to a 10-AEQ-page article are 0.0030 to 0.0033, which is similar to our findings.

Experience is positively associated with higher earnings across all columns. The estimates for the male dummy are statistically insignificant and suggest that there is no substantial gender difference. For full professors, the estimates are statistically significant across all columns. However, the estimated coefficient of associated professors is significant only when we add university fixed effect. The estimates for the merit pay and no salary cap dummy variable are never statistically significant. The estimates for the union dummy are statistically significant in columns (1), (3), (5), and (7). However, the estimated coefficients are statistically insignificant in the other columns.

Productivity is measured by the total number of citations associated with a publication

We separately estimate the effects of citation and co-authorship adjusted citation. For co-authorship adjusted citation, we weight each citation by $\frac{1}{n}$, where n is the total number of authors of a paper. The first two columns of Table 2.23 show the empirical estimates of the effects of citation, and the last two columns present the effect of co-authorship-adjusted citation. The estimates of citation are statistically significant across all specifications, and the estimates imply that a 1 percent increase in research productivity is associated with an approximately 0.008 percent to 0.01 percent increase in annual salary.

A limitation of OLS in our case is that when the variable was censored, it provided inconsistent estimates of the parameters. Because we do not observe salaries below \$100,000, we have to either impute a wage for all economists although they are not on the "sunshine list" and use this imputed salary to estimate our equation, or we can apply tobit regression. Following Sen et al. (2014), we use tobit regression to overcome the limitation of OLS estimates. The marginal effects obtained from tobit regression are presented in Tables A.8 and Table A.9 in the appendix. There is no substantial difference between OLS estimates and the derived tobit estimates.

In this subsection, we evaluate the relationship between pay and productivity. Using two of the most commonly used measures of research productivity, number of publications and citation counts, our analysis suggests that salary is associated with higher research productivity.

2.5.4 Promotion and productivity

In this subsection, we match the likelihood of promotion of all professors in 16 Ontario economics departments to their past performance and other individual characteristics to determine how past productivity affects the likelihood of their promotion.

Table 2.24 presents the weighted average of publications before and after promotion. The first panel represents the summary statistics for all job ranks. We then focus on promotions from assistant professor to associate professor and from associate professor to full professor in the second and third panel, respectively. The three panels show that the weighted average of publications over three years before promotion is higher than the weighted average of publications over three years after promotion. The result is similar for the five-year weighted average.

Following Coupé et al. (2005), we begin with the following probit regression:

$$PROM^{*}_{iut} = \beta_0 + \beta_1 * PP_{iut-1} + \beta_2 * EXP_{iut} * PP_{iut-1} + \beta_3 * EXP_{iut} + \beta_4 * EXP^{2}_{iut} + \epsilon_{iut} \tag{2.6}$$

where $PROM_{iut} = 1$ if $PROM^{*}_{ist} > 0$

$PROM_{iut} = 0$ if $PROM^{*}_{iut} \leq 0$.

PP_{iut-1} is individual i's past performance. We first estimate the effect of short-run past performance where PP_{iut-1} is the average of publications from year t-3 to t-1. Then, we estimate the impact of relative long-run past performance where PP_{iut-1} is defined as the average of publications from year t-5 to t-1.

Table 2.25 shows the marginal effects from the probit estimates of the impact of past performance on promotion. The first two columns display the results for all job ranks. We then focused on those promotions from assistant professor to associate professor in column (3) and column (4) [28] and from associate professor to full professor in the last two columns. For each category, the differences between the first and second columns are that the estimates in the first column are based on a model where we estimate equation 2.6 with short-run past performance. The estimates in column (2) use relative long-run past performance. The estimate of past performance in columns (1) and (2) suggest that past performance is positively associated with the likelihood of promotion. We also observe that

[28] We restrict our data to those who were never promoted to full professor in our sample. In other words, we focus on economists at the level of associate professor at the last stage of their careers in our sample period.

the interaction of past performance with experience is negatively related to the promotion probability in columns (1) and (2). The results in columns (3) and (4) are similar to the results in columns (1) and (2). The estimates from the regression where we focused on promotions from associate professor to full professor are different. The estimates of past performance, as well as the interaction of past performance with experience, are not statistically significant associated with the likelihood of promotion in columns (5) and (6). Examining the mobility and promotion patterns of 1,000 top economists over 30 years, Coupé et al. (2005) find that the probability of promotion is positively related to an individual's past production in the early stages of their career. The authors' estimates are approximately 3 percent to 9 percent, and our findings are in accord with their conclusions.

The relative effect of co-authored publications and solo-authored publications on the likelihood of promotion

We investigate the relative effect of co-authored publications and solo-authored publications on the likelihood of promotion. A recent working paper [29] by Sarsons (2015) suggests that an additional co-authored paper will increase the probability of gaining tenure by 8 percent, and another single-authored publication will increase the tenure probability by 7.3 percent. Additionally, the author notes that co-authored publications are detrimental to women but benefit men. We determine if there is any difference between male and female concerning the impact of the number of co-authored publications on the likelihood of promotion in Ontario.

Following Sarsons (2015), we estimate the following econometric model:

$$\begin{aligned} PROM^*_{iut} = \beta_0 + \beta_1 * TOTCO_{iut-1} + \beta_2 * TOTSO_{iut-1} + \beta_3 * (TOTCO_{iut-1} * Male) + \\ \beta_5 * (TOTSO_{iut-1} * Male) + \beta_6 * EXP_{iut} + \beta_7 * EXP^2_{iut} + \beta_8 * Male + \alpha + T + \epsilon_{iut} \end{aligned} \tag{2.7}$$

where $PROM_{iut} = 1$ if $PROM^*_{ist} > 0$

$PROM_{iut} = 0$ if $PROM^*_{iut} \leq 0$.

$TOTCO_{iut-1}$ and $TOTSO_{iut-1}$ are the total co-authored publications at time t-1 and total solo-authored publications at t-1, respectively.

Table 2.26 shows the marginal effect estimates from probit regression. For each category, the differences between the first and second column are that we added the interactions of the number of co-authored/solo-authored publications to the male dummy in the second column. The estimated coefficient in column (1) suggests that an additional co-authored

[29] Sarsons (2015) uses curriculum vitae data from economists who had tenure between 1985 and 2014 in 30 universities in the United States.

publication is associated with an approximate 8 percent increase in the probability of being promoted and an additional solo-authored publication is associated with an approximately 6 percent increase. The estimates in column (2) suggest that some gender differences in terms of the impact of co-authored publication on the likelihood of promotion do exist in Ontario since the estimates of solo-authored publication are higher than that of co-authored publication for women.

2.5.5 The effect of research and researcher characteristics on the likelihood of collaboration

In the literature, some studies use graph theory to study the effect of co-authorship networks on the likelihood of collaboration among researchers. Employing a database of all publications in economic journals over a 30-year period, Fafchamps et al. (2006) find that controlling for all other factors, the likelihood of a new co-authorship occurs faster if the two authors are more closely connected through collaboration with others. The authors also find that the greater the differences in research output between two authors, the greater the likelihood that they are working together. In this subsection, we adopt the methodology proposed by Fafchamps et al. (2006) to analyze the effect of research and researcher characteristics on the likelihood of collaboration among Ontario economists.

Variable construction

Likelihood of co-authorship

We begin our analysis by constructing co-authorship variable y_t^{ij} by pairing economists. If Ontario economist i and j co-authored a paper at time t, then, we say the co-authorship variable $y_t^{ij} = 1$ at time t and 0 otherwise. We then find out the overlapping period of each pair of economists. The beginning year of each pair of economics is determined by: $t_0^{ij} = max(t_0^i, t_0^j)$, where t_0^i is individual i's first year of a publication with an Ontario university as affiliation. A similar identification strategy for the last possible year of each pair of economists to co-author is determined by the last year of publication with an Ontario university as an affiliation.

Network

For each economist, we construct the network variable N_t^i by looking at all co-authors from all their peer-reviewed publications in all years before t. Based on N_t^i, we then construct the overlapping network variable D_t^{ij}. D_t^{ij} takes the value of 1 if i and j both published a paper with k before they published together, and it equals 0 if i and j are unconnected by others [30].

D_t^{ij} is identified by the following procedure: we first construct the network variable N_t^i and N_t^j for each pair of economists, then, for each name in N_t^i, we conduct a grid search in N_t^j to determine out whether it appeared. If i and j both co-authored with k, we then determine in which year i co-authored with k and in which year j co-authored with k. Then, we compare it to t^{ij}. If t^{ik} and t^{jk} are both smaller that t^{ij}, we assign value 1 to the variable D_t^{ij} and 0 otherwise.

Overlapping research interest

Following Fafchamps et al. (2006), the measure of overlapping research interest γ_t^{ij} is constructed based on the following equation:

$$\gamma_t^{ij} = \frac{\sum_{f=1}^{F}(x_{t,f}^i x_{t,f}^j)}{\sqrt{(\sum_{f=1}^{F}(x_{t,f}^i)^2)(\sum_{f=1}^{F}(x_{t,f}^j)^2)}}$$

where $x_{t,f}^i$ is the total fraction of articles published by author i in the field f at year t and F is the number of fields. We first collect the Journal of Economic Literature (JEL) codes of each paper from EconLit and categorize all articles into nine sub-fields using the first two digits of the JEL codes. We then compute $\sum_{f=1}^{F}(x_{t,f}^i)^2$ for each economists and $\sum_{f=1}^{F}(x_{t,f}^i x_{t,f}^j)$ for each pair of economists over each year. We finally construct the overlapping research interest variable γ_t^{ij} based on above equation.

Same gender

Some studies in the gender-sorting literature find that researchers are more likely to select co-authors of the same gender [31]. Hence, we capture this effect by including G^{ij}, which is a variable indicating whether a pair of economists are of the same gender. The gender of each individual is identified by the economists profile photo. If two people of the same gender co-authored a paper, we assign the value of one to the variable G^{ij} and zero otherwise.

Research ability

Following Fafchamps et al. (2006), we include $\bar{Q_t}^{ij}$, which is the average number of published papers by i and j at t-1 and $\triangle Q_t^{ij} \equiv \mid Q_t^i - Q_t^j \mid$, which is a measure of differences in research ability of i and j in our analysis.

[30] Any direct link between i and j is ignored.
[31] McDowell and Smith (1992)

Same affiliation

The propensity to co-author with others may also vary with affiliation. We include a variable COM_t^{ij} to indicate whether i and j are from the same university. Table 2.27 shows the percentage of types of co-authorship. Within Ontario, co-authorship among colleges is the most common type of collaboration [32].

Research experience

To test the propensity of junior-senior economist collaboration and capture the effect that more experienced individuals may have more contacts and are, therefore, more likely to collaborate with others [33], we include $\bar{Exp_t}^{ij}$ and $\triangle EXP_t^{ij}$ in our analysis. $\bar{Exp_t}^{ij}$ is the average number of years of experience of i and j. $\triangle EXP_t^{ij} \equiv \mid EXP_t^i - EXP_t^j \mid$. Experience is defined as the number of years since an individual obtained their PhD.

Same graduate school

We consider that individuals who attend the same graduate school may have a greater propensity to collaborate. Hence, we construct a dummy variable $GRAD^{ij}$ to capture this effect. $GRAD^{ij} = 1$ if individual i and j went to the same graduate school and 0 otherwise.

Table 2.28 displays the summary statistics for the main variables used in this analysis.

Empirical model and results

Following Fafchamps et al. (2006), we estimate the following model:

$$\begin{aligned} PROB(y_t^{ij} = 1) = \beta_0 + \beta_1 * D_t^{ij} + \beta_2 * \bar{Q_t}^{ij} + \beta_3 * \triangle Q_t^{ij} + \beta_4 * \gamma_t^{ij} + \beta_5 * \bar{Exp_t}^{ij} \\ + \beta_6 * \triangle EXP_t^{ij} + \beta_7 COM_t^{ij} + \beta_8 * GRAD^{ij} + \beta_9 * G^{ij} + \epsilon_t \end{aligned} \tag{2.8}$$

where $PROB(y_t^{ij} = 1)$ is the likelihood of co-authorship; $y_t^{ij} = 1$ is defined as follows: if authors 1 and 2 co-author a paper at time t, then $y_t^{ij} = 1$ takes the value 1 at t_1^{ij} and 0 otherwise; D_t^{ij} is the network variable; $\bar{Q_t}^{ij}$ is the average number of published papers by i and j at t-1; $\triangle Q_t^{ij} \equiv \mid Q_t^i - Q_t^j \mid$; γ_t^{ij} is a measure of overlapping research interest; $GRAD^{ij}$ indicates whether i and j went to a same graduate school or not; we also include COM_t^{ij} to indicate whether i and j are having a same affiliation; G_{ij} is the same gender variable. It takes the value of 1 if i and j have the same gender, 0 otherwise; $\bar{Exp_t}^{ij}$ is the

[32] Nearby university in this subsection means universities located in the same city.

[33] Ductor (2015)

average number of years of experience of i and j; $\triangle EXP_t^{ij} \equiv \mid EXP_t^i - EXP_t^j \mid$; ϵ_t is an idiosyncratic error term.

The marginal effects from probit regression are presented in Table 2.29. In the first column, we include all the variables described. We then add year fixed effects in column (2). In column (3), we add pair fixed dummies.

Our results suggest that in Ontario, same university collaboration is more likely because the estimates of "the same affiliation" are positive and statistically significant. In addition, highly productive economists are more likely to work together since the estimates of $\bar{Q_t}^{ij}$ are positive and statistically significant, and the estimates of $\triangle Q_t^{ij}$, which is a measure of differences in research ability, is negative and statistically significant. The estimated coefficient of overlapping research interest is positive and statistically significant across all columns, which suggests that when the degree of research overlapping increases by one unit, the likelihood of collaboration increases by approximately 0.1 to 0.2 percentage points. Moreover, our results suggest that there is no gender-sorting effect among Ontario economists.

2.6 Conclusion

This study has shown that there is a significant return to co-authored publications relative to solo-authored publications in Ontario universities. The investigation of the relationship between co-authorship and productivity has shown that co-authored publications are associated with higher citation counts, U.S. collaboration, and international collaboration is significantly related to higher research productivity. Our research has also demonstrated that higher quality publications have a greater effect on salary.

By matching the likelihood of promotion of each professors with their past research performance and other individual characteristics, we show that the likelihood of promotion is positively associated with past performance and the number of co-authored publications matters differently for males and females. This finding is consistent with the relevant literature. Using curriculum vitae data from economists who had tenure between 1985 and 2014 in 30 universities in the United States, Sarsons (2015) suggests that an additional co-authored paper will increase the probability of gaining tenure by 8 percent, and another single-authored publication will increase the tenure probability by 7.3 percent.

Finally, we pair up each professor in Ontario and find that in Ontario, economists are more likely to co-author with their colleagues, who have the similar ability and research interest. We found no gender-sorting effect among Ontario economists.

Does co-authorship lead to higher research productivity, higher pay and a higher likelihood of promotion? The answer to this question matters to the policy maker. The results from this chapter do provide some suggestive evidence that co-authorship is associated with higher research productivity which has a market value. Further work needs to be done to address the potential endogeneity of teamwork formation. Ductor (2015) [34] notes that the choice of co-authorship and solo-authorship relies on the partnering opportunities available to researchers.

[34] Ductor (2015) uses the amount of co-authorship by the common research interest between an author and her potential co-authors as an instrument for teamwork formation.

2.7 Tables

Table 2.1: Studies related to academic economists in the US

Authors	Main RH variable	Data	Other RH variables	Method	Main finding
The determinants of co-authorship in economics					
Barnett et.al (1988)	Number of authors	all papers published in AER from 1960 to 1985	dummy for notes dummy for theoretical paper/empirical paper total number of submissions to AER number of people mentioned in the acknowledgment	OLS	the growth in co-authorship can be explained by the increasing of specialization, the rising opportunity cost of time, the growing incentive of avoid uncertainty of editorial review process.
Laband et.al (2000)	Co-authored or not	articles published in the A.E.R, J.P.E and Q.J.E during 1886-1995	JEL subject dummy gender Article length	Probit	the increasing incidence of co-authorship may resulted from greater quantitative content of papers
Co-authorship and pay					
Sauer (1988)	# of citations AEQ pages in each of top 100 journals(single authored/co-authored) $QPAGES = \sum_{j=1}^{100} p_j * w_j^{\alpha}$ pages of other papers # of books	140 academic economists 7 Econ departments 1982-1983	Years since PhD years of admin service department dummy	NLS	a 10-AEQ-pages paper in the top journals is associated with a 4% increase in salary, return to a co-authored paper with n people is $\frac{1}{n}$ times that of a single authored paper.
Hilmer et.al (2005)	total number of sole,co- and multi-authored articles	326 faculty members from top-ranked PhD granting programs	Years since PhD Gender Department dummy	OLS	higher quality papers have a higher impact on annual earnings and single authored articles have a higher return than multi-authored papers.
Co-authorship and the output of academic economists					
Hollis (2001)	output is measured by: $\sum_j \frac{p_j * q_j}{n_j}$, where p is number of AER length pages, q is quality index, n is number of authors. co-authorship is measured by the arithmetic mean of the number of authors for all papers published during the period .	339 members of the American Economic Association in 1981 whose surnames begin with a, b, c, d, s, t u or v	years since graduation school quality	OLS Tobit	more co-authorship is associated with higher quality of paper. However, after controlling for the number of authors, the net relation between collaboration and output is negative.
Medoff (2003)	number of citations an article receives	Every paper published during 1990 from 8 top economics journals	quality of the author subject area journal quality pages	Tobit maximum likelihood	co-authorship does not enhance research quality of economists
Borjas et.al (2012)	research productivity is measured by number of papers and total number of citations. co-authorship is measured by co-author status	data constructed from American Mathematical Society, Web of Science, Mathematics Genealogy Project.	research overlap index year field		collapse of Soviet union has a negative effect on research productivity of US mathematicians, co-authoring with a Soviet reduce this negative effect.

Table 1 Continued: Studies related to academic economists in the US

Authors	Main RH variable	Data	Other RH variables	Method	Main finding
Mobility and promotion patterns of academic economists					
Coupé	Number of publications	1000 top economics	Experience job rank	Probit	the probability of promotion and upward mobility is positively related to past publications.
Professional achievement and gender differences					
Maske et.al (2003)	total publication	a survey of members of American Economics Association	experience co-authorship rate gender institution type teaching loads	OLS OAXACA decomposition	co-authorship can increase the production of a paper. Males have around 7 more papers than females.
Pay and research productivity					
Hansen et.al (1978)	sum of published articles and books	the 1966 survey of economists undertaken by the National Register of Scientific and Technical Personnel 863 economists	job quality age degree quality gender experience experience squared	multi-equation Three stage LS	an additional unit of research productivity is associated with a 8% increase in annual salary.
Hamermesh et.al (1982)	Frequency of references # of books # of articles	148 full profs of economics 7 large public Universities 1979-1980	experience administrator dummy	OLS	an additional reference adds more to salary than an additional book or an article.
Moore et.al (2007)	# of level 1 articles # of level 2 articles # of other publications	181 faculty members 9 Econ department 1999-2000	experience experience squared seniority gender PhD quality Chair years rank	OLS	a level 1 publication is associated with 2.5% increase in salary.
Wang et.al (2013)	publications in top journals, top field journals, good general interest journals, and total publications in journals outside these top rankings	economics faculty at University of California	field university seniority seniority squared	OLS	a top-10 journal publication is associated with 1.5% increase in pay

Table 2.2: Canadian studies of research productivity of academic economists

Authors	Research productivity	Data	Other RH variables	Method	Main finding
Pay and research productivity					
Sen et.al (2014)	# of publications in peer-reviewed journals # of pubs on top 10 # of pubs on top 21 # of books # of total citations	543 tenure/tenure-track profs 16 universities 1996-2006 Panel data	SSHRC dummy experience experience squared teaching quality full prof dummy gender dummy for chair individual fixed effect year fixed effect	OLS Tobit	a top journal publication is associated with a 1% to 3% increase in annual salary.
Publication lags and research productivity					
Conley et.al (2012)	# of AER-equivalent Publications # of AER-equivalent pages	14271 PhD graduates between 1986 and 2000 in US and CA Economics departments. panel data	time polynomials dummy for grad year dummy for grad from top 30 department	Tobit	a downward trend in publication records.
Publication activity (and assessment of economics department)					
Lucas (1995)	pages of an article divide the number of authors(single-author-equivalent measure)	733 economists holding tenured or probationary appointment 1981-1990		Descriptive	on average, economists published one SAE article every 2.5 years during 1980s.
Davies et.al (2008)	# of publications in Top-10 journals # of publications in CJE	1980-2000		Descriptive	Canadian economists contributed to 5% of publications in the Top-10 journals and around half of publications in CJE, they also provided a ranking of Econ departments.
Simpson et.al (2012)	# of publications Canadian content of each article	258 economists with career starting between 1967 and 2010 panel data	stage of career country of PhD university type	Descriptive Logit	there is a declining interest in publishing a paper with Canadian content for new faculty hired since 1990.

Table 2.3: Data source and definition

Variable	Data source	Definition
Salary	Ministry of finance	salary
Co-authorship	EconLit	a co-authored publication is defined as a publication with at least two authors.
Type of co-authorship	EconLit	
1. inter-department collaboration		Number of co-authors from the same department on a paper excluding the focused economics himself
2. intra-departments collaboration		Number of co-authors from other Econ departments within Ontario
3. domestic collaboration		Number of co-authors from other province of Canada on a paper, including authors from other departments in the same University
4. US collaboration		Number of co-authors from US on a paper
5. international collaboration		Number of co-authors from other countries on a paper
Research productivity		
1. number of articles	EconLit	total number of articles published by a economist in previous year
2. total citation	Thomson Reuters Institute for Scientific Information-Web of Science archive and Google Scholar	total citation of each publication
Other individual characteristics		
Experience	PROQUEST Dissertations and Thesis database	number of years since an individual complete PhD
Male	Online CV	dummy for gender
Job rank	Online CV	dummy for assistant prof, associated prof, prof
US PhD	PROQUEST Dissertations and Thesis database	PhD is obtained from a US or non-US university
SSHRC	Social Sciences and information Humanities Research Council of Canada Award Search Engine	whether holding a SSHRC for a given year
Teaching performance	www.ratemyprofessor.com	average teacher ratings over all years
Other publication characteristics		
Subject	EconLit subject code	dummy for research area
Methodology		dummy for empirical/theoretical paper
Other university characteristics		
Type of university		dummy for medical/doctoral, comprehensive or primarily undergraduate institution
Merit pay and no salary cap		dummy for offer merit pay and no salary cap
Unionization		dummy for with/without faculty unionization

Table 2.4: Descriptive statistics of main variables

		Observations	Salary	Experience	Publications	Top 10	Top 21
All		2274	99,834	20.39	0.644	0.0484	0.065
			(19,249)	(10.46)	(1.039)	(0.236)	(0.27)
by individual characteristics							
Gender	male	2,416	100,890	21.39	0.658	0.0489	0.0658
			(19,614.67)	(10.37)	(1.063)	(0.239)	(0.273)
	female	328	92,014	12.95	0.530	0.0457	0.0579
			(14,033.74)	(7.80)	(0.838)	(0.209)	(0.246)
Cohort	grad before 1980	821	105,599	32.014	0.587	0.0146	0.0255
			(19,746)	4.91	(1.137)	(0.129)	(0.172)
	1980-1989	776	103,780	23.05	0.602	0.058	0.071
			(18,573)	(4.23)	(1.005)	(0.254)	(0.281)
	1990-1999	659	98,581	13.61	0.768	0.0714	0.0925
			(19,181)	(4.10)	(1.008)	(0.274)	(0.310)
	2000-2012	488	85,550	5.73	0.641	0.059	0.084
			(9,845)	(2.839)	(0.946)	(0.276)	(0.318)
by university characteristics							
Ranked or not	Ranked	1723	103,106	20.76	0.667	0.0656	0.0853
			(21,589)	(10.91)	(1.049)	(0.277)	(0.3109)
	Non-ranked	1021	94,312	19.76	0.606	0.0196	0.0304
			(12,686)	(9.63)	(1.023)	(0.139)	(0.177)
Type	Primarily Undergraduate	411	94,327	19.98	0.608	0.019	0.0292
			(13,539)	(9.24)	(1.13)	(0.138)	(0.182)
	Comprehensive	563	92,195	21.11	0.500	0.0196	0.0373
			(11,826)	(9.78)	(0.86)	(0.1386)	(0.189)
	Medical/Doctoral	1511	104,575	20.73	0.733	0.071	0.089
			(22,020)	(10.93)	(1.095)	(0.288)	(0.32)

Source: authors' own calculation.

Table 2.5: Total number of journal publications, by year and journal rank

	All			Co-authored		
Year	Total	Top 21	Top 10	Total	Top 21	Top 10
1996	234	34	24	159	28	20
1997	198	40	28	133	28	18
1998	205	30	17	143	22	13
1999	182	41	21	129	33	16
2000	143	21	14	95	16	11
2001	156	27	19	108	18	12
2002	187	35	17	138	27	13
2003	181	33	16	125	21	9
2004	174	22	10	122	15	7
2005	176	19	9	131	13	7
2006	203	31	15	145	27	12
2007	190	31	22	140	26	18
2008	209	31	18	151	26	18
2009	198	31	20	145	21	12
2010	246	32	21	191	24	15
2011	201	25	14	149	20	12
2012	257	36	22	206	32	20

Source: authors' own calculation.

Table 2.6: Summary statistics for citation counts (2005 - 2013)

Variables	Mean	Std. Dev	Min	Max
Citation counts from Google Scholar	52.83	119.04	0	2,137
Citation counts from WOS	14.72	28.86	0	563
Journal Impact Factor	1.87	1.39	0.152	9.67

Source: authors' own calculation.

Table 2.7: Number of citations, by number of authors

	Total number of citations	Average number of citations
single work	34,964	40
two authors	83,613	54
three authors	39,361	68
more than 3 authors	5,777	53

Source: authors' own calculation.

Table 2.8: Definition of types of collaboration

Type of collaboration	Definition
inter-department collaboration	Number of co-authors from the same department on a paper excluding the focus economist herself
intra-departments collaboration	Number of co-authors from other Econ departments within Ontario
domestic collaboration	Number of co-authors from other province of Canada on a paper, including authors from other departments in the same University
US collaboration	Number of co-authors from US on a paper
international collaboration	Number of co-authors from other countries on a paper

Table 2.9: Other university characteristics used

	Type of university	Merit pay and no salary cap	Unionization
Brock University	Primary Undergraduate	No	before 1996
Carleton University	Comprehensive	No	before 1996
University of Guelph	Comprehensive	Yes	2006
Lakehead University	Primary Undergraduate	No	before 1996
Laurentian University	Primary Undergraduate	Yes	before 1996
McMaster University	Medical/Doctoral	Yes	No
University of Ottawa	Medical/Doctoral	No	before 1996
Queen's University	Medical/Doctoral	Yes	before 1996
Ryerson University	Primary Undergraduate	No	before 1996
University of Toronto	Medical/Doctoral	Yes	No
Trent University	Primary Undergraduate	No	before 1996
University of Waterloo	Comprehensive	Yes	No
Western University	Medical/Doctoral	Yes	1998
Wilfrid Laurier University	Primary Undergraduate	No	before 1996
University of Windsor	Comprehensive	No	before 1996
York University	Comprehensive	No	before 1996

Source: Sen et al. (2014).

Table 2.10: Distribution of number of authors

N	Total number	Percentage
1	930	0.28
2	1642	0.49
3	619	0.18
4	117	0.04
> 4	32	0.01
total	3340	1

Table 2.11: Number of authors, by year

	Number				Percentage			
Year	Single	Two authors	Three authors	> 4 authors	Single	Two authors	Three authors	> 4 authors
1996	75	115	41	3	32.05%	49.15%	17.52%	1.28%
1997	65	99	26	8	32.83%	50.00%	13.13%	4.04%
1998	62	105	33	4	30.39%	51.47%	16.18%	1.96%
1999	53	98	26	5	29.12%	53.85%	14.29%	2.75%
2000	48	67	23	5	33.57%	46.85%	16.08%	3.50%
2001	48	82	20	6	30.77%	52.56%	12.82%	3.85%
2002	49	98	28	12	26.20%	52.41%	14.97%	6.42%
2003	56	81	41	3	30.94%	44.75%	22.65%	1.66%
2004	52	71	41	10	29.89%	40.80%	23.56%	5.75%
2005	45	79	40	12	25.57%	44.89%	22.73%	6.82%
2006	58	104	3	4	34.32%	61.54%	1.78%	2.37%
2007	50	91	37	12	26.32%	47.89%	19.47%	6.32%
2008	58	106	35	10	27.75%	50.72%	16.75%	4.78%
2009	54	98	36	11	27.14%	49.25%	18.09%	5.53%
2010	54	130	56	5	22.04%	53.06%	22.86%	2.04%
2011	52	92	42	15	25.87%	45.77%	20.90%	7.46%
2012	51	126	57	23	19.84%	49.03%	22.18%	8.95%

Table 2.12: Distribution of different types of co-authored publications, by quality

	Co-authored-Same University		Co-authored-US University	
	Total Number	Percentage	Total Number	Percentage
Top 10	48	0.21	82	0.35
Top 21	92	0.23	121	0.30

Table 2.13: Percentage of different types of publication, by gender

Type	All	Female	Male
same university	39.38%	41.78%	39.09%
other Canada university	23.57%	24.44%	23.46%
US university	16.54%	19.11%	16.22%
international university	20.51%	14.67%	21.23%

Table 2.14: OLS estimates of the effects of co-authored publications in the "TOP-10", "TOP-21" and non-top ranked journals

	publications in top 10 journals		publications in top 21 journals		publications in other journals	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of co-authored publication	-0.1369	-0.1400	0.7745**	0.6616***	0.1809*	0.2372**
	(0.1723)	(0.1681)	(0.3650)	(0.3419)	(0.1230)	(0.1183)
Individual characteristics						
Male	0.0117	-0.0282	-0.3263	-0.7046	0.2778	0.2066
	(0.2961)	(0.3193)	(0.5445)	(0.4720)	(0.2635)	(0.2862)
Associate professor	-0.2593	-0.1837	-0.0819	0.0086	-0.5429	−0.5768*
	(0.3339)	(0.3233)	(0.2969)	(0.3032)	(0.3656)	(0.3511)
Full professor	-0.0643	-0.0695	0.2146	0.1113	-0.1350	-0.0590
	(0.3916)	(0.3849)	(0.3423)	(0.3531)	(0.5447)	(0.5602)
Experience	0.0934*	0.0807	0.0289	0.0318	0.0106	0.0336
	(0.0568)	(0.0597)	(0.0420)	(0.0415)	(0.0773)	(0.0755)
Experience squared	−0.0027*	-0.0024	-0.00064	-0.00057	-0.0011	-0.0016
	(0.0014)	(0.0014)	(0.0010)	(0.00095)	(0.0019)	(0.0018)
PhD from US	0.1659	0.1422	-0.1461	-0.1249	-1.1488	-1.0665
	(0.2106)	(0.2274)	(0.2309)	(0.2325)	(0.7658)	(0.7701)
PhD from Canada	0.3395	0.3540	0.4978	0.3373	0.3209	0.3741
	(0.2305)	(0.2689)	(0.3648)	(0.3923)	(0.3067)	(0.3724)
Average rating on teaching	−0.1103***	−0.1210***	-0.0514	-0.0436	−0.1941*	−0.1506*
	(0.0539)	(0.0590)	(0.0915)	(0.0896)	(0.1197)	(0.0941)
SSHRC	0.1438	0.2041	0.3376	0.4003***	0.4061*	0.4319*
	(0.1380)	(0.1505)	(0.2240)	(0.2082)	(0.2570)	(0.2479)
University characteristics						
Medical schools	0.6664***	0.9457***	0.0156	0.0183	0.0279	0.0041
	(0.2913)	(0.2586)	(0.3185)	(0.4365)	(0.5050)	(0.9543)
Comprehensive universities	−0.9166***	−0.9100***	-0.5479	-0.5470	-0.1208	-0.1258
	(0.3363)	(0.3360)	(0.3970)	(0.3970)	(0.5028)	(0.8128)
Union	-0.0884		0.1902		-0.0448	
	(0.2301)		(0.2596)		(0.3757)	
Nocap	0.6198***		0.4094		-0.6062	
	(0.2770)		(0.3385)		(0.5071)	
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes	No	Yes
R-squared	0.3415	0.3644	0.2327	0.3925	0.1127	0.1500
Observations	159	159	305	305	1,517	1,517

Note: Dependent variable in all columns is the total number of corresponding publications aggregated over 3 year period. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.15: OLS estimates of the effects of co-authored publications relatives to solo-publications

	(1)	(2)	(3)	(4)
Single-authored publication	0.6087***	0.6199***	0.5886***	0.5774***
	(0.0535)	(0.0500)	(0.0413)	(0.0419)
Co-authored publication	0.7497***	0.7645***	0.7732***	0.7864***
	(0.0642)	(0.0631)	(0.0570)	(0.0577)
Individual characteristics				
Male	-0.1021	-0.1552	−0.2187*	-0.1619
	(0.1330)	(0.1290)	(0.1231)	(0.1281)
Associate professor	0.0270	0.0386	0.0573	0.0491
	(0.1080)	(0.1040)	(0.0974)	(0.0991)
Full professor	0.1265	0.0355	0.0191	0.0778
	(0.1451)	(0.1407)	(0.1315)	(0.1306)
Experience	−0.0320*	−0.0296*	-0.0224	−0.0274*
	(0.0175)	(0.0163)	(0.0148)	(0.01537)
Experience squared	-0.00016	-0.00013	-0.00031	-0.00019
	(0.00041)	(0.00039)	(0.00034)	(0.00035)
PhD from US	-0.1257	-0.1377	-0.1483	-0.0783
	(0.1737)	(0.1715)	(0.1564)	(0.1576)
PhD from Canada	−0.5545***	−0.5478***	-0.2688	-0.0934
	(0.1727)	(0.1713)	(0.1651)	(0.1769)
Average rating on teaching	−0.0666*	-0.0479	-0.0298	-0.0253
	(0.0348)	(0.0343)	(0.0298)	(0.0292)
SSHRC			0.4120***	0.3588***
			(0.0893)	(0.0840)
University characteristics				
Medical schools			0.3815***	0.3155
			(0.1520)	(0.2858)
Comprehensive universities			0.0031	-0.1116
			(0.1347)	(0.3451)
Union			-0.1631	
			(0.1249)	
Nocap			0.2403	
			(0.1471)	
Year fixed effect	No	Yes	Yes	Yes
University fixed effect	No	No	No	Yes
R-squared	0.2148	0.2512	0.3077	0.3319
Observations	1,962	1,962	1,962	1,962

Note: Dependent variable in all columns is the natural log of the total number of citations adjusted by year. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.16: OLS estimates of the effects of co-authored publications relatives to solo-publications in the "TOP-10" and "TOP-21" ranked journals

	publications in top 10 journals		publications in top 21 journals	
	(1)	(2)	(3)	(4)
Top journals				
Single-authored publication	1.3328***	1.2740***	1.0581***	0.9948***
	(0.1543)	(0.1503)	(0.1235)	(0.1187)
Co-authored publication	1.3944***	1.3324***	1.5051***	1.4536***
	(0.1132)	(0.1095)	(0.1083)	(0.1052)
Non-top journals				
Single-authored publication	0.5475***	0.5415***	0.6974***	0.5244***
	(0.0417)	(0.0419)	(0.0554)	(0.0396)
Co-authored publication	0.7393***	0.7501***	0.5308***	0.7059***
	(0.0561)	(0.0576)	(0.0400)	(0.0567)
Individual characteristics				
Male	−0.2462**	-0.1909	-0.2365	-0.1833
	(0.1231)	(0.1287)	(0.1218)	(0.1269)
Associate professor	0.0572	0.0480	0.0382	0.0286
	(0.0978)	(0.0989)	(0.0969)	(0.0973)
Full professor	0.0061	0.0674	-0.0103	0.0484
	(0.1296)	(0.1299)	(0.1271)	(0.1265)
Experience	−0.0249*	−0.0292***	−0.0247*	−0.0288*
	(0.0146)	(0.0151)	(0.01440)	(0.0148)
Experience squared	-0.00015	-0.000059	-0.00011	-0.000018
	(0.00034)	(0.00034)	(0.00033)	(0.00034)
PhD from US	-0.1513	-0.0835	-0.1440	-0.0750
	(0.1567)	(0.1594)	(0.1578)	(0.1599)
PhD from Canada	-0.2635	-0.1167	-0.2422	-0.1066
	(0.1641)	(0.1781)	(0.1647)	(0.1782)
Average rating on teaching	-0.01430	-0.0090	-0.0135	-0.0082
	(0.0296)	(0.0295)	(0.0292)	(0.0291)
SSHRC	0.3848***	0.3527***	0.3536***	0.3304***
	(0.0882)	(0.0833)	(0.0873)	(0.0829)
University characteristics				
Medical schools	0.3802***	0.3492	0.3642***	0.3413
	(0.1502)	(0.2597)	(0.1497)	(0.2607)
Comprehensive universities	0.0361	-0.1636	0.0370	-0.1416
	(0.1326)	(0.3402)	(0.1311)	(0.3471)
Union	-0.1306		-0.1453	
	(0.1235)		(0.1207)	
Nocap	0.1980		0.1517	
	(0.1422)		(0.1392)	
Year fixed effect	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes
R-squared	0.3295	0.3491	0.3407	0.3590
Observations	1,962	1,962	1,962	1,962

Note: Dependent variable in all columns is the natural log of the total number of citations adjusted by year. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.17: OLS estimates of different types of co-authorship based on publication level data (1996-2012)

	(1)	(2)	(3)	(4)
Inter-department collaboration	-0.0559	0.00047	0.0135	-0.0765
	(0.0774)	(0.0697)	(0.0655)	(0.0674)
Intra-departments collaboration	-0.0536	0.0839	0.1299	0.0550
	(0.1059)	(0.1053)	(0.1008)	(0.1171)
Domestic collaboration	0.0919	0.1197^{**}	0.1098^{*}	0.1069^{*}
	(0.066))	(0.0595)	(0.0599)	(0.0668)
US collaboration	0.2939^{***}	0.3251^{***}	0.3336^{***}	0.3194^{***}
	(0.1292)	(0.109)	(0.1044)	(0.089)
international collaboration	0.1998^{***}	0.2505^{***}	0.2386^{***}	0.1178^{*}
	(0.0942)	(0.076)	(0.0792)	(0.0791)
Individual characteristics				
SSHRC dummy	0.558^{***}	0.543^{***}	0.499^{***}	
	(0.131)	(0.130)	(0.1075)	
Male	0.0139	-0.0728	-0.0151	
	(0.125)	(0.1092)	(0.1202)	
Associate professor	0.174	0.092	0.137	
	(0.121)	(0.101)	(0.1002)	
Full professor	0.299^{***}	0.037	0.173^{*}	
	(0.134)	(0.113)	(0.102)	
Experience	-0.0463^{***}	-0.0196^{*}	-0.0311^{***}	-0.1032^{***}
	(0.0154)	(0.0118)	(0.0114)	(0.0176)
Experience squared	0.00032	-0.00015	0.000085	0.0005
	(0.00038)	(0.00027)	(0.00026)	(0.00035)
PhD from US	-0.0857	0.002	0.0481	
	(0.150)	(0.126)	(0.136)	
PhD from Canada	-0.520^{***}	-0.144	-0.0117	
	(0.145)	(0.131)	(0.146)	
Average rating on teaching	-0.0342	-0.0016	-0.0067	
	(0.0368)	(0.0271)	(0.0261)	
University characteristics				
Medical schools				1.248^{***}
				(0.614)
Comprehensive universities				0.803
				(0.606)
Year fixed effect	No	Yes	Yes	Yes
University fixed effect	No	No	Yes	No
Individual fixed effect	No	No	No	Yes
R-squared	0.097	0.212	0.249	0.446
Observations	2005	2005	2005	2005

Note: Research productivity is measured by citation counts from Google Scholar. Dependent variable in all columns is the natural log of citations. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.18: OLS estimates of different type of co-authorship based on publication level data (1996-2012)

	Journal Impact Factor				citation counts from WOS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inter-department collaboration	0.0201	-0.0128	0.002892	-0.0448	-0.0224	0.0261	0.0460	0.0132
	(0.059)	(0.056)	(0.0510)	(0.0574)	(0.0527)	(0.0504)	(0.0482)	(0.0570)
Intra-departments collaboration	0.0085	0.094	0.1437	0.1627	0.0093	0.0625	0.1215	0.1392
	(0.111)	(0.107)	(0.1041)	(0.1079)	(0.1090)	(0.0986)	(0.0944)	(0.108)
Domestic collaboration	0.065	0.033	0.0086	0.0706	0.0339	0.0658	0.0543	0.1110*
	(0.054)	(0.0506)	(0.0506)	(0.0651)	(0.0615)	(0.0599)	(0.0562)	(0.0662)
US collaboration	0.2117***	0.1846***	0.1922***	0.1820***	0.1964***	0.2361***	0.2621***	0.2689***
	(0.0673)	(0.069)	(0.0679)	(0.0677)	(0.0850)	(0.0732)	(0.0686)	(0.0689)
International collaboration	0.1537**	0.1425**	0.1166*	0.1211*	0.1670***	0.1989***	0.1940***	0.1281**
	(0.074)	(0.0698)	(0.0678)	(0.0755)	(0.0747)	(0.0689)	(0.0711)	(0.0656)
Year fixed effect	No	Yes	Yes	Yes	No	Yes	Yes	Yes
University fixed effect	No	No	Yes	No	No	No	Yes	No
Individual fixed effect	No	No	No	Yes	No	No	No	Yes
Observations	2323	2323	2323	2323	2323	2323	2323	2323
R-squared	0.08	0.15	0.20	0.41	0.036	0.103	0.138	0.400
	Publications in top 10 journals				Publications in top 21 journals			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inter-department collaboration	-0.0460	0.3470	0.3470	0.2802	0.0510	0.1476	0.1034	0.2076
	(0.2174)	(0.2111)	(0.2111)	(0.2880)	(0.2067)	(0.2057)	(0.2140)	(0.3016)
Intra-departments collaboration	0.2671	0.4393	0.4393	0.5912	0.1668	0.3235	0.3613	0.6700
	(0.2786)	(0.2731)	(0.2731)	(0.3815)	(0.2727)	(0.2404)	(0.2562)	(0.4219)
Domestic collaboration	0.4275*	0.2608	0.2608	0.1028	0.3863**	0.2302	0.2410	0.0403
	(0.2255)	(0.2586)	(0.2586)	(0.2042)	(0.1970)	(0.2014)	(0.2134)	(0.2086)
US collaboration	-0.2518	-0.1993	-0.1993	-0.0356	-0.1139	-0.0738	-0.0846	0.0505
	(0.2999)	(0.2673)	(0.2673)	(0.3720)	(0.2977)	(0.2707)	(0.2821)	(0.3997)
International collaboration	0.9811***	1.4437***	1.4437***	1.4770***	0.5379*	0.6121*	0.6501*	1.4027**
	(0.3447)	(0.2835)	(0.2835)	(0.3309)	(0.3520)	(0.3950)	(0.4220)	(0.3477)
Year fixed effect	No	Yes	Yes	Yes	No	Yes	Yes	Yes
University fixed effect	No	No	Yes	No	No	No	Yes	No
Individual fixed effect	No	No	No	Yes	No	No	No	Yes
Observations	193	193	193	193	252	252	252	252
R-squared	0.1771	0.3637	0.3637	0.5920	0.0988	0.2835	0.2960	0.6891

Note: Research productivity is measured by citation counts and Journal Impact Factor from WOS in the lower panel. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.19: Empirical studies on collaboration and research productivity

authors	country	research productivity	sample	results
Adam et.al (2005)	US	1.number of citation 2.number of papers	2.4 million scientific papers written in 110 top U.S. universities over 1981 1999	Collaboration decrease research productivity, increases research research quality .
He et al. (2009)	New Zealand	1.citation counts a paper received in a 2-year window 2. Journal impact factor	65 bio-medical scientists	their coefficient estimates of international collaboration domestic collaboration and within-university collaboration are around 15%, 1.6% and 19%.
Tang (2013)	China	1.total citation 2.Journal impact factor	77 Chinese nanoscientist	collaboration across national boundaries has a positive effect on Chinas nano research quality

Table 2.20: OLS estimates of returns to co-authorship based on individual level data

	all data					restricted data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Single-authored publication	0.0073	0.0092	0.0019	0.0047	0.0064	0.0128	0.0165
	(0.0111)	(0.0116)	(0.0084)	(0.0073)	(0.0083)	(0.0124)	(0.0140)
Co-authored publication	0.0219***	0.0197***	0.0098**	0.0109***	0.0133***	0.0164***	0.0184***
	(0.0046)	(0.0047)	(0.0044)	(0.0041)	(0.0042)	(0.0066)	(0.0067)
Individual characteristics							
Male	0.0167	0.0250	0.0194	0.0279	0.0279	0.0340	0.0340
	(0.0199)	(0.0216)	(0.0194)	(0.0175)	(0.0179)	(0.0176)	(0.0184)
Associate professor	0.020	0.0139	0.022	0.0168	0.0108	0.0219	0.0225
	(0.0236)	(0.0247)	(0.0205)	(0.0196)	(0.0209)	(0.0215)	(0.0225)
Full professor	0.0619**	0.0679**	0.0568**	0.0518**	0.0469*	0.0518*	0.0524*
	(0.0325)	(0.0343)	(0.0280)	(0.0263)	(0.0278)	(0.0297)	(0.0308)
Experience	0.0117***	0.0139***	0.0167***	0.0200***	0.0210***	0.0217***	0.0223***
	(0.003)	(0.0031)	(0.0028)	(0.00268)	(0.0028)	(0.0031)	(0.0032)
Experience squared	−0.00016***	−0.0002***	−0.00024***	−0.0003***	−0.00033***	−0.00034***	−0.00036***
	(0.00007)	(0.00007)	(0.00007)	(0.00006)	(0.00006)	(0.00007)	(0.00007)
PhD from US	-0.015	-0.0149	-0.0219	-0.0183	-0.0175	-0.0198	-0.0214
	(0.0307)	(0.0319)	(0.026)	(0.0249)	(0.0275)	(0.0268)	(0.0290)
PhD from Canada	−0.064**	−0.0702***	-0.0386	-0.0063	-0.0016	-0.0123	-0.0105
	(0.0306)	(0.032)	0.0269	(0.0261)	(0.0287)	(0.0273)	(0.0293)
Average rating on teaching	0.0061	0.005	0.0041	0.0028	0.0029	0.0015	0.0016
	(0.0049)	(0.0052)	(0.0047)	(0.0046)	(0.0049)	(0.0044)	(0.0048)
SSHRC			0.084***	0.0778***	0.0793***	0.0573***	0.0599***
			(0.0148)	(0.0133)	(0.0144)	(0.0152)	(0.0165)
University characteristics							
Medical schools			0.0337*	0.023	0.0944	0.030	0.0300
			(0.182)	(0.0195)	(0.0681)	(0.0318)	(0.0661)
Comprehensive universities			-0.0230	-0.0125	-0.0512	−0.0830*	0.0083
			(0.0159)	(0.0308)	(0.0419)	(0.0452)	(0.0604)
Union			−0.0631***		0.0871**		-0.0051
			(0.0191)		(0.0390)		(0.0398)
Nocap			-0.0036		0.2952***		0.1013***
			(0.0162)		(0.0838)		(0.0434)
Year fixed effect	No	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	No	No	Yes	Yes	Yes	Yes
School specific time trend	No	No	No	No	Yes	No	Yes
R-squared	0.25	0.29	0.41	0.49	0.52	0.52	0.54
Observations	2717	2717	2717	2717	2717	789	789

Note: Dependent variable in all columns is the natural log of salary. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. Column 6 and 7 replicate column 4 and 5 but with a smaller number of years: 1996, 1999, 2003, 2008 and 2012. "TOP-21" ranked journal including Canadian Journal of Economics.

Table 2.21: OLS estimates of returns to co-authored publications in the top ranked journals based on individual level data (1996-2012) and year-university fixed effects

	all data				restricted data			
	publications in top 10 journals		publications in top 21 journals		publications in top 10 journals		publications in top 21 journals	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Top journals								
Single-authored publication	0.0678**	0.0568*	0.0430	0.0302	0.0973***	0.0772**	0.0197	0.0209
	(0.0330)	(0.0326)	(0.0282)	(0.0245)	(0.0418)	(0.0378)	(0.0315)	(0.0198)
Co-authored publication	0.0546***	0.0546***	0.0296**	0.0213*	0.0878***	0.0849***	0.0393***	0.0320*
	(0.0170)	(0.0148)	(0.0128)	(0.0117)	(0.0238)	(0.0202)	(0.0184)	(0.0173)
Non-top journals								
Single-authored publication	0.0020	0.0047	-0.0047	-0.00025	0.0093	0.0125	-0.00021	0.0054
	(0.0084)	(0.0073)	(0.0071)	(0.0062)	(0.0130)	(0.0120)	(0.0127)	(0.0121)
Co-authored publication	0.0095**	0.0105***	0.0058	0.0091**	0.0151***	0.0158***	0.0127*	0.0144**
	(0.0044)	(0.0040)	(0.0050)	(0.0043)	(0.0072)	(0.0068)	(0.0079)	(0.0072)
Individual characteristics								
Male	0.0189	0.0278	0.0228	0.0275	0.0252	0.0314*	0.0261	0.0326*
	(0.0195)	(0.0175)	(0.0184)	(0.0177)	(0.0175)	(0.0174)	(0.0177)	(0.0175)
Associate professor	0.0223	0.0168	0.0172	0.0163	0.0261	0.0211	0.0264	0.0215
	(0.0206)	(0.0196)	(0.0205)	(0.0197)	(0.0226)	(0.0216)	(0.0224)	(0.0215)
Full professor	0.0565***	0.0519**	0.0443*	0.0508*	0.0498	0.0474*	0.0501*	0.0489*
	(0.0280)	(0.0263)	(0.0280)	(0.0264)	(0.0318)	(0.0299)	(0.0312)	(0.0294)
Experience	0.0166***	0.0199***	0.0184***	0.0202***	0.0199***	0.0221***	0.0196***	0.0218***
	(0.0028)	(0.0026)	(0.0027)	(0.0026)	(0.0032)	(0.0031)	(0.00318)	(0.0031)
Experience squared	−0.00024***	−0.0003***	−0.00027***	−0.00031***	−0.0003***	−0.00035***	−0.00029***	−0.00034***
	(0.00006)	(0.00006)	(0.00006)	(0.00006)	(0.00007)	(0.00007)	(0.00007)	(0.00007)
PhD from US	-0.0220	-0.0183	-0.0260	-0.0254	-0.0215	-0.0188	-0.0235	-0.0205
	(0.0265)	(0.0249)	(0.0265)	(0.0250)	(0.0276)	(0.0268)	(0.0272)	(0.0266)
PhD from Canada	-0.0385	-0.00651	-0.0320	-0.0067	−0.0473*	-0.0124	−0.0478*	-0.0135
	(0.0268)	(0.0261)	(0.0259)	(0.0263)	(0.0275)	(0.0270)	(0.0271)	(0.0269)
Average rating on teaching	0.0044	0.0029	0.0035	0.0032	0.0017	0.00098	0.0025	0.0016
	(0.0046)	(0.0045)	(0.0045)	(0.00454)	(0.0046)	(0.0044)	(0.0045)	(0.0044)
SSHRC	0.0818***	0.0772***	0.0835***	0.0766***	0.0624***	0.0571***	0.0580***	0.0544***
	(0.0146)	(0.0132)	(0.0144)	(0.0131)	(0.0163)	(0.0150)	(0.0159)	(0.0148)
University characteristics								
Medical schools	0.0332*	0.0235	0.0384	0.0256	0.0306	0.382	0.309	0.303
	(0.0181)	(0.0193)	(0.0171)	(0.0177)	(0.0178)	(0.0176)	(0.0174)	(0.0175)
Comprehensive universities	-0.0234	-0.0128	-0.0276	-0.0116	−0.1225***	−0.0848*	−0.1233***	-0.0815
	(0.0158)	(0.0310)	(0.0149)	(0.0385)	(0.0161)	(0.0453)	(0.0159)	(0.0451)
Union	−0.0620***		−0.0532***		−0.0610***		−0.0586***	
	(0.0190)		(0.0199)		(0.0189)		(0.0184)	
Nocap	-0.004		0.0204		0.0630***		0.0640***	
	(0.0162)		(0.0196)		(0.0159)		(0.0157)	
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.41	0.49	0.43	0.49	0.46	0.52	0.46	0.52
Observations	2717	2717	2717	2717	789	789	789	789

Note: Dependent variable in all regressions is the natural log of salary. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. Column 5 to 8 replicate column 1 to 4 but with a smaller number of years: 1996, 1999, 2003, 2008 and 2012. “TOP-21” ranked journal including Canadian Journal of Economics.

Table 2.22: OLS estimates based on individual level data (1996-2012) and with year-university fixed effects, using number of publications as measure for research productivity

	publications in top 10 journals		publications in top 21 journals		all publications		AER	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Publications (Top journals)	0.0319***	0.016	0.0307***	0.017	0.008***	0.008***	0.0033***	0.0023***
	(0.014)	(0.014)	(0.013)	(0.012)	(0.003)	(0.003)	(0.0012)	(0.0011)
Non-top journals	0.0063***	0.0077***	0.006***	0.0075***				
	(0.003)	(0.003)	(0.003)	(0.0027)				
Individual characteristics								
Experience	0.008***	0.008***	0.0078***	0.008***	0.0079***	0.008***	0.0079***	0.008***
	(0.0016)	(0.0015)	(0.0016)	(0.0015)	(0.0016)	(0.0015)	(0.0016)	(0.0015)
Experience squared	-0.00006	−0.00008***	-0.00006	−0.00008***	-0.00006	−0.00008***	-0.00006	−0.00008***
	(0.00004)	(0.000038)	(0.00004)	(0.000038)	(0.00004)	(0.000039)	(0.00004)	(0.00004)
Male	0.015	0.018	0.015	0.018	0.0158	0.019	0.015	0.019
	(0.01)	(0.0098)	(0.01)	(0.0099)	(0.01)	(0.01)	(0.01)	(0.01)
associate professor	0.02	0.024***	0.021	0.024***	0.021	0.024***	0.021	0.024***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Full professor	0.048***	0.054***	0.048***	0.054***	0.048***	0.054***	0.048***	0.055***
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)	(0.012)	(0.017)	(0.016)
Average rating on teaching	0.007***	0.0057***	0.007***	0.0057***	0.0068***	0.0056***	0.007***	0.0058***
	(0.0028)	(0.0027)	(0.0028)	(0.0027)	(0.0028)	(0.0028)	(0.0028)	(0.0027)
SSHRC dummy	0.075***	0.065***	0.075***	0.065***	0.075***	0.065***	0.075***	0.067***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
University characteristics								
Medical schools	0.0017		0.0018		0.0018		0.0013	
	(0.014)		(0.0137)		(0.014)		(0.014)	
Comprehensive universities	−0.055***		−0.055***		−0.055***		−0.055***	
	(0.012)		(0.012)		(0.012)		(0.012)	
Union	−0.031***	-0.021	−0.0306***	-0.021	−0.031***	-0.021	−0.031***	-0.022
	(0.013)	(0.017)	(0.013)	(0.017)	(0.013)	(0.017)	(0.013)	(0.017)
Merit/no salary cap dummy	0.01		0.01		0.011		0.01	
	(0.01)		(0.01)		(0.01)		(0.01)	
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.438	0.494	0.438	0.494	0.495	0.437	0.494	0.495
Observations	2717	2717	2717	2717	2717	2717	2717	2717

Note: Dependent variable is the natural log of salary. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.23: OLS estimates based on individual level data (1996-2012), using citation counts as measure for research productivity

	without co-authorship adjustment		with co-authorship adjustment	
	(1)	(2)	(3)	(4)
Citations($ln(citation+1)$)	0.008***	0.0074***	0.0098***	0.0088***
	(0.0022)	(0.0020)	(0.0025)	(0.0023)
Individual characteristics				
experience	0.0184***	0.0202***	0.0184***	0.0202***
	(0.0027)	(0.0026)	(0.0027)	(0.0026)
experience squared	−0.00026***	−0.00031***	−0.00026***	−0.00031***
	(0.00006)	(0.00006)	(0.00006)	(0.00006)
Male	0.0223	0.0272	0.0220	0.0270
	(0.0182)	(0.0175)	(0.0182)	(0.0175)
Average rating on teaching	0.0034	0.0031	0.0034	0.0031
	(0.0046)	(0.0045)	(0.0046)	(0.0045)
SSHRC dummy	0.0812***	0.0742***	0.0809***	0.0741***
	(0.0143)	(0.0129)	(0.0142)	(0.0128)
PhD from US	-0.0185	-0.0184	-0.0184	-0.0185
	(0.0261)	(0.0252)	(0.0260)	(0.0252)
PhD from Canada	-0.0293	-0.0051	-0.0290	-0.0050
	(0.0261)	(0.0263)	(0.0261)	(0.0263)
Associate professor	0.0155	0.0148	0.0155	0.0148
	(0.0203)	(0.0196)	(0.0203)	(0.0196)
Full professor	0.0431	0.0500*	0.0431	0.0501***
	(0.0277)	(0.0262)	(0.0277)	(0.0261)
University characteristics				
Comprehensive universities	−0.0514***		−0.574***	
	(0.0150)		(0.0150)	
Medical schools	-0.0116		-0.0115	
	(0.0175)		(0.0175)	
Union	−0.0526***	−0.0459***	−0.0523***	−0.0457***
	(0.0199)	(0.0160)	(0.0198)	(0.0159)
Year fixed effect	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes
R-squared	0.4153	0.4729	0.4173	0.4735
Observations	2717	2717	2717	2717

Note: Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.24: weighted average of publications before and after promotion

	mean	standard error	min	max
All				
weighted average of publications from year t-3 to t-1	0.75	0.58	0	2.67
weighted average of publications from year t-5 to t-1	0.66	0.47	0	2.2
weighted average of publications from year t+3 to t+1	0.67	0.64	0	2.67
weighted average of publications from year t+5 to t+1	0.64	0.48	0	1.8
from asst to assoc				
weighted average of publications from year t-3 to t-1	0.76	0.53	0	2
weighted average of publications from year t-5 to t-1	0.62	0.43	0	2
weighted average of publications from year t+3 to t+1	0.67	0.67	0	2.67
weighted average of publications from year t+5 to t+1	0.60	0.47	0	1.8
from assoc to prof				
weighted average of publications from year t-3 to t-1	0.74	0.69	0	2.67
weighted average of publications from year t-5 to t-1	0.77	0.53	0	2.2
weighted average of publications from year t+3 to t+1	0.68	0.59	0	2
weighted average of publications from year t+5 to t+1	0.723	0.51	0	1.6

Note: where t is the year of promotion.

Table 2.25: Marginal effects from Probit estimates of effect of past performance on promotion

	All		Assis to Assoc		Assoc to Full	
	(1)	(2)	(3)	(4)	(5)	(6)
Past performance (PP_{iut-1})	0.0492***	0.0452***	0.0634***	0.0348***	0.0145	0.0038
	(0.017)	(0.0134)	(0.0164)	(0.0116)	(0.010)	(0.0073)
$EXP_{iut} * PP_{iut-1}$	−0.0042***	−0.0044***	−0.00497***	−0.0031**	-0.00015	0.0006
	(0.0015)	(0.0014)	(0.0017)	(0.0014)	(0.0006)	(0.00054)
EXP_{iut}	0.0076***	-0.00214	0.0071**	−0.00039***	0.0056***	0.0041***
	(0.0029)	(0.0021)	(0.0033)	(0.0001)	(0.00197)	(0.00145)
EXP^2_{iut}	0.00009	-0.00003	−0.0003***	−0.00039***	−0.00012***	−0.00009***
	(0.00008)	(0.00006)	(0.00012)	(0.0001)	(0.00005)	(0.000036)
$ASST_{iut}$	0.144***	0.119***				
	(0.0215)	(0.0164)				
$ASSO_{iut}$	0.228***	0.193***				
	(0.026)	(0.0208)				
Gender	-0.0012	0.0028	0.00036	0.00402	-0.0076	-0.0054
	(0.011)	(0.0088)	(0.0098)	(0.0080)	(0.0061)	(0.0047)
Observations	2,136	2,805	1,180	1,303	301	363
Log Likelihood	-341.988	-382.662	-308.688	-354.02153	-148.82615	-155.16884
Pseudo R^2	0.2001	0.2027	0.078	0.062	0.081	0.113

Note: The dependent variable is the likelihood of promotion and takes the value one or zero. All specifications are estimated using a probit model and the marginal effects are presented. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.26: Marginal effects from Probit estimates of effect of past co-authored publications on promotion

	All		Assis to Assoc		Assoc to Full	
	(1)	(2)	(3)	(4)	(5)	(6)
EXP_{iut}	-0.0057^{**}	-0.0056^{**}	0.0072^{**}	0.00711^{**}	0.0091^{***}	0.0084^{***}
	(0.0029)	(0.0029)	(0.00339)	(0.0033)	(0.0029)	(0.0027)
EXP_{iut}^2	0.00008	0.000079	-0.00029^{**}	-0.00029^{**}	-0.00016^{***}	-0.00015^{***}
	(0.000078)	(0.000078)	(0.00012)	(0.00012)	(0.00006)	(0.00006)
Number of co-authored paper	0.0882^{***}	0.0741^{**}	0.0808^{***}	0.0826^{***}	0.0312	0.0077
	(0.0246)	(0.0373)	(0.0220)	(0.0331)	(0.0211)	(0.0298)
Number of solo-authored paper	0.0634^{***}	0.0864^{***}	0.0791^{***}	0.1065^{***}	0.0215	-0.0794
	(0.0192)	(0.0322)	(0.0182)	(0.0289)	(0.0173)	(0.0572)
Number of co paper * male		0.0201		-0.00039		0.0217
		(0.0357)		(0.0312)		(0.0281)
Number of solo paper * male		-0.0263		-0.0354		0.1075^{*}
		(0.0297)		(0.0252)		(0.0565)
Male	-0.0099	-0.0066	-0.0019	0.0180	-0.0146	-0.0357^{***}
	(0.0115)	(0.0170)	(0.0102)	(0.0163)	(0.0092)	(0.0137)
SSHRC	0.0021	0.0028	-0.0088	-0.0077	0.0382^{***}	0.0381^{***}
	(0.0112)	(0.0113)	(0.0101)	(0.0101)	(0.0109)	(0.0108)
$ASSO_{iut}$	-0.0486^{***}	-0.0483^{***}				
	(0.0135)	(0.0135)				
$ASST_{iut}$	-0.1674^{***}	-0.1672^{***}				
	(0.0231)	(0.0231)				
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,139	2,139	1,180	1,180	301	301
Log Likelihood	-352.46	-352.20	-296.21	-294.041	-120.32	-116.41
Pseudo R^2	0.1759	0.1765	0.1154	0.1219	0.2001	0.2261

Note: The dependent variable is the likelihood of promotion and takes the value one or zero. All specifications are estimated using a probit model and the marginal effects are presented. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 2.27: Percentage of types of co-authorship

Category	percentage
same university	67.67%
nearby university	8.49%
other university	23.84%

Table 2.28: Summary statistics of main variables

Variables	Observations	Mean	SD	Min	Max
Likelihood of co-authorship	402,667	0.001145	0.0338	0	1
Network	402,667	0.000057	0.0076	0	1
overlapping research interest	402,667	0.1157	0.1825	0	1
Same gender	402,667	0.7733	0.4187	0	1
Same affiliation	402,667	0.0836	0.2767	0	1
Same grad school	402,667	0.0278	0.1644	0	1
Average research ability	402,667	0.9840	0.7918	0	7.5
Difference in research ability	402,667	1.0633	1.1126	0	9
Average experience	402,667	16.4045	7.2222	0	44.5
Difference in experience	402,667	11.6206	8.5266	0	51

Table 2.29: Marginal effects from Probit estimates of the likelihood of co-authorship

	(1)	(2)	(3)
Same affiliation	0.0032***	0.0033***	0.0049***
	(0.00017)	(0.00017)	(0.00025)
Overlapping research interest	0.0017***	0.0018***	0.0032***
	(0.00023)	(0.00023)	(0.00037)
Same gender	0.00004	0.00004	0.00027
	(0.00013)	(0.00013)	(0.00024)
Average research ability	0.0013***	0.0013***	0.0022***
	0.00078)	(0.00078)	(0.00013)
Difference in research ability	−0.00067***	−0.00068***	−0.00098***
	(0.0.00059)	(0.000059)	(0.000086)
Average research experience	−0.000022***	−0.000022***	−0.000054***
	(0.0000000)	(0.0000000)	(0.000015)
Difference in research experience	-0.00001	-0.00001	-0.00001
	(0.0000000)	(0.0000000)	(0.0000000)
Same grad school	0.0017***	0.0017***	0.0023***
	(0.00018)	(0.00018)	(0.00027)
Network	omitted from the estimation: just 23 occurrences		
Year fixed effect	No	Yes	Yes
Pair fixed effect	No	No	Yes
Number of obs	402644	402644	402644

2.8 Figures

Figure 2.1: Percentage of co-authored papers, 1996 to 2012

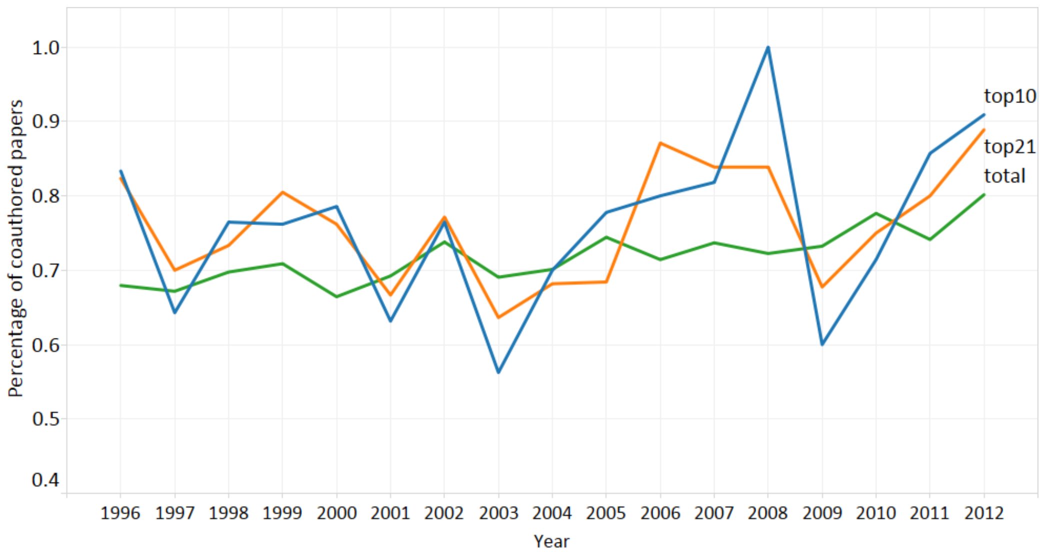


Figure 2.2: Single and co-authored papers, by university

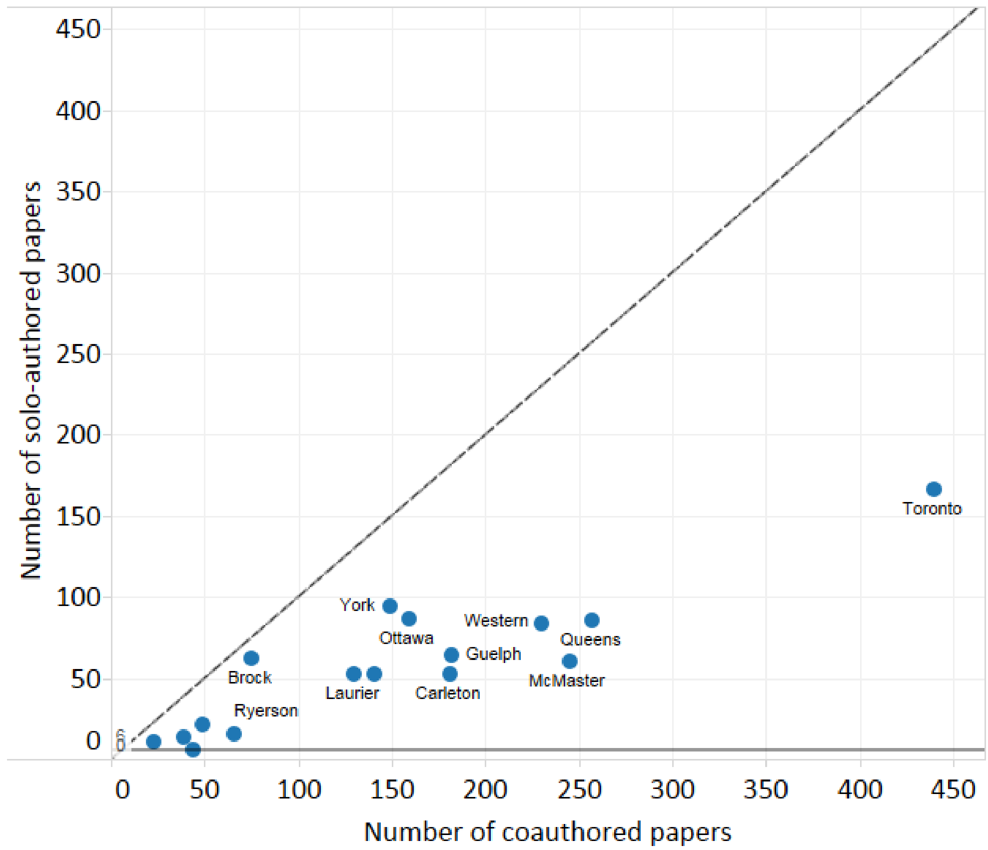


Figure 2.3: Single and co-authored papers, by gender

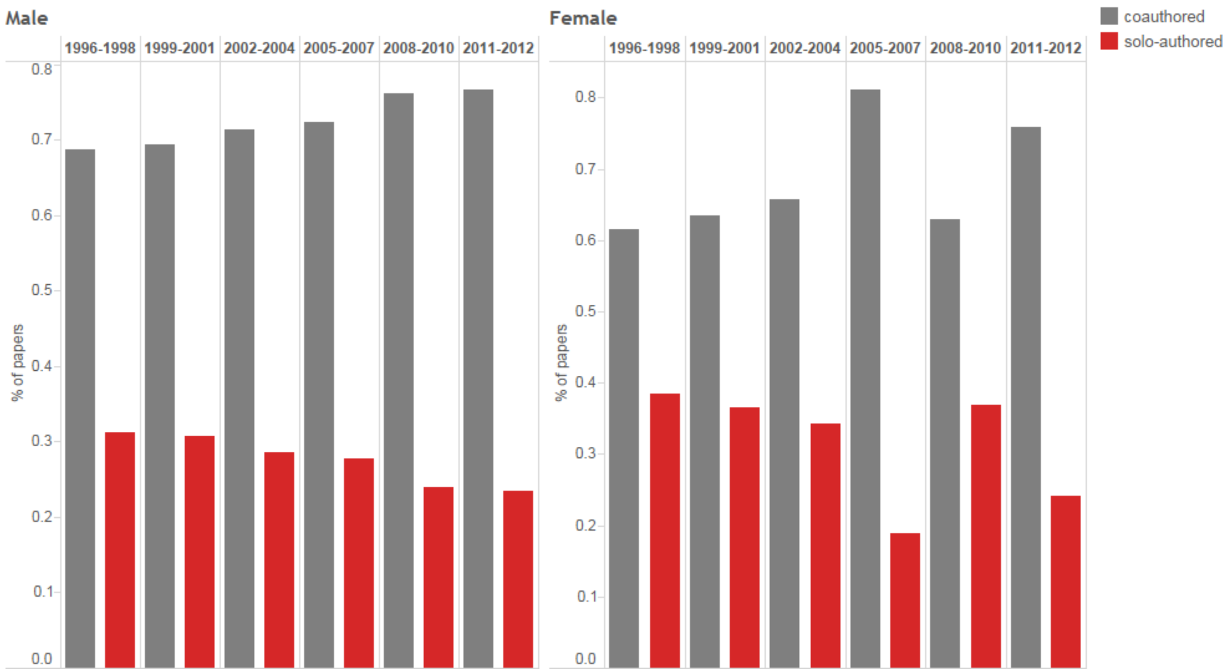


Figure 2.4: Percentage of different types of co-authored papers from 1996 to 2012

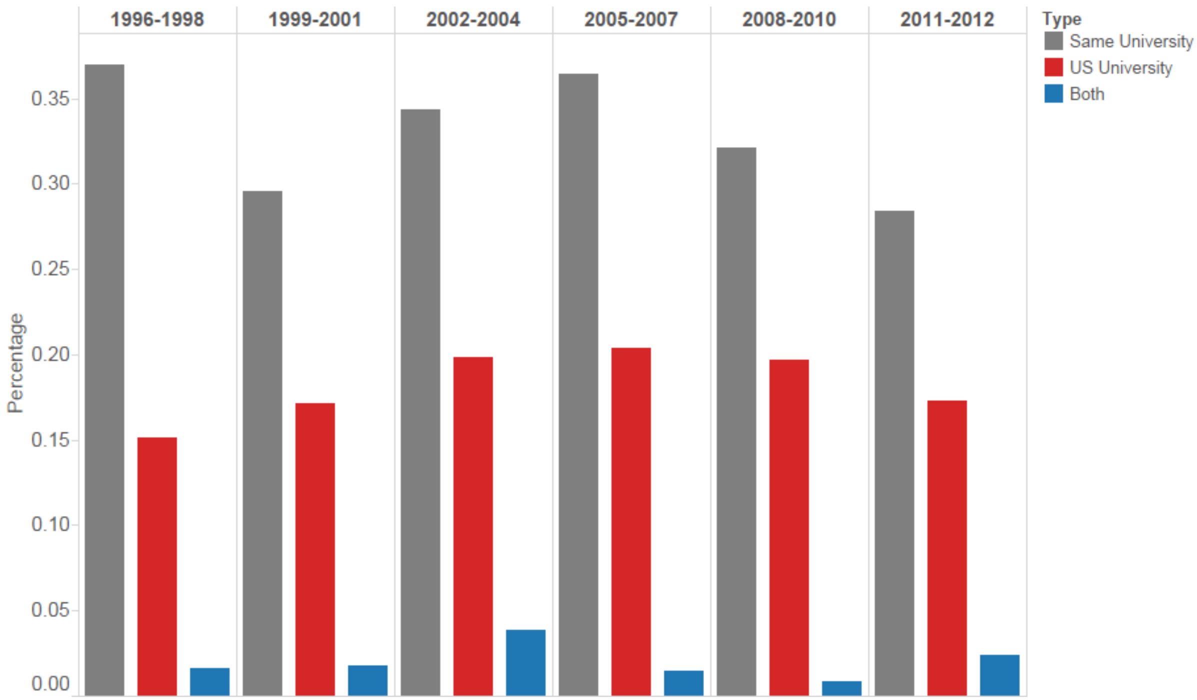
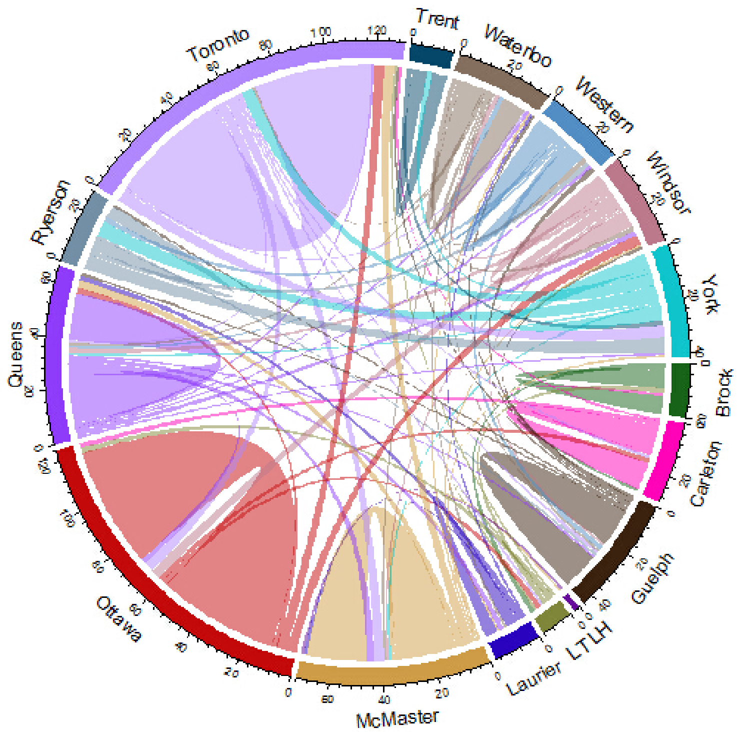
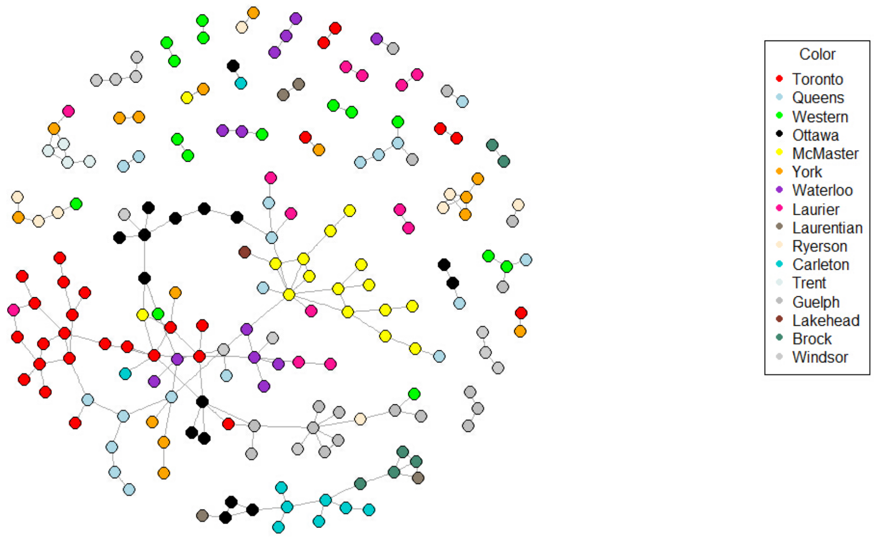


Figure 2.5: The co-authorship network among Ontario universities



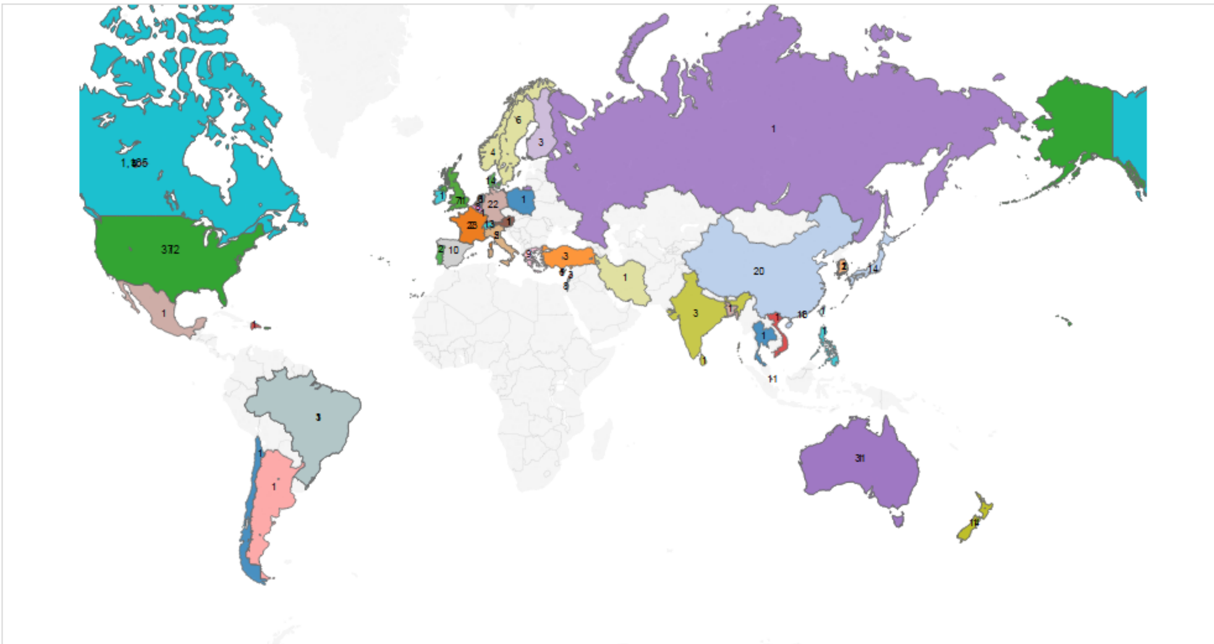
Note: colors identify different universities and the thickness of links identify the number of collaborations.

Figure 2.6: Co-authorship network among Ontario economists.



Note: colors identify different universities and each node in the figure represents an Econ professor who co-authored a publication with another Econ professor from Ontario.

Figure 2.7: The global distribution of Ontario economists' coauthors



Note: colors identify different countries.

Chapter 3

Nearby School Competition and School Performance in Ontario

3.1 Introduction

Given the wide concern about the effectiveness of input policies such as added resources and increased attention to teacher qualifications [1], more and more education policies around the globe focus on proposals to increase parental choices or enhance competition among schools. There is an increasing volume of studies investigating the impact of school choice and school competition [2] on school performance. However, the results derived from the relevant studies are mixed, and most of them are U.S.-based. There is much less research based on Canadian data that evaluates school academic performance in the context of the direction and magnitude of variation of competition from nearby schools. From a simple perspective, this paper aims to answer the following question: Do public schools facing more competition (measured by the total number of nearby schools within a certain distance) perform better than similar schools facing less competition? I think this is an important research question that has remained relatively unaddressed in Canada. The conclusion drawn from this study might shed light on understanding the potential effect of specific school reforms in the context of the debate about whether or not to adopt an "open enrollment" policy.

[1] Hanushek (2003)

[2] According to Gibbons et al. (2008), school choice is a property of residential location and competition is a property of school location.

Competition pressure between schools can create some incentives for schools to improve their educational performance, and this competition has become more pronounced with the transparency of different measures of school performance, such as the Grade 9 Assessment of Mathematics (by schools) and the Ontario Secondary School Literacy Test (OSSLT). Alternatively, it is also plausible that schools react to competition by improving their competitive positions through particular marketing or promotional activities [3] rather than improving educational performance. Schools might also respond to an increase in competition by enrolling only those students who have weak preferences for school performance rather than making an effort to enhance performance [4]. Given this theoretically ambiguous effect of nearby school competition on school performance, solid empirical evidence is needed.

The only recent study that has focused on the effects of local competition in Ontario is by Card et al. (2010) who compare test score gains between third and sixth grade of English-language Catholic public schools and public elementary schools, employing student-level data from 1998 to 2005. In contrast to Card et al. (2010), I expand the measurement of competitive effects of schools by looking at the impacts of changes in the number of all public schools (irrespective of the school board) on certain measures of academic performance for each public school in urban Ontario. Schools may not only compete with one another by geographic location but also through perceptions of quality. Most research has emphasized the use of geographic market concentration as measures for competitive effects, but not the quality of other schools. In this analysis, the average performance of nearby schools is also included to account for the fact that quality competition may have a different effect on school performance. In contrast to Card et al. (2010), I mainly focus on secondary schools and use more recent data, from 2004 to 2013. However, I do employ a similar identification strategy by exploiting openings and closures of secondary schools in local areas.

The key reasons to consider secondary education in Ontario can be listed as follows: First, given that secondary education is frequently discussed and the issue of high school dropout rates has received much concern, analysis of the competition among secondary schools is quite critical. Second, Catholic schools at secondary school level are required to accept non-Catholic students as a condition of their public funding [5], I think this may introduce a greater degree of competition than in the elementary school market. Third, students in Ontario can attend schools other than the assigned ones within a school board depending on the capacity of the selected schools. I believe that students at the

[3] Lubienski (2005)
[4] Gibbons et al. (2008)
[5] Leonard (2015)

secondary level are more likely to choose not to attend assigned schools since they are more independent and distance from home to school poses less restriction than for students at elementary level. Finally, to the best of my knowledge, no previous study has investigated the relationship between secondary school competition and school performance in Ontario.

This analysis uses a school-level panel data set obtained from the Ontario Ministry of Education and the Educational Quality and Accountability Office (EQAO) to estimate whether average school performance in standardized tests is affected by competition from nearby schools. The availability of data on a panel of schools allows me to control for the potentially confounding effects of unobserved school-specific attributes. I employ fixed effect, random trend, and instrumental variables (IV) estimation to eliminate the potential bias associated with competition between schools. Following Gibbons et al. (2008), I use proximity to school board boundaries as an instrumental variable for local school competition. The ordinary least squares (OLS) estimates show that there is a negative association between school competition and school performance, controlling for other relevant factors. However, the fixed effect estimates show that performance improved slightly for schools facing more competition. IV estimates suggest a positive but statistically insignificant association between competition and performance. Another important finding of this analysis is that both OLS and IV estimates are positive and statistically significant when the sample is restricted to the Toronto District School Board, which suggest that competition may improve school performance in schools where students are given more freedom to choose schools other than those to which they are assigned.

Due to data availability, there are several limitations in this paper. Firstly, the unit of analysis in this paper is a school; although the majority of research also uses school-level data, the limitation remains as this cannot account for student fixed effect to reduce heterogeneity bias. Next, I do not have precise test score data. Based on the information obtained from the Education Quality and Accountability Office (EQAO), in this analysis, the percentage of students at level 3 and level 4 of the Grade 9 Assessment of Mathematics and the percentage of students passing the Ontario Secondary School Literacy Test (OSSLT) are used as the measurement for secondary school performance. To enhance the credibility of empirical results, in the sensitivity analysis, I use another data set collected from the Fraser Institute, which does have data on average scores for grade 9 mathematics of each school over the period 2009 to 2013. Finally, school markets in this analysis are defined by the fixed radius approach. If I had information on detailed student home addresses, I would be able to choose the radius of the circle of a given school by calculating the average school-home distance of students enrolled in that school.

The remainder of this paper is organized as follows: Section 2 contains a review of the related literature. Section 3 displays a simple model of nearby school competition and school performance. Section 4 briefly introduces the Ontario secondary education system. Following that is a description of the construction of the data used. An empirical model and results are detailed in Section 6. Section 7 concludes this paper.

3.2 Literature Review

Tables 3.1 and 3.2 present a survey of studies that analyze how school competition affects school performance in North America and Europe, respectively. While the issue has been studied by many researchers around the world using different approaches and data sets, no consistent conclusion has yet been drawn.

In general, the most common empirical framework in the school competition and performance analysis uses spatial-based school competition measures to explore the effects of competition in different school markets. The main challenges that have arisen are that there is no consensus on how to appropriately define the market area, nor on how to measure competition intensity within a certain market area. Due to the differences in units of analysis, research methodology, as well as the different educational programs in different contexts, the debate continues about how spatial competition affects traditional public schools.

Several studies explore the number of schools within a fixed radius circle. By defining school market area as a region enclosed by a 5-mile circle centered on a given school, and applying school-level data in Michigan between 1996 to 1997, Bettinger (2005) suggests that charter schools have no significant effect on test scores in neighboring public schools. Using a similar strategy, Sass (2006) investigates the competitive effect from a charter school in Florida, and concludes that its effect has a small positive impact on the performance of Florida's traditional public schools. Using fixed effect estimation and six different measures of competition pressure, including the number of charter schools within 2.5 miles of a given regular public school in California, Zimmer and Buddin (2009) also find charter schools have a small positive impact on traditional public schools.

Other researchers evaluate the competition effect using school board districts to define school market areas. They define competition either by the percentage of students who transferred to a charter school, or the percentage of a school district's total enrollment attending charter schools. By creating a dummy variable taking the value of one if the percentage of charter school enrollment exceeds 6% as the charter school competition measure, and employing school level data in Michigan from 1994 to 2004, Ni (2009) argues that there is a statistically significant adverse effect of nearby charter schools on traditional public school efficiency in the long-run. Ni concludes that a one standard deviation increase in measured competition is associated with a 0.2 standard deviation decrease in satisfactory rates in math and 0.5 standard deviations decrease in reading. Using a similar strategy, Carr and Ritter (2007) show that charter schools have a small but statistically significant negative effect on traditional public schools in Ohio.

Another challenge that has arisen in attempts to exploit school competition effect in the literature is how to address the endogeneity and omitted variable bias problem which is raised by the fact that schools are not randomly located [6]. Using the number of streams and rivers in a metropolitan area natural boundaries to derive instruments for school competition, Hoxby (2000) finds a one standard deviation increase in school competition is associated with 0.27 standard deviation increase in average public school students' educational attainment. Similarly, using the number of buildings with an area between 30,000 and 60,000 square feet and the number of shopping malls within a certain radius as instruments for charter school location, Imberman (2011) finds a negative impact of neighboring charter schools on math and language test scores of traditional public schools in the southwest. Utilizing the distance to district boundaries as an instrumental variable, Gibbons et al. (2008) suggest the impact from greater school competition is limited in England. Using distance between schools and the city center as an IV for competition, Noailly et al. (2012) conduct the school competition analysis in the Netherlands, finding school competition has a small positive significant effect on a competing school's performance. Böhlmark and Lindahl (2015) focus on the effect of competition from the expanding independent school sector on average educational performance in Sweden. They find the greater the proportion of independent school students, the better the average academic performance of all students at both the end of compulsory schooling and in long-run educational outcomes.

In addition to looking at the spatial competition effect, to date, there are also few studies investigating how the quality of neighboring schools affects the performance of a given public school. By focusing on school districts within the seven largest metropolitan areas in Ohio, Staley and Blair (1995) conclude that higher unweighted test scores of neighboring districts are associated with the higher performance of a given public school. However, the effect is small. The mechanisms of the effect of quality competition have been summarized in their paper as follows: First, a poorly-performing district which is surrounded by better-performed districts will experience emigration. Second, citizens in a poorly-performing district will put pressure on their officials to improve performance. Then last, but not least, the better-performed district may have a demonstration effect on a poorly-performed district, and this effect is more pronounced in nearby districts. Using a multi-dimensional approach and employing panel data from 1990 to 2000, Millimet and Rangaprasad (2007) analyze the strategic competition among public schools in Illinois. They analyze five district

[6] Ni (2009)

level inputs which potentially impact educational quality: student-teacher ratio, average teacher salary, current expenditure per student, capital expenditure per student, and school size of neighboring school districts. These are applied on a given public school district and show that there is substantial evidence of public school district responses to competition from the neighboring districts.

The evidence in Canada is relatively sparse. Exploring competition between publicly-funded Catholic school and public school in Ontario, Card et al. (2010) find out that the test score gains between third and sixth grades is 0.03 to 0.05 of a standard deviation higher when there is a 40 percent increase in the proportion of individuals who can choose between education systems. A recent report [7] by the C.D. Howe Institute, which explores the effect of greater competition driven by British Columbia's "open enrollment" policy suggests that increased competition has increased the numeracy and reading scores of grade 4 students.

Altogether, a considerable amount of U.S.-based literature has been published studying the impact of school competition on school performance. To date, there is no research based on Ontario data addressing the relationship between school competition and school performance at the secondary school level. By adopting the commonly-used methodology in the literature of creating a circle of certain difference and focusing on the effect of nearby school competition in Ontario, Canada, my paper aims to contribute to the growing literature by filling this gap.

3.3 A simple model of nearby school competition and school performance

In this section I adopt a simple version of a "circular city" model to link nearby school competition and school performance, and to motivate our empirical study.

Assume n schools position themselves symmetrically around a unit circle, there are m families uniformly distributed, and each family has one school-aged child. An individual chooses to attend school i, with the following utility function:

$$U = \alpha + \beta q_i - \delta(x - Z_i)^2$$

Where q_i is school i's quality, x is individual's home address, Z_i indicates school i's

[7] Friesen et al. (2015)

location, $\delta(x-Z_i)^2$ measures the dis-utility of the distance between home and school. A family located between school i and $i+1$ will be indifferent to school i and $i+1$ if:

$$\alpha + \beta q_i - \delta(x-Z_i)^2 = \alpha + \beta q_{i+1} - \delta(x-Z_{i+1})^2$$

Since it has been assumed that n schools position themselves symmetrically around a unit circle, if the distance of family x to school $i+1$ is X, the distance of x to school i should be $\frac{1}{n} - X$. Consequently:

$$\beta q_i + \delta X^2 = \beta q_{i+1} + \delta(\frac{1}{n} - X)^2$$

$$X = \frac{n\beta}{2\delta} * (q_{i+1} - q_i) + \frac{1}{2n}$$

Similarly, a family located between school i and $i-1$ will be indifferent to school i and $i-1$ if:

$$X = \frac{n\beta}{2\delta} * (q_{i-1} - q_i) + \frac{1}{2n}$$

Then the total demand of school i is given by:

$$d = m[\frac{1}{n} + \frac{n}{2}(q_{i+1} - q_i + q_{i-1} - q_i)]$$

Given the demand and the fact that the funding for all publicly-funded schools in Ontario is linked to enrollments, and assuming the quality of school i's competitors is $E(q)$, school i choose e_i to maximize the following profit function:

$$\pi_i = (\bar{V} - C(e_i)) * d_i \;\; = (\bar{V} - C(e_i)) * m * [\frac{1}{n} + n * (E(q) - q_i)]$$

Where $\bar{V}$ is per-pupil funding, C is a convex cost function, and assuming school quality is a function of effort, then the optimal level of e_i^* should fulfill the following condition:

$$\frac{\partial \pi}{\partial e} = (\bar{V} - C(e)) * \frac{\partial d(e)}{\partial e} - d(e) * \frac{\partial C(e)}{\partial e} = 0 \tag{3.1}$$

Rearranging Equation 3.1, we have:

$$[\frac{1}{n} + n * (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} - (\bar{V} - C(e)) * n * \frac{\partial q(e)}{\partial e} = 0 \tag{3.2}$$

Total differentiating Equation 3.2 with respect to n and e, we have:

$$\begin{aligned} D = -n * \frac{\partial C(e)}{\partial e} * \frac{\partial q(e)}{\partial e} * de^* + [\frac{1}{n} + n * (E(q) - q(e))] * \frac{\partial^2 C(e)}{\partial^2 e} * de^* \\ +n * \frac{\partial q(e)}{\partial e} * \frac{\partial C(e)}{\partial e} * de^* - (\bar{V} - C(e)) * n * \frac{\partial^2 q(e)}{\partial^2 e} * de^* \\ +[n^{-2} + (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} * dn - (\bar{V} - C(e)) * \frac{\partial q(e)}{\partial e} * dn \end{aligned} \tag{3.3}$$

Rearranging Equation 3.3, we have:

$$\frac{de^*}{dn} = \frac{[n^{-2} + (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} - (\bar{V} - C(e)) * \frac{\partial q(e)}{\partial e}}{[\frac{1}{n} + n * (E(q) - q(e))] * \frac{\partial^2 C(e)}{\partial^2 e} - (\bar{V} - C(e)) * n * \frac{\partial^2 q(e)}{\partial^2 e}} \tag{3.4}$$

Since C is a convex cost function and q is a concave function, $\frac{\partial^2 C(e)}{\partial^2 e} < 0$, and $\frac{\partial^2 q(e)}{\partial^2 e} > 0$. The sign of $\frac{de^*}{dn}$ depends on $[n^{-2} + (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} - (\bar{V} - C(e)) * \frac{\partial q(e)}{\partial e}$. If $[n^{-2} + (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} < (\bar{V} - C(e)) * \frac{\partial q(e)}{\partial e}$, then $\frac{\partial e^*}{\partial n} > 0$, which means when the nearby school competition increases, competing schools do exert more effort in order to retain enrollment. On the other hand, if $[n^{-2} + (E(q) - q(e))] * \frac{\partial C(e)}{\partial e} > (\bar{V} - C(e)) * \frac{\partial q(e)}{\partial e}$, which means that schools also can respond to the increase in competition by reducing costly effort, and serve only those with weak preferences for school performance [8]. Given this theoretically ambiguous effect of nearby school competition on school performance, solid empirical evidence is needed.

3.4 A brief introduction to the secondary education system in Ontario

To understand the potential sources of incentives for schools to compete with one another in Ontario, it is necessary to present some details of the Ontario's secondary education market before proceeding to the description of data and empirical results.

[8] McMillan (2005)

According to the Ontario Ministry of Education, publicly-funded secondary education in Ontario is administered by Ministry of Education, and there are approximately 700,000 students attending more than 850 publicly-funded secondary schools [9] with about 5% of these 700,000 students enrolled in private schools [10]. Generally speaking, publicly-funded English school boards in Ontario are divided into two systems: Public School Board and Catholic Separate School Board [11]. Figure 3.1 displays total enrollment and enrollment growth rate of public as well as Catholic schools from 2000 to 2012, respectively. As shown, there is a clear trend of increasing enrollment in Catholic secondary schools over time, as the enrollment growth rate of Catholic secondary school is higher than that of public schools. The period between 2002 and 2004 saw a dramatic decrease in the enrollment growth rate. This greatest decline was due to Ontario's Tuition Tax Credit, which allowed parents to offset certain parts of the cost of tuition for a child attending a private school at the elementary and secondary school level. This policy was introduced in 2002 but canceled by the new Liberal government in 2003. This tax credit policy apparently impacted the 2001-2002 and 2002-2003 parental schooling choices, and thus affected the enrollment rate of all public and Catholic schools [12].

Figure 3.2 displays the total number of secondary schools, as well as the number of new secondary schools opened from 2000 to 2012. As can be seen, 107 new secondary schools operated by English language boards opened in urban Ontario during 2000-2012.

There are several potential sources of competition incentives in Ontario's secondary public school market. First, in some school boards, students are allowed to attend schools other than assigned schools within the school board, depending on the capacity of applied schools, although some school boards have differing policies regarding this issue [13]. This certain degree of freedom for students to access other schools, along with the fact that funding is linked to school enrollment volume, may provide some incentive for each school to attract students by improving their performance, or by using other strategies. Second, the existence of two public school boards, along with the fact that-at the secondary school level-Catholic schools are required to accept non-Catholic students as a condition of their

[9] Ministry of Education website.

[10] Leonard (2015)

[11] Ministry of Education website.

[12] I am not conducting research on private schools in this analysis.

[13] For example, in Toronto District School Board (TDSB), there is the TDSB Optional Attendance Policy and Procedure which gives opportunities for students to access schools outside of the designated attendance area in which they reside. In the Waterloo Region District School Board, according to Student out of Boundary Transfer Requests, it is hard for students to attend another school other than the assigned school.

public funding [14] provide another source of competition incentive for Ontario public schools. Another factor is the transparency of external exam results may also provide some incentive for schools to compete.

It is necessary here to clarify exactly where the variation in the level of competition is derived in this analysis. First, a different number of nearby schools would have differential competitive effects, which provides some variation in spatial competition across schools. Public schools in regions with more nearby schools would face greater pressure on their enrollment than those facing less competition. The differences in the quality of nearby schools may provide another source of variation in quality competition across schools. A customer's reservation quality level will be higher when competitor's quality improves and, to maintain market share, previously inefficient firms have to improve quality as well [15]. Also, as shown in Figure 3.2, 107 secondary new schools opened in urban Ontario between 2000-2012 [16], which provide some time-series variation in the level of competition.

Based on these ideas, this paper makes use of these variations in the degree of competition in different school markets to study the relationship between school competition and school performance. In this analysis, school markets are defined under the fixed radius approach, which is a commonly-used method in the related literature. Based on the fixed radius approach, every school has a unique market, which is the area enclosed by a circle centered on the given school with a fixed radius.

3.5 Data

This analysis employs a provincial-wide school-level panel data set of all public schools in urban Ontario from 2004 to 2013. The data set is combined with Ontario Ministry of Education and the Educational Quality and Accountability Office (EQAO) data. The finalized data includes school-specific information on school performance, nearby school competition, and other school and neighborhood characteristics.

Measurement of school competition

In general, two measures of school competition are used in this analysis: (1) the number of nearby secondary schools within a circle; (2) the average academic performance of nearby secondary schools.

[14] Leonard (2015)

[15] McMillan (2005)

[16] There are also some exits happened in our sample period, however, the number of closures are very limited.

1. Identification of new school openings/closures

The first step in constructing data was to find out when and where a new school opened/closed, and which schools are affected by the opening/closure. To identify an opening school, I obtained the contact information for publicly-funded schools in Ontario from the Ministry of Education. The data includes school level, school type, school board, school open date, school address and postcode, etc. From the information on "school open date" and "school address," I can determine when and where a given school opened. To identify when and where a school closure happened, I obtained further data from Ministry of Education on the list of open publicly-funded schools in Ontario for each academic year between 1998-1999 and 2012-2013. To identify closed schools, I matched the data by school number on a year-by-year basis. If a given school was not on the list for a given year, it was coded as a closed school.

2. Identification of affected schools

As previously stated, since not every school in the same neighborhood as the opened/-closed school will be affected by it, I defined an area of influence based on a widely-used method in the related literature. Following this approach, I identified affected schools within a circle of 5 km, since that is the average distance students travel from home to high school in Ontario [17]. I converted address and postal code information of each school into longitude and latitude, and then located every school on a map [18]. Using this self-created map, I could rapidly identify the number of affected schools within a circle of any radius value. As pointed out by Leonard (2015), due to school board policies, students usually cannot cross school board boundaries to attend other schools. Hence, when I calculate the number of affected schools within a circle, I exclude those schools in other school boards.

To clarify how the number of affected school is identified through the self-created map, I take Gary Allan High School - Burlington as an example, Figure 3.3 is a snapshot of the map showing schools with a 2 km circle of Gary Allan High School - Burlington. As can be seen, there are one Catholic school and one traditional public school located within the 2 km circle of Gary Allan High School - Burlington.

Figure 3.4 indicates the distribution of Ontario's publicly-funded secondary schools. As of 2012-2013, there were 3,978 elementary and 913 secondary schools in Ontario [19]. I focus on English language publicly-funded secondary schools [20] in urban Ontario for my

[17] Leonard (2015)

[18] http://www.easymapmaker.com/map/51c2292133e17f28ab7321626525cb25

[19] Ministry of Education website.

[20] The reason is that if I had a children went to a English school, I would be more interested in the

analysis, and exclude summer schools, night schools, care/treatment schools, schools for the blind and deaf, vocational/occupational schools, and others.

Measurement of school performance

Test score results from the Educational Quality and Accountability Office (EQAO) are used as a measurement of school performance. EQAO is an independent government agency in Ontario that administers provincial standardized tests, including assessments of reading, writing, and mathematics in primary and junior divisions, the Grade 9 assessment of academic and applied mathematics, and the Ontario Secondary School Literacy Test (OSSLT) [21].

In Ontario, for Grades 9 and 10, mathematics courses are designated as either academic or applied. This system was introduced by the Ministry of Education in 1999. In this paper, terms used to refer Grade 9 assessment of academic and applied mathematics are AP and AC, respectively. According to the EQAO website, if a student entered Grade 9 in September of a year and is working toward an Ontario Secondary School Diploma (OSSD), he/she is required to take the OSSLT in that year for the first time. However, if a student is absent, deferred or not successful during a previous administration, he/she can still take the OSSLT in the following year, and those students are termed "previously eligible." The terms FTE and PE are used in this analysis to refer to "first time eligible" and "previously eligible," respectively.

The EQAO test is a standardized test which is comparable over time and across all public schools. EQAO uses a four-level scale to report student performance, where Level 3 and Level 4 indicate the student performance is above the provincial standard, and Level 1 and Level 2 mean achievement is below the provincial standard. I use the percentage of students at levels 3 and 4 of the Grade 9 Assessment of Mathematics, and the percentage of students passing the OSSLT as the secondary school performance measures. According to EQAO's website, school performance reports are displayed using two different methods. Method 1 expresses "the number of students attaining each level as a percentage of all of the students in that grade, including students who were exempted and those who took part in the assessment but did not produce enough work to be scored." Method 2 expresses "the distribution of student results as a percentage of those students who took part in the assessment and produced work that could be scored." Our measure follows EQAO's Method 1 of reporting school achievement, since it is EQAO's primary method of reporting performance. Schools and school boards are required to use it to ensure consistency of

performance of other English schools, however, further research should be undertaken to incorporate French schools.

[21] EQAO website.

reporting.

Figure 3.5 provides the kernel density plot of school achievement from 2004 to 2013, by different EQAO assessments. The first two graphs show the kernel density plot of the percentage of students passing the OSSLT, and the last two graphs present the kernel density of proportion of students above Level 3 and Level 4 of the Grade 9 Assessment of Mathematics. As can be seen, the average percentage of students passing the provincial standard in both assessments varies from year to year. The average number of nearby schools and average school performance of Ontario public school boards are presented in Figure A.2 in the Appendix.

Neighborhood characteristics

Following Card, Dooley, and Payne (2010), I match the publicly-funded school information to Canadian Census data for 2001, 2006, and 2011, aggregated to the "Forward Sortation Areas" (FSA) level to obtain neighborhood characteristics information. For a given school, I use the 3-digit postal code to identify its FSA, and then match the census data to schools. The 2011 census data tabulated at FSA level can be obtained directly from Statistics Canada, and for 2006 and 2001 census data, I use the 2010 Postal Code Conversion File (PCCF) to convert data on Dissemination Area level to FSA. Neighborhood characteristics include, among others, median household income, median dwelling value, population density, and educational characteristics.

School characteristics

Besides school test results, EQAO also provides several other school characteristics, including the percentage of students taking ESL (English as a Second Language) and the percentage of gifted students. I combine the EQAO data with data obtained from Ministry of Education to control for school characteristics. All in all, school characteristics in this analysis include the number of years of school existence, total enrollment, and the percentage of gifted students.

Table 3.4 presents summary statistics of main variables used in my analysis, and Table 3.5 displays the detailed data collection and organization procedures. As can be seen from Table 3.4, the average proportion of students achieving the standard in Grade 9 Academic Mathematical Assessment is 0.74 and the standard deviation is 0.16. And the average proportion of students achieving the standard in Grade 9 Applied Mathematical Assessment is 0.44 which is lower than that of academic mathematics.

3.6 Empirical model and results

Having discussed the construction of the data set used, I now discuss the model and some of the empirical results. As baseline estimates, I rely on pooled OLS to determine how nearby school competition affects school performance. I also separate the competition effect of Catholic schools from that of traditional public schools. I then employ other alternative estimation methods, including fixed effect and random trend model, to account for the possibility that time-invariant unobserved school characteristics are correlated with the number of nearby schools, and different schools have distinct time trends, respectively. Finally, instrumental variable estimation is employed to eliminate the endogeneity bias.

3.6.1 Empirical model

In general, separate test specifications in the Grade 9 assessment of mathematics (both AC and AP) and the OSSLT (both PE and FTE) are estimated using the following econometric model:

$$Y_{it} = \theta * X_{it} + \gamma * Z_{it} + \beta * C_{it} + \alpha_i + T + \epsilon_{it} \tag{3.5}$$

Where Y_{it} is the corresponding EQAO test passing rate of affected school i in year t; Z_{it} and X_{it} are vectors of time varying area and school attributes, and C_{it} is a vector of competition measures; α is a school-specific fixed effect, counting for factors that differ across schools but are time invariant which may partially account for school heterogeneity; T is a vector of year-fixed effect; ϵ_{it} is an idiosyncratic error term.

Firstly, the competition measure C_{it} is captured by the number of secondary schools within a circle. In order to fully understand how nearby school competition affects the performance of a competing school, I then extend C_{it} to two dimensions: (1) the number of secondary schools within a circle; (2) the average academic performance of nearby secondary schools.

3.6.2 Benchmark estimates

As baseline estimates, I pooled the data over years and across schools to estimate the school performance as a function of the number of nearby schools using OLS. Table 3.6 presents the pooled OLS estimators for both the OSSLT and Grade 9 Assessment of Mathematics. For each test score, the first column shows the effect of all nearby schools, and the second column lists the results of the model where we distinguish between the number of nearby

Catholic schools and the number of regular public schools. Standard errors are clustered at the school level. All regressions include the full set of controls and year-fixed effect to allow for any systematic variation in school performance over time, and that is common to all schools but not attributed to other controls.

I then add school board fixed effects in the second panel of Table 3.6. As stated earlier, there are important differences between school boards regarding the freedom of students to attend schools other than those to which they are assigned, and Toronto District School Board (TDSB) is more flexible than other school boards. Hence, a dummy for TDSB is added in the third panel to study whether the school competition effect is different for TDSB than others. In the fourth panel, I limit the sample to TDSB. And in the last panel, I restrict my sample to main regions in Ontario including Peel, Halton, Toronto, York and Ottawa.

Empirical estimates in the first panel of Table 3.6 suggest that there is a negative association between the number of nearby schools and school performance. The specific estimate in column (1) reveals that an additional nearby school is associated with an approximate decrease of 0.29 percent in the passing rate of OSSLT (PE). However, it is only statistically significant at 10% level. When I separate the number of nearby schools by traditional public schools and Catholic schools in column (2), it seems that the negative association between the passing rate and school competition comes from traditional public schools, since the estimated coefficient associated with Catholic schools is positive, although it remains statistically insignificant. Results obtained for other tests are quite similar.

School board fixed effects are added in the second panel. As can be seen, all the estimates decrease slightly, and they are statistically insignificant across all columns. The estimated coefficient in column (1) suggests that the passing rate of OSSLT (PE) is predicted to decrease by 0.18 percent when there is an additional nearby school. In column (2), where I separate the number of nearby schools into traditional public schools and Catholic schools, the estimated coefficient associated with public schools is negative while the coefficient related to Catholic schools is positive.

In the third panel, the estimates of the TDSB dummy is negative and statistically significant across all columns, which suggests that compared to other school boards, the competition effect on school performance is different for TDSB, where students have more freedom to attend schools other than their assigned ones within the school board, depending on the capacity of applied schools.

In the fourth panel, I limit the sample to TDSB. All the estimates are positive and most of them are statistically significant. The coefficient estimates of Grade 9 assessment of mathematics (columns (5) and (7)) suggest that an additional nearby school increases the passing rate of the Grade 9 mathematics test by between 5% to 6%. The estimated coefficients associated with the model where I separate the number of nearby schools into traditional public schools and Catholic schools in columns (2), (4), (6), and (8) are also different from the results using all data. The results reveal that there is a positive and statistically significant association between the passing rate and the number of traditional public schools. The estimated coefficient associated with Catholic schools is positive as well, however, they are statistically insignificant. The estimates abstained in this specification suggest that competition may improve school performance in schools where students are given more freedom to select their school. In the last panel, I focus on main regions. All the estimates are positive, however, they are statistically insignificant. Finally, estimates of a log-log specification is presented in Table A.10 in the Appendix. The results are comparable to those in Table 3.6.

3.6.3 Sensitivity analysis

Fixed effect and random trend estimation

It is possible that the unobserved school characteristics are correlated with the number of nearby schools. If this is true, the OLS estimates described earlier would be biased. Thus, I adopt fixed effect estimation [22] to account for unobserved school characteristics. The upper panel of Table 3.7 presents fixed effect estimates. In contrast to the OLS estimates in Table 3.6, all the coefficient estimates of nearby school competition are positive, and they are statistically significant across all columns. Taking column (5) as an example, the estimated coefficient implies that the number of Grade 9 students attaining the provincial standard of academic mathematics is predicted to increase by 3.11 percent when the number of nearby schools increases by one unit. Further, the estimates in column (6) reveal that the number of public schools, as well as the number of Catholic schools, are all positively associated with school performance. The estimates are higher with traditional public schools than that of Catholic schools.

Fixed effect estimation can pick up all the time-invariant unobserved school characteristics. I also apply a random trend model to control for the possibility that each school has its own trend over time. According to Ni (2009), the random trend model allows the time trend of schools which have a quick response when facing more competition to be different from those schools which have no or little response to competition [23]. The lower

[22] To decide whether to use a fixed effects or random effects estimator, the correlated random effect (CRE) model was applied. The results suggest that time-invariant unobservables are correlated with the regressors and that the fixed effects model is preferred over the random effects model.

[23] Ni (2009)

panel of Table 3.7 shows the coefficient estimates derived from the random trend model are all positive, but most of them are statistically insignificant. In general, the random trend estimates suggest that there is a positive but weak association between school competition and school performance. Also, the number of public schools, as well as the number of Catholic schools, are all positively associated with school performance.

Fraser data

The EQAO data set did not allow me to control for average family income, and only provides the percentage of students achieving level 3 and level 4 of the Grade 9 Assessment of Mathematics test, and the percentage of students passing the OSSLT as the measurement for secondary school performance. To enhance the credibility of our empirical results, I utilize another data set collected from the Fraser Institute [24]. The Fraser Institute data is publicly available at the Fraser Institute website [25] in the form of yearly school ranking. The data obtained have a shorter time span of only five years. It is important to note that the Fraser data have average scores data for each school for the Grade 9 mathematics test, and average family income over 2009 to 2013 [26]. Recall that EQAO uses a four-level scale to report student performance, where level 3 and level 4 indicate student performance is above the provincial standard. We previously used the percentage of students at level 3 and level 4 of the mathematics assessment as the secondary school performance measures. However, the Fraser data reports the average test level of each school.

I merge the Fraser test data with the measurement of competition, other school characteristics, and neighborhood characteristics as another type of sensitivity analysis. The estimates obtained from the analysis of Fraser data are presented in Table 3.8. The results appear consistent with my EQAO analysis. The top half of Table 3.8 shows the OLS estimates. No increase in school performance was detected. Most of the estimates are negative and statistically insignificant. The estimated coefficient in column (1) suggests that the number of students passing the OSSLT test in a public school is predicted to decrease by 0.12 percent when the number of nearby schools increases by one unit. The coefficient estimates in column (3) imply that an additional nearby school is associated with a 0.006 point decrease of the average attainment level in the Grade 9 mathematics assessment. In addition, the coefficient estimates of average family income are positive and statistically significant across all columns, which implies that average family income is positively associated with school performance. The fixed effect estimates are presented in the second panel

[24] The data description is presented in Table A.11 in the Appendix.

[25] https://www.fraserinstitute.org/

[26] For OSSLT, the Fraser Institute data also use the percentage of students passing OSSLT to report school performance, but for the Grade 9 Assessment of Mathematics, Fraser have average scores of each school.

of Table 3.8, and the results indicate that there is a positive association between school competition and school performance. In the third and fourth panel, I restrict the sample to Toronto District School Board and school boards in main regions of Ontario, respectively. The estimates in the third panel suggest that there is a positive association between school performance and number of nearby schools in Toronto. However, the estimates in the last panel are statistically insignificant.

All in all, different regression techniques revealed different results. The OLS results of this study indicate that there is a negative association between school competition and school performance, however, the fixed effect and random trend estimates show that there is a positive competition effect. A possible explanation for this might be that the school competition measure is correlated with some unobserved time-varying school/neighborhood attributes. Following the literature, I will later use the distance of a school to a school board boundary as an instrumental variable to overcome the potential endogeneity problem.

Expanding competition measure

In previous tables the competition measure C_{it} is captured by the total number of nearby schools. I now expand the competition measure into two dimensions: the total number of nearby schools, and the average academic performance of nearby schools.

Table 3.9 displays the estimates of the expanded competition measure on the passing rates in both the OSSLT and the Grade 9 Assessment of Mathematics. The upper and lower panel of Table 3.9 display OLS and fixed effect estimates, respectively. To conserve space, we do not report all the coefficient estimates. The empirical estimates are organized similarly to previous tables.

Adding school quality results in some changes. However, all the estimates are comparable to those presented in Table 3.6. The estimates obtained from OLS suggest that there is a negative association between the number of nearby schools and school performance. The estimates for nearby school quality are all positive and statistically significant.

The results for fixed effect estimates are presented in the second panel of Table 3.9. Echoing previous results, the fixed effect estimates suggest that there is a positive and statistically significant association between the number of nearby schools and school performance. For nearby school quality, the FE estimates are positive and considerably higher than OLS estimates. In the third panel, I restrict my data into Peel, Halton, Toronto, York and Ottawa. The estimates of a specification in which I limited my sample to TDSB schools are presented in the last panel of Table 3.9. In accordance with my previous findings, the results of the number of nearby schools are positive in the last two panels. For school quality, all estimates are statistically significant at either the 1%, 5% or 10% level.

To generate nonlinear effects from school competition on school performance, I group the number of nearby schools into 10 categories. The summary statistics by number of nearby schools is presented in Table 3.10 and the OLS estimates of different numbers of nearby schools on school performance with other controls are shown in Table 3.11. It is clear that no significant effect has been identified [27].

3.6.4 Instrumental Variable Estimates

The objective of our empirical research is to identify the effect of competition from nearby schools. The fixed effect and random trend model used in the previous subsection can eliminate any school heterogeneity bias. There is another potential bias arising from the concern that the number of nearby schools may correlate with some unobserved time-varying school/neighborhood attributes. Hence, instrument variable estimation (IV) is applied to address the endogeneity problem in this subsection.

I follow Gibbons et al. (2008), using proximity to school board boundaries [28] as an IV for nearby school competition. The intuition is that schools near a boundary face less competition than schools located in the middle of a school district, since families near a school board boundary face longer journeys to schools other than their closest (and likely assigned) school. Hence, families near a boundary are more likely to attend the assigned school, which then further implies that schools near a boundary face less competition. Figure 3.6 demonstrates the rationality described by Gibbons et al. (2008). Assume we have four schools a, b, c, d within the same school board, evenly distributed in a linear local school market, and school a and d are located near the boundary, while school b and c are located in the center. Each colored dot in Figure 3.6 represents the average distance of family i located along the linear local school market to schools other than the closest one. As shown, the average distance of families to schools other than the closest one is higher for those who live near boundaries than others.

The distance of a given school to its school board boundary is identified using ArcGIS. I first obtained the shapefile of Ontario public school board boundaries from Geospatial Centre [29], and the shapefile-along with the longitude and latitude information of each school-was input into ArcGIS. The Ontario public school board boundary is shown in Figure 3.7. Figure 3.8 presents an example demonstrating how to determine the distance from a given school to a school board boundary using ArcGIS.

[27] I also restricted my data into TDSB, all estimates are statistically insignificant.

[28] Using data pertaining to English, Gibbons, Machin and Silva show that the school-school board boundary distance is a good IV for their competition index.

[29] University of Waterloo.

The first stage results obtained from IV estimation is represented in Table 3.12, and Table 3.13 displays the IV estimates. The first stage results show that a 10% increase in the distance from a school to a school board boundary increases the number of nearby schools by 0.045. The minimum eigenvalue statistic is around "56.20" [30]. Comparing this to the critical value (at the 5% level) of "16.38" [31], we can reject the presence of weak instruments since the statistic greatly exceeded the critical values.

Table 3.13 presents the IV estimates. The first panel shows the results with all data while the second and third panel consist of estimates based on restricted samples. From the first panel of Table 3.13, we can see that the coefficient estimates of the number of nearby schools are positive but statistically insignificant across all columns, which suggests there is a small association between school competition and school performance. The results are different when we limit the sample to Peel, Halton, Toronto, York and Ottawa in the second panel and only TDSB in the last panel, where the coefficient estimates for OSSLT (columns (1) and (2)) and Grade 9 mathematics (columns (3) and (4)) are all positive and statistically significant at the 1% level.

3.7 Conclusion

This paper estimates whether school performance of a given school is affected by competition from other nearby schools. The fixed effect, random trend, and instrumental variable evidence show that school performance improved slightly for schools facing more competition, controlling for other relevant factors. The estimates imply that a one unit increase in the number of nearby schools is associated with a small increase in average public school performance. This finding is consistent with the literature. Exploring competition from Catholic school on the performance of traditional public school in Ontario, Card et al. (2010) find that the test score gains between third and sixth grades is 0.03 to 0.05 of a standard deviation higher when there is a 40 percent increase in the proportion of individuals who can choose between education systems. Leonard (2015) also suggests there is a small positive effect of school choices on student university applications in Ontario.

Another important finding is that the estimated coefficient is stronger when the sample is restricted to the Toronto District School Board, which suggest that competition may improve school performance where students are given more freedom to choose their school

[30] The Stock and Yogo test for weak instruments.

[31] Decision rule: reject the presence of weak instruments if the minimum eigenvalue statistic is larger than the critical value.

since the freedom of choices may provide a better matching for students with schools based on their own needs [32]. This finding matters to the recent discussion about whether nearby school competition can improve school performance and it may lend support to the current policy which is designed to improve public school performance in Ontario. There are some other studies need to done. Based on a principal-agent problem, increased competition may increase private allocative efficiency and school productivity; however, on the other hand, if there are human capital spillovers, increased competition may decrease social allocative efficiency (Benabou Model) in the sense that increased self-sorting through more choices may impair educational outcomes of those who are supposed to benefit from their good peers. As mentioned previously, my research is based on school-level data, which means I could not control for the possibility of a changing student population in this study. Further research is needed to examine the student sorting issue.

[32] Gibbons et al. (2008)

3.8 Tables

Table 3.1: Reviews of studies on the effect of spatial competition on public school performance, North America

Author	Data	Dependent variable	How to define Market area	How to define competition intensity within a market area	Main finding
US					
Hoxby (2000)	National Longitudinal Survy of Youth 1998	test score	a metropolitan area	number of streams and rivers in an metropolitan	there is a 0.27 standard deviation increase in test scores when there is a sd increase in competition.
Bettinger (2004)	Michigan Educational Assessment Program school-level data 1996-1997	reading and math test score in 4th grades	region enclosed by a 5 mile circle centered on a given school	number of chart schools	charter schools have no significant effect.
Sass (2006)	"Sunshine State Standards" Florida Norm-Reference Test 1999-2001	test score gains	region within 2.5, 5 and 10 mile radius of a school	number and enrollment shares of different nearby schools	charter competition has a modest effect on math but no effect on reading scores on traditional public shcools (TPSs) in Florida.
Carr et.al (2007)	standardized math and reading scores 2002-2006	change in average school building proficiency passing rates in reading and math	school district	whether at least one charter school located in the same district, number and market share of charter school in the same district	charter schools has a small negative effect on traditional public schools in Ohio.
Buddin et.al (2009)	Stanford 9 reading and math test scores 1997-2001	student test score	region enclosed by a 2.5 mile circle centered on a given school	distance to charter or other schools, presence of charter or other schools, number of charter or other schools, share of charter or other schools, student lose to charter or other schools.	charter schools have no significant effect on test scores in neighboring public schools in California.
Ni (2009)	Michigan Educational Assessment Program school-level data 1994-2004	reading and math test score in 4th and 7th grades	school district	the degree of competition = 1 if the percentage of charter enrollment reaches 6% and 0 otherwise.	using fixed effect, first difference and IV, they find charter competition has a negative impact on traditional public school efficiency in Michigan.
Imberman (2011)	the Stanford Achievement Test versions 9 and 10 score 1993-2004	student test score and test score gains	region within a specified radius of the regular public school	number of buildings between 30,000 and 60,000 square feet and number of shopping malls within a certain radius	using characteristics of the building stock near TPSs as IV for charter location, the author finds charter schools have a negative impact on math and language score of TPS in the southwest.
Canada					
Card et.al (2010)	Elementary EQAO 1998-2005	test score gains between third and sixth grades	school district	local fraction of Catholic and the neighborhood growth rate	test score gains between third and sixth grades is 0.03 to 0.05 of a standard deviation higher when there is a 40% increase in fraction of individual who can choose between education system.

Table 3.2: Reviews of studies on the effect of spatial competition on public school performance, Europe

Author	Data	Dependent v	How to define Market area	How to define competition intensity within a market area	Main finding
England					
Gibbons et.al (2008)	National Pupil Database 1996-2003	gain in test score from ages 6-7 to ages 10-11	school district	average number of schools accessible to students in the school	using the distance to district boundaries as IV, the authors find the impact from greater school competition is limited.
Netherlands					
Noailly et.al (2009)	the nationwide uniform Cito test 1999-2003	the performance of students at Cito test in the final year of primary education	region enclosed by a 1.5 km circle centered on a given school	number of alternative schools and inverted Herfindahl index	using distance between school and city center as an IV for competition, they find school competition has a small positive significant effect on competing school's performance.
Sweden					
Lindahl (2015)	test scores from 9th grade achievement test	average score	municipalities	share of students attending independent schools	an increase in the share of independent school students is associated with better school performance.

Table 3.3: Reviews of studies on the effect of quality competition on public school performance

Author	Data	Dependent variable	How to define quality competition within a market area	Main finding
US				
Blair et.al (1995)	Standard Achievement Test, the California Achievement Test and the Lowa test	unweighted average of the 4th, 6th and 8th test score	unweighted average of 4th, 6th and 8th test score of neighboring schools	better performance of neighboring districts is associated with better performance of a given public school district, but the effect is small.
Millimet et.al (2007)	Common Core of Data 1990-2000	student teacher ratio, average teacher salary, current per student expenditure, capital PSE and school size.	student teacher ratio,average teacher salary,current and capital per student expenditure and the school size of neighboring school districts.	using multi-dimensional approach, the authors find robust evidence that public school compete with each other in Illinois .

Table 3.4: Summary statistics of main variables used in my analysis

Variables	Observations	Mean	SD	Min	Max
Measurement of school performance					
OSSLT passed PE	3525	0.61	0.18	0	1
OSSLT passed FTE	3525	0.74	0.20	0	1
Grade 9 level 3 math-academic	3525	0.74	0.16	0	1
Grade 9 level 3 math-applied	3525	0.44	0.17	0	0.97
number of nearby schools	3525	6.86	6.39	0	35
neighborhood characteristics					
log average household income	3525	10.61	0.27	10.11	11.82
log average dwelling value	3525	12.74	0.40	11.60	13.95
percentage of children aged 15 to 19	3525	12.73	0.40	11.60	13.95
percentage of Canada citizen	3525	0.91	0.053	0.76	1
percentage with post secondary degree	3525	0.45	0.08	0.24	0.78
percentage of Catholics	3525	0.31	0.11	0.094	0.81
percentage of total immigrant	3525	0.27	0.18	0.027	0.72
percentage of recent immigrant	3525	0.037	0.036	0	0.17
percentage of people of Asian origin	3525	0.19	0.198	0.045	0.87
school characteristics					
number of years of existence	3525	40.97	11.04	4	46
total enrollment	3525	1066.9	398.85	24	2327
percentage of gifted students	3525	0.031	0.059	0	0.643

Source: Author's own calculation.

Table 3.5: Detailed data collecting and organizing procedures

Variable	Source	Procedure
Measurement of competition		
1. Identification of school openings	Ministry of Education	1. obtain contact information of all publicly funded schools in Ontario from Ministry of Education. 2. using the variable "school open date" and "school address" to identify when and where a given school is opened.
2. Identification of school closures	Ministry of Education	1. obtain a list of open publicly funded schools in Ontario for each academic year between 1998/1999 and 2012/2013 from Ministry of Education. 2. match yearly data by school number to identify a given school is on the list for a given year or not. 3. if the school is not on the list, we indicate it as closed.
3. Number of nearby schools	Ministry of Education Self-created map	1. convert address and postal code information of each school into longitude and latitude. 2. locate every school on a self-created map based on longitude and latitude. 3. using this map to identify number of schools within a circle of any radius value.
4. Quality of nearby schools	EQAO	1. calculate average performance of nearby schools
School performance	EQAO	1. request data from EQAO. 2. merge the yearly data by school number.
Neighborhood characteristics	Statistics Canada	1. download Census data from Statistics Canada: the 2011 census data tabulated at FSA level is obtained directly from Statistics Canada; for 2006 and 2001 census, we use 2010 Postal Code Conversion File (PCCF) to convert data on Dissemination Area level to FSA. 2. use 3-digit postal code to identify the FSA of a given school. 3. based on the FSA, we match the census data to schools.
School characteristics	Ministry of Education EQAO	1. directly obtained from Ministry of Education and EQAO.

Table 3.6: OLS estimates based on school level EQAO data (2004 - 2013)

	OSSLT passed PE		OSSLT passed FTE		Grade 9 level 3 math AC		Grade 9 level 3 math AP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OLS with year fixed effect								
Number of all nearby school	-0.0029^{*}		-0.0012		-0.0041^{**}		-0.0024^{*}	
	(0.0015)		(0.0013)		(0.0015)		(0.0014)	
Number of nearby public school		-0.0036^{*}		-0.0017		-0.0037^{**}		-0.0038^{**}
		(0.0019)		(0.0017)		(0.0019)		(0.0018)
Number of nearby Catholic school		0.0009		0.0008		-0.0028		0.0021
		(0.0029)		(0.0023)		(0.0034)		(0.0032)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.3426	0.3423	0.5875	0.5876	0.2153	0.2132	0.2096	0.2107
Observations	3,515	3,515	3,515	3,515	3,515	3,515	3,515	3,515
School Board Fixed effect								
Number of all nearby school	-0.0018		-0.00093		-0.0013		-0.00038	
	(0.0016)		(0.00135)		(0.0015)		(0.00148)	
Number of nearby public school		-0.00081		0.00015		0.00107		0.00137
		(0.0020)		(0.0018)		(0.0020)		(0.0019)
Number of nearby Catholic school		-0.00141		-0.0021		-0.0040		-0.00043
		(0.0027)		(0.0022)		(0.0034)		(0.0028)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.3722	0.3715	0.6012	0.6012	0.2790	0.2800	0.2819	0.2822
Observations	3,515	3,515	3,515	3,515	3,515	3,515	3,515	3,515
Adding a dummy for TDSB								
Number of all nearby school	-0.0015		-0.0005		-0.0021		-0.0007	
	(0.0018)		(0.0013)		(0.0016)		(0.0015)	
Number of nearby public school		-0.0016		-0.00077		-0.00076		-0.00145
		(0.0020)		(0.0018)		(0.0021)		(0.0019)
Number of nearby Catholic school		0.0010		0.00085		-0.0027		0.00215
		(0.0028)		(0.0023)		(0.0034)		(0.0032)
TDSB	-0.0487^{**}	-0.0502^{**}	-0.0341^{*}	-0.0337^{*}	-0.0674^{***}	-0.0739^{***}	-0.0591^{***}	-0.0577^{***}
	(0.021)	(0.022)	(0.0182)	(0.0192)	(0.0218)	(0.0223)	(0.0227)	(0.0232)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.346	0.346	0.588	0.5883	0.2247	0.2239	0.2156	0.2161
Observations	3,515	3,515	3,515	3,515	3,515	3,515	3,515	3,515
Only TDSB								
Number of all nearby school	0.0048^{*}		0.0064^{***}		0.0062^{**}		0.0053^{*}	
	(0.0028)		(0.0027)		(0.0028)		(0.0028)	
Number of nearby public school		0.0063^{***}		0.0075^{***}		0.0084^{***}		0.0062^{**}
		(0.0025)		(0.0024)		(0.0028)		(0.0028)
Number of nearby Catholic school		0.0045		0.0039		0.0051		0.0023
		(0.0069)		(0.0063)		(0.0068)		(0.0086)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.4660	0.4724	0.5204	0.5256	0.3463	0.3569	0.2925	0.2945
Observations	750	750	750	750	750	750	750	750
Restrict data to Peel, Halton, Toronto, York and Ottawa								
Number of all nearby school	0.00065		0.00002		0.0012		0.0013	
	(0.0018)		(0.0014)		(0.0018)		(0.0017)	
Number of nearby public school		-0.00005		0.0020		0.00092		0.00007
		(0.0023)		(0.0017)		(0.0021)		(0.0021)
Number of nearby Catholic school		-0.0040		-0.0048		-0.0027		-0.0036
		(0.0045)		(0.0040)		(0.0050)		(0.0050)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.4339	0.4247	0.5411	0.5427	0.3103	0.3169	0.1993	0.1992
Observations	1,700	1,700	1,700	1,700	1,700	1,700	1,700	1,700

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics.

Table 3.7: Fixed effect and Random Trend estimates based on school level EQAO data (2004 - 2013)

	OSSLT passed PE		OSSLT passed FTE		Grade 9 level 3 math AC		Grade 9 level 3 math AP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fixed effect								
Number of all nearby school	0.0301***		0.0284***		0.0311***		0.0179***	
	(0.0064)		(0.0046)		(0.0058)		(0.0064)	
Number of nearby public school		0.0331***		0.0324***		0.0427***		0.0216**
		(0.0093)		(0.0071)		(0.0085)		(0.0099)
Number of nearby Catholic school		0.0250***		0.0254***		0.0215***		0.0149
		(0.0098)		(0.0087)		(0.0083)		(0.0101)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,515	3,515	3,515	3,515	3,515	3,515	3,515	3,515
Random trend								
Number of all nearby school	0.0071		0.0156*		0.0114		0.0011	
	(0.0065)		(0.0088)		(0.0079)		(0.0105)	
Number of nearby public school		0.0186**		0.0070		0.0091		0.00093
		(0.0088)		(0.0131)		(0.0137)		(0.0165)
Number of nearby Catholic school		0.0019		0.0312***		0.0140		0.0071
		(0.0081)		(0.0118)		(0.0114)		(0.0132)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics.

Table 3.8: OLS and Fixed effect estimates based on school level Fraser data (2009 - 2013)

	OSSLT passed PE		OSSLT passed FTE		Ave Level Grade 9 math AC		Ave Level Grade 9 math AP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OLS								
Number of all nearby school	-0.0012		-0.0007		-0.0061		-0.0161^{***}	
	(0.0022)		(0.0017)		(0.0049)		(0.0058)	
Number of nearby public school		-0.00047		0.0019		-0.00063		-0.0210^{***}
		(0.0028)		(0.0020)		(0.0063)		(0.0089)
Number of nearby Catholic school		0.0038		-0.0013		-0.0150		-0.0081
		(0.0045)		(0.0029)		(0.0086)		(0.0076)
average parents income	0.000002***	0.000002***	0.000002***	0.000002***	0.000005***	0.000005***	0.000005***	0.000005***
	(0.0000004)	(0.0000004)	(0.0000004)	(0.0000004)	(0.0000013)	(0.0000013)	(0.0000013)	(0.0000013)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.1444	0.1452	0.3298	0.3309	0.2286	0.2311	0.1580	0.1592
Observations	1,647	1,647	1,647	1,647	1,647	1,647	1,647	1,647
Fixed effect								
Number of all nearby school	0.0168		0.0128		0.0398		0.0772*	
	(0.0114)		(0.0105)		(0.0321)		(0.0402)	
Number of nearby public school		0.1157***		0.1295***		0.0477		0.1072***
		(0.0116)		(0.0102)		(0.0317)		(0.0397)
Number of nearby Catholic school		0.0166		0.0059		0.0399		0.0772*
		(0.0114)		(0.0105)		(0.0322)		(0.0402)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,647	1,647	1,647	1,647	1,647	1,647	1,647	1,647
Restrict data to Peel, Halton, Toronto, York and Ottawa								
Number of all nearby school	0.0012		0.0021		0.0043		0.0072	
	(0.0025)		(0.0021)		(0.0065)		(0.0070)	
Number of nearby public school		0.0034		0.0036		0.0098		0.0072
		(0.0030)		(0.0022)		(0.0071)		(0.0094)
Number of nearby Catholic school		-0.0034		0.0071		0.0033		-0.0069
		(0.0043)		(0.0040)		(0.0118)		(0.0120)
average parents income	0.000002***	0.000002***	0.000002***	0.000002***	0.000004***	0.000004***	0.000002***	0.000002***
	(0.0000006)	(0.0000006)	(0.0000004)	(0.0000004)	(0.0000015)	(0.0000015)	(0.000002)	(0.000002)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.1930	0.1953	0.4188	0.4393	0.3176	0.3402	0.2224	0.2224
Observations	890	890	890	890	890	890	890	890
Only TDSB								
Number of all nearby school	0.0047*		0.0058*		0.0059*		0.0058*	
	(0.0031)		(0.0034)		(0.0032)		(0.0035)	
Number of nearby public school		0.0067*		0.0069*		0.0086*		0.060*
		(0.0040)		(0.0036)		(0.0037)		(0.0034)
Number of nearby Catholic school		0.0024		-0.0081		-0.0032		-0.0008
		(0.0043)		(0.0060)		(0.0155)		(0.0160)
average parents income	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***
	(0.0000001)	(0.0000001)	(0.0000005)	(0.0000005)	(0.0000013)	(0.0000013)	(0.000002)	(0.000002)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.1757	0.1757	0.4815	0.5189	0.4135	0.4413	0.1803	0.1841
Observations	345	345	345	345	345	345	345	345

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics.

Table 3.9: OLS estimates based on school level EQAO data (2004 - 2013)

	OSSLT passed PE (1)	OSSLT passed FTE (2)	Grade 9 level 3 math AC (3)	Grade 9 level 3 math AP (4)
OLS with year fixed effect				
Number of nearby school	-0.0037^{***}	-0.0018	-0.0047^{***}	-0.0031^{**}
	(0.0016)	(0.0013)	(0.0015)	(0.0014)
Nearby school quality	0.0068^{***}	0.0207^{*}	0.0439^{***}	0.0787^{***}
	(0.0172)	(0.0115)	(0.0193)	(0.0318)
Year fixed effect	Yes	Yes	Yes	Yes
R-squared	0.3523	0.5885	0.2191	0.2139
Observations	3,515	3,515	3515	3,515
Fixed effect				
Number of nearby school	0.0114^{*}	0.0261^{***}	0.0209^{***}	0.0109^{*}
	(0.0065)	(0.0046)	(0.0058)	(0.0064)
Nearby school quality	0.3197^{***}	0.1799^{***}	0.2984^{***}	0.1862^{***}
	(0.0228)	(0.0154)	(0.0538)	(0.0335)
Year fixed effect	Yes	Yes	Yes	Yes
R-squared	0.6806	0.7535	0.5687	0.5203
Observations	3,515	3,515	3515	3,515
Restrict data to Peel, Halton, Toronto, York and Ottawa				
Number of nearby school	0.0010	0.00015	0.0013	0.0013
	(0.0018)	(0.0016)	(0.0020)	(0.0019)
Nearby school quality	0.0669^{***}	0.0340^{*}	0.1058^{**}	0.1565^{***}
	(0.0271)	(0.0185)	(0.0527)	(0.0549)
School board school	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
R-squared	0.4407	0.5441	0.3499	0.2315
Observations	1,700	1,700	1,700	1,700
Only TDSB				
Number of nearby school	0.0046^{*}	0.0061^{**}	0.0063^{**}	0.0049^{*}
	(0.0025)	(0.0027)	(0.0029)	(0.0028)
Nearby school quality	0.1274^{***}	0.0376^{*}	0.2163^{*}	0.1861^{*}
	(0.0401)	(0.0280)	(0.1236)	(0.1017)
Year fixed effect	Yes	Yes	Yes	Yes
R-squared	0.4837	0.5225	0.3543	0.1927
Observations	750	750	750	750

Note: Standard errors are in parenthesis and clustered at the school level for OLS estimates. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics.

Table 3.10: Average school performance and number of observations, by number of nearby school

	0	1	2	3	4	5	5 to 10	10 to 20	more than 20
Average math-AC	0.7407	0.7612	0.7225	0.7411	0.7453	0.7896	0.7422	0.7398	0.6572
Average math-AP	0.4711	0.4622	0.4481	0.4241	0.4524	0.4388	0.4338	0.4204	0.4016
Average OSSLT FTE	0.7218	0.7245	0.7310	0.7416	0.7504	0.7824	0.7363	0.7500	0.6967
Average OSSLT PE	0.6050	0.5895	0.6236	0.6310	0.6405	0.6333	0.6028	0.5846	0.5357
Number of obs	439	287	324	276	321	266	820	630	211

Author's own calculation .

Table 3.11: OLS estimates of different numbers of nearby school on school performance based on school level EQAO data (2004 - 2013)

	OSSLT passed PE	OSSLT passed FTE	Grade 9 level 3 math AC	Grade 9 level 3 math AP
	(1)	(2)	(3)	(4)
0	0.0637	0.0653	0.0899	0.0676
	(0.0463)	(0.0484)	(0.0565)	(0.0417)
1	0.0562	0.0448	0.0595	0.0180
	(0.0462)	(0.0474)	(0.0439)	(0.0388)
2	0.0740	0.0410	0.0108	-0.0009
	(0.0477)	(0.0476)	(0.0435)	(0.0388)
3	0.0677	0.0394	0.0235	-0.0158
	(0.0468)	(0.0477)	(0.0424)	(0.0379)
4	0.0507	0.0371	0.0170	-0.0087
	(0.0447)	(0.0470)	(0.0429)	(0.0372)
5	0.0472	0.0469	0.0454	-0.0214
	(0.0449)	(0.0453)	(0.0410)	(0.0368)
5-10	0.0385	0.0238	0.0209	-0.0009
	(0.0416)	(0.0442)	(0.0401)	(0.0339)
10-20	0.0549	0.0438	0.0390	0.0130
	(0.0401)	(0.0426)	(0.0378)	(0.0322)
school quality	0.0848***	0.0328***	0.1966***	0.1790***
	(0.0197)	(0.0131)	(0.0514)	(0.0347)
School characteristics	Yes	Yes	Yes	Yes
Neighborhood characteristics	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
School board fixed effect	Yes	Yes	Yes	Yes
R-squared	0.3865	0.4051	0.2971	0.2943
Observations	3,515	3,515	3,515	3,515

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics. The omitted category is “more than 20”.

Table 3.12: First stage statistics of instrumental estimates based on school level EQAO data (2004 - 2013)

First stage statistics	
Number of nearby schools	
ln (distance)	0.4651***
	(0.1807)
R^2	0.6380
Adjusted R^2	0.6357
partial R^2	0.0158
Under-identification test	
Anderson canon. corr. LM statistic	55.677
P value	0.0000
Weak identification test	
Cragg-Donald Wald F statistic	56.203
Stock-Yogo weak ID test critical values: 10% maximal IV size	16.38
15% maximal IV size	8.96
20% maximal IV size	6.66
25% maximal IV size	5.53

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table 3.13: Instrumental estimates based on school level EQAO data (2004 - 2013)

	OSSLT passed PE (1)	OSSLT passed FTE (2)	Grade 9 level 3 math AC (3)	Grade 9 level 3 math AP (4)
IV estimates				
Number of nearby school	0.0064	0.0050	0.0096	0.0183
	(0.0111)	(0.0086)	(0.0091)	(0.0112)
Year fixed effect	Yes	Yes	Yes	Yes
Observations	3,515	3,515	3,515	3,515
R^2	0.3391	0.5871	0.2068	0.2094
IV estimates with data restricted to Peel, Halton, Toronto, York and Ottawa				
Number of nearby school	0.0126*	0.0090	0.0105*	0.00089
	(0.0072)	(0.0066)	(0.0060)	(0.0069)
Year fixed effect	Yes	Yes	Yes	Yes
Observations	1,700	1,700	1,700	1,700
R^2	0.3466	0.5077	0.2557	0.1991
IV estimates with only TDSB				
Number of nearby school	0.0132***	0.0229***	0.0198***	0.0187***
	(0.0059)	(0.0074)	(0.0060)	(0.0071)
Year fixed effect	Yes	Yes	Yes	Yes
Observations	750	750	750	750
R^2	0.3603	0.2759	0.2528	0.1800

Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

3.9 Figures

Figure 3.1: Enrollment of secondary schools from 2000 to 2012, by school type

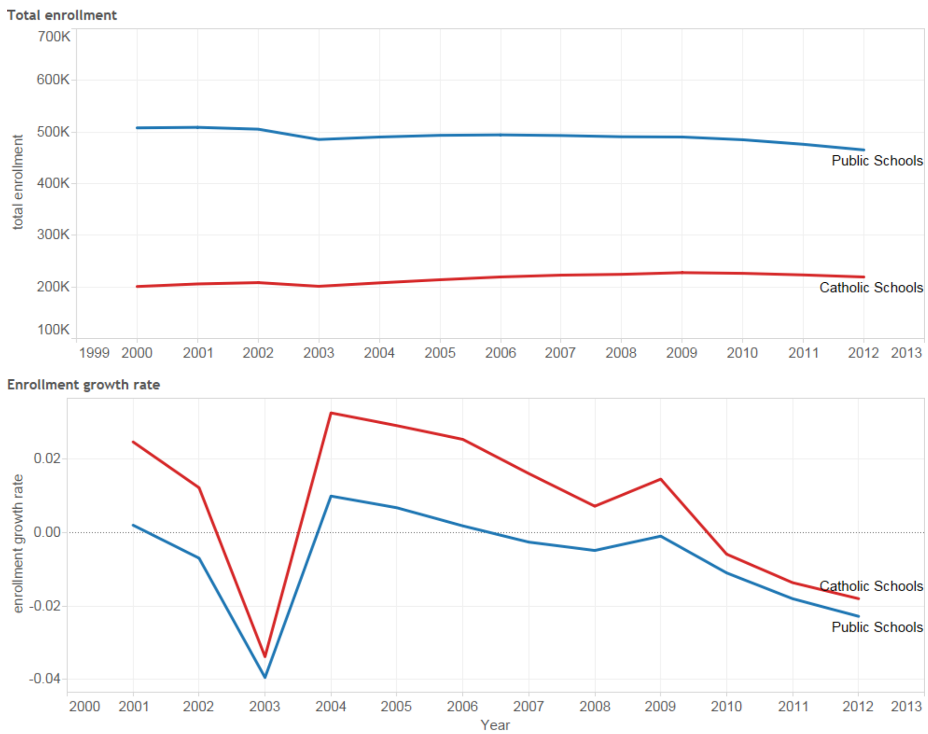
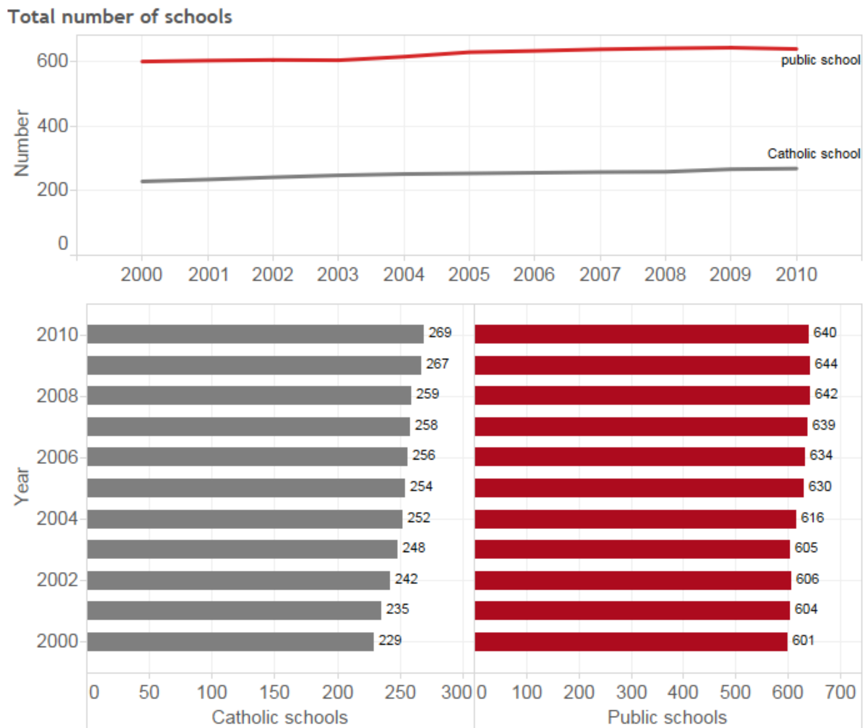
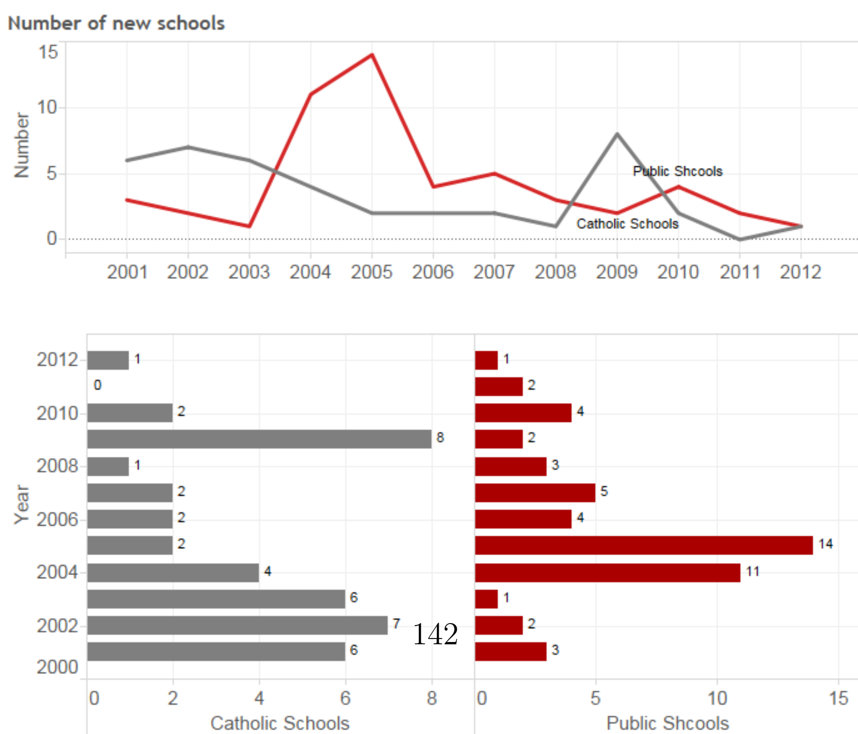


Figure 3.2: Total number of secondary schools and number of new secondary schools opened from 2000 to 2012, by school type



(a) Total number of secondary schools



(b) Number of new secondary schools

Figure 3.3: Schools within a circle of 2 kilometers of Gary Allan High School - Burlington

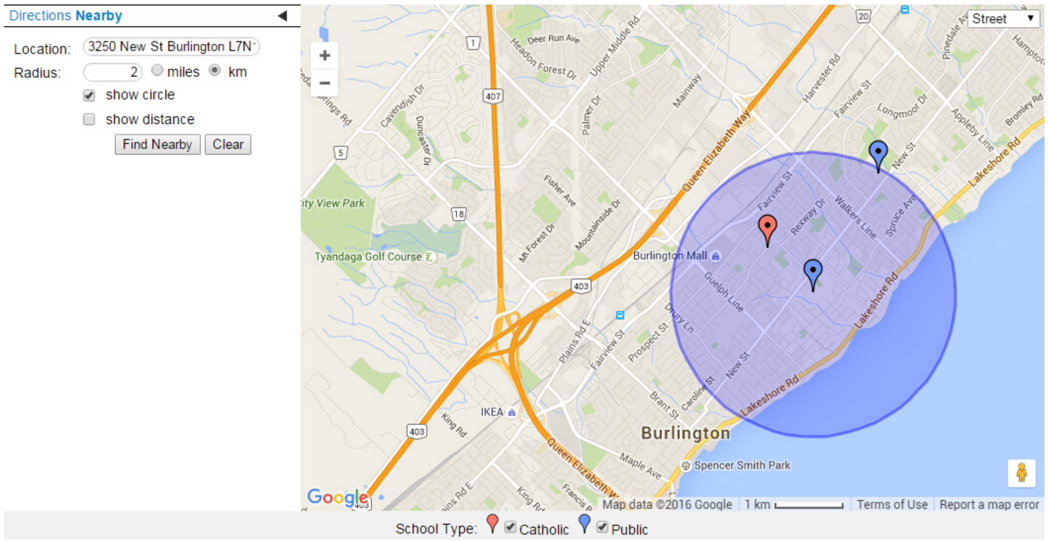


Figure 3.4: The distribution of Ontario publicly funded secondary schools

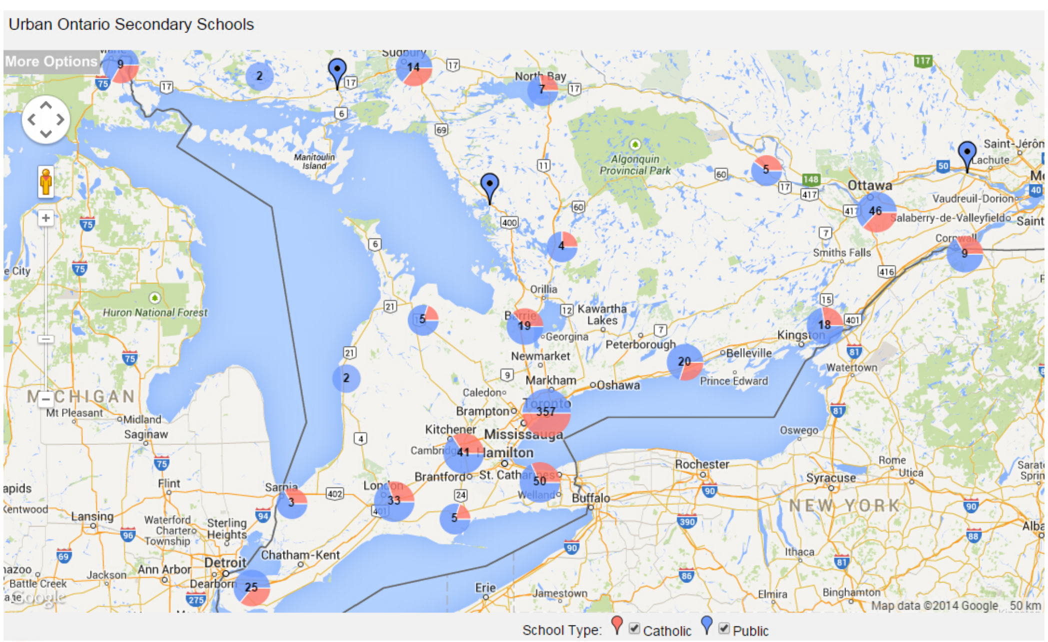
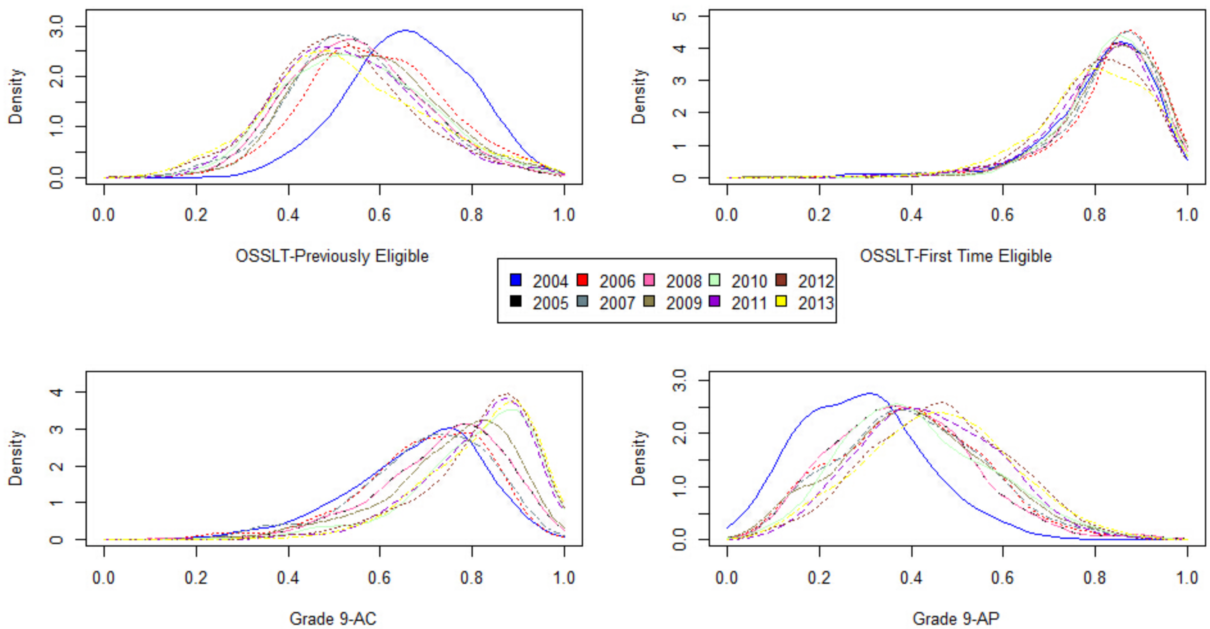


Figure 3.5: The kernel density plot of students passing provincial standard in different EQAO assessment from 2004 to 2013



Note: "AP" and "AC" refer to the terms Grade 9 assessment of academic and applied Mathematics, respectively. "FTE" and "PE" refer to "first time eligible" and "previously eligible".

Figure 3.6: Average distance of families to schools other than the closest one

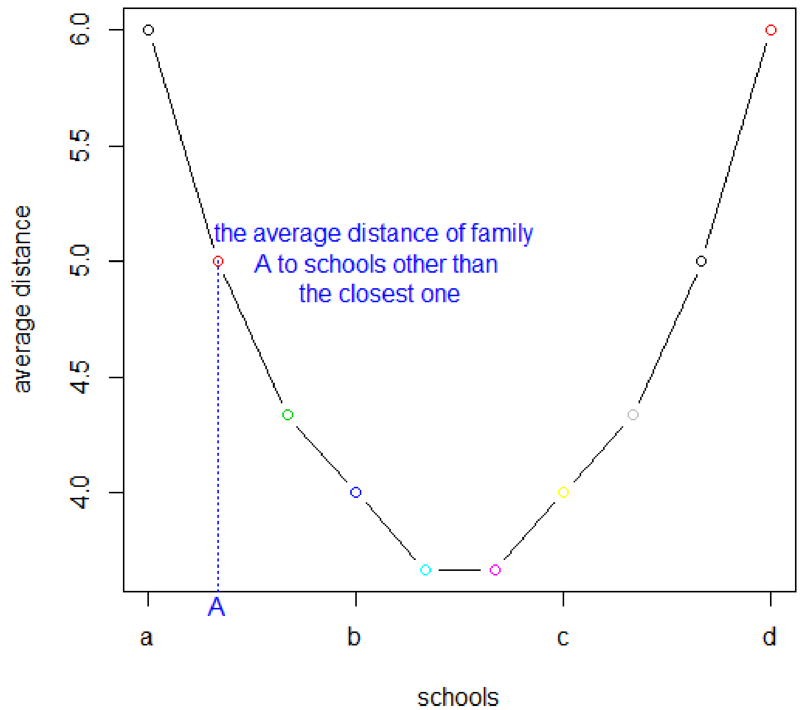
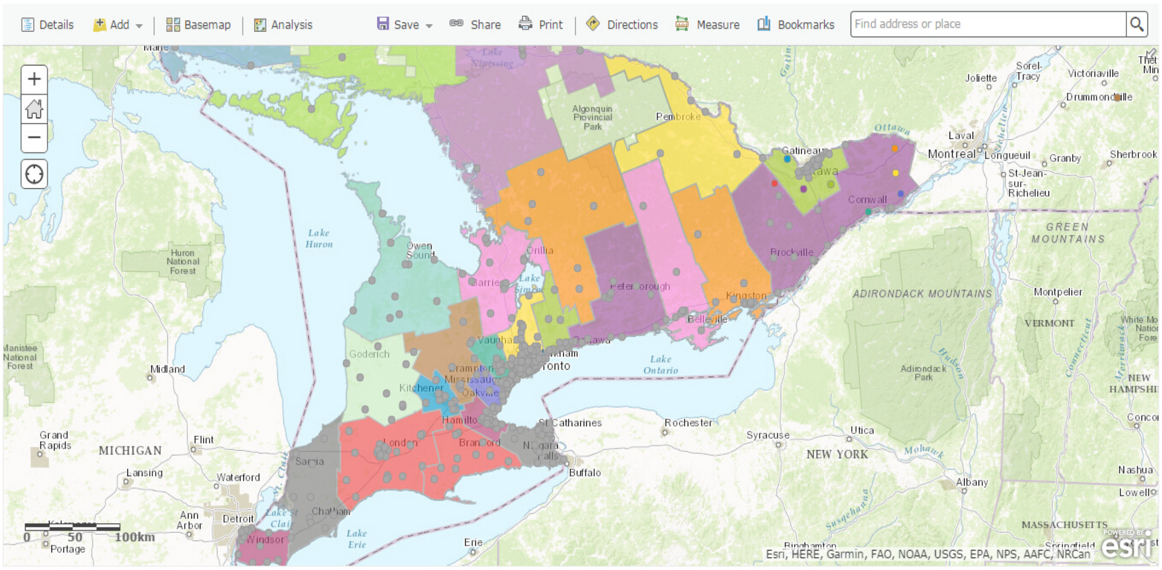
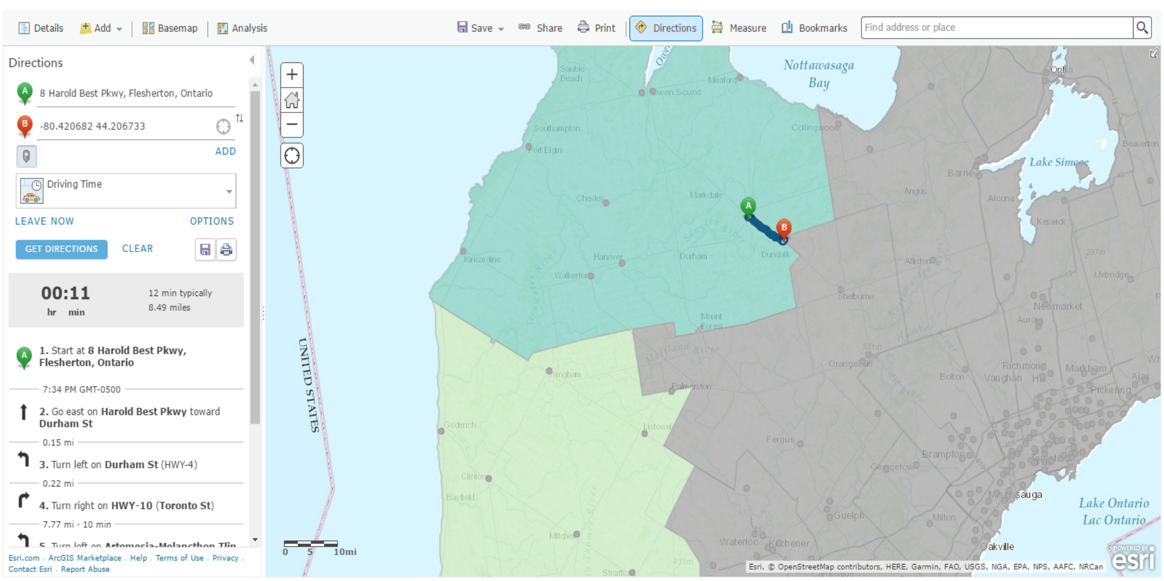


Figure 3.7: School board boundary



Note: colors identify different school boards.

Figure 3.8: An example demonstrating how to identify the distance from a given school to school board boundary.



Bibliography

Joshua D Angrist and Alan B Keueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1):343–375, 2004.

Yuval Bar-Or, John Burbidge, Lonnie Magee, and A Leslie Robb. The wage premium to a university education in canada, 1971-1991. *Journal of Labor Economics*, 13(4):762–794, 1995.

Andy H Barnett, Richard W Ault, and David L Kaserman. The rising incidence of co-authorship in economics: Further evidence. *The review of Economics and Statistics*, pages 539–543, 1988.

Paul Beaudry and David Green. What is driving us and canadian wages: exogenous technical change or endogenous choice of technique? Technical report, National bureau of economic research, 1998.

Eric P Bettinger. The effect of charter schools on charter students and public schools. *Economics of Education Review*, 24(2):133–147, 2005.

Anders Böhlmark and Mikael Lindahl. Independent schools and long-run educational outcomes: Evidence from sweden's large-scale voucher reform. *Economica*, 2015.

Daniel Boothby and Torben Drewes. Postsecondary education in canada: Returns to university, college and trades education. *Canadian Public Policy/Analyse de politiques*, pages 1–21, 2006.

George J Borjas and Kirk B Doran. The collapse of the soviet union and the productivity of american mathematicians. *The Quarterly Journal of Economics*, 127(3):1143–1203, 2012.

Brahim Boudarbat, Thomas Lemieux, and W Craig Riddell. Recent trends in wage inequality and the wage structure in canada. *Dimensions of inequality in Canada*, pages 273–306, 2006.

Brahim Boudarbat, Thomas Lemieux, and W Craig Riddell. The evolution of the returns to human capital in canada, 1980-2005. *Canadian Public Policy*, 36(1):63–89, 2010.

Ivy E Broder. Professional achievements and gender differences among academic economists. *Economic Inquiry*, 31(1):116–127, 1993.

John B Burbidge, Lonnie Magee, and A Leslie Robb. The education premium in canada and the united states. *Canadian Public Policy/Analyse de politiques*, pages 203–217, 2002.

David Card. Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research, 1993.

David Card. The causal effect of education on earnings. *Handbook of labor economics*, 3: 1801–1863, 1999.

David Card, Martin D Dooley, and A Abigail Payne. School competition and efficiency with publicly funded catholic schools. *American Economic Journal: Applied Economics*, 2(4):150–176, 2010.

Matthew Carr and Gary Ritter. Measuring the competitive effect of charter schools on student achievement in ohios traditional public schools. *National Center for the Study of Privatization in Education (Columbia University) Research Paper*, 146, 2007.

Pamela Christie and Michael Shannon. Educational attainment and the gender wage gap: evidence from the 1986 and 1991 canadian censuses. *Economics of Education Review*, 20(2):165–180, 2001.

John P Conley and Ali Sina Önder. The research productivity of new phds in economics: the surprisingly high non-success of the successful. *The Journal of Economic Perspectives*, pages 205–215, 2014.

John P Conley, Mario J Crucini, Robert A Driskill, and Ali Sina Önder. The effects of publication lags on life-cycle research productivity in economics. *Economic Inquiry*, 51 (2):1251–1276, 2013.

Tom Coupé, Valérie Smeets, and Frédéric Warzynski. Incentives, sorting and productivity along the career: Evidence from a sample of top economists. *Journal of Law, Economics, and Organization*, 22(1):137–167, 2005.

James B Davies, Martin G Kocher, and Matthias Sutter. Economics research in canada: a long-run assessment of journal publications. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(1):22–45, 2008.

Alison F Del Rossi and Joni Hersch. Double your major, double your return? *Economics of Education Review*, 27(4):375–386, 2008.

Marie Drolet. The persistent gap: New evidence on the canadian gender wage gap. 2001.

Lorenzo Ductor. Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385–407, 2015.

Marcel Fafchamps, Marco J Van der Leij, and Sanjeev Goyal. Scientific networks and co-authorship. *University of Oxford Economics Discussion Paper*, (256), 2006.

Marianne A Ferber and Michelle Teiman. Are women economists at a disadvantage in publishing journal articles? *Eastern Economic Journal*, 6(3/4):189–193, 1980.

Ana Ferrer and W Craig Riddell. Education, credentials, and immigrant earnings. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(1):186–216, 2008.

Ana M Ferrer and W Craig Riddell. The role of credentials in the canadian labour market. *Canadian Journal of Economics/Revue canadienne d'économique*, 35(4):879–905, 2002.

Kelly Foley and David A Green. Why more education will not solve rising inequality (and may make it worse). *Income Inequality: The Canadian Story. Montreal: Institute for Research on Public Policy*, 2016.

Nicole M Fortin and Thomas Lemieux. Changes in wage inequality in canada: An interprovincial perspective. *Canadian Journal of Economics/Revue canadienne d'économique*, 48(2):682–713, 2015.

Mary Frank Fox and Catherine A Faver. Independence and cooperation in research: The motivations and costs of collaboration. *The Journal of Higher Education*, pages 347–359, 1984.

Richard B Freeman and Wei Huang. Collaborating with people like me: Ethnic co-authorship within the us. Technical report, National Bureau of Economic Research, 2014.

Richard B Freeman and Karen Needels. Skill differentials in canada in an era of rising labor market inequality. Technical report, National Bureau of Economic Research, 1991.

Marc Frenette and Simon Coulombe. *Has higher education among young women substantially reduced the gender gap in employment and earnings?* Business and Labour Market Analysis, Statistics Canada, 2007.

Marc Frenette and René Morissette. Wages and full-time employment rates of young high school graduates and bachelor's degree holders, 1997 to 2012. *Statistics Canada Analytical Studies Branch Research Paper*, (360), 2014.

Jane Friesen, Benjamin Harris, and Simon D Woodcock. Expanding school choice through open enrolment: Lessons from british columbia. *CD Howe Institute Commentary*, (418), 2015.

Stephen Gibbons, Stephen Machin, and Olmo Silva. Choice, competition, and pupil achievement. *Journal of the European Economic Association*, 6(4):912–947, 2008.

John W Graham and Steven A Smith. Gender differences in employment and earnings in science and engineering in the us. *Economics of education review*, 24(3):341–354, 2005.

Morley Gunderson. Earnings differentials between the public and private sectors. *Canadian Journal of Economics*, pages 228–242, 1979.

Daniel S Hamermesh, George E Johnson, and Burton A Weisbrod. Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, pages 472–481, 1982.

W Lee Hansen, Burton A Weisbrod, and Robert P Strauss. Modeling the earnings and research productivity of academic economists. *The Journal of Political Economy*, pages 729–741, 1978.

Eric A Hanushek. The failure of input-based schooling policies*. *The economic journal*, 113(485):F64–F98, 2003.

Zi-Lin He, Xue-Song Geng, and Colin Campbell-Hunt. Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a new zealand university. *Research policy*, 38(2):306–317, 2009.

Christiana E Hilmer and Michael J Hilmer. How do journal quality, co-authorship, and author order affect agricultural economists' salaries? *American Journal of Agricultural Economics*, 87(2):509–523, 2005.

Michael J Hilmer, Michael R Ransom, and Christiana E Hilmer. Fame and the fortune of academic economists: How the market rewards influential research in economics. *Southern Economic Journal*, 82(2):430–452, 2015.

Aidan Hollis. Co-authorship and the output of academic economists. *Labour Economics*, 8(4):503–530, 2001.

Caroline M Hoxby. Does competition among public schools benefit students and taxpayers? *American Economic Review*, pages 1209–1238, 2000.

John Hudson. Trends in multi-authored papers in economics. *The Journal of Economic Perspectives*, 10(3):153–158, 1996.

Scott A Imberman. The effect of charter schools on achievement and behavior of public school students. *Journal of Public Economics*, 95(7):850–863, 2011.

David A Jaeger and Marianne E Page. Degrees matter: New evidence on sheepskin effects in the returns to education. *The review of economics and statistics*, pages 733–740, 1996.

Christopher Jepsen, Kenneth Troske, and Paul Coomes. The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics*, 32 (1):95–121, 2014.

J Sylvan Katz and Ben R Martin. What is research collaboration? *Research policy*, 26(1): 1–18, 1997.

Yolanda K Kodrzycki and Pingkang Yu. New approaches to ranking economics journals. *The BE Journal of Economic Analysis & Policy*, 5(1), 2006.

David N Laband and Robert D Tollison. Intellectual collaboration. *Journal of Political economy*, 108(3):632–662, 2000.

Thomas Lemieux. Occupations, fields of study and returns to education. *Canadian Journal of Economics/Revue canadienne d'économique*, 47(4):1047–1077, 2014.

Thomas Lemieux and David Card. Education, earnings, and the canadian gi bill. *Canadian Journal of Economics/Revue canadienne d'économique*, 34(2):313–344, 2001.

Philip SJ Leonard. Choice of ontario high schools and its impact on university applications. *Education Economics*, 23(4):433–454, 2015.

Christopher Lubienski. Public schools in marketized environments: Shifting incentives and unintended consequences of competition-based educational reforms. *American Journal of Education*, 111(4):464–486, 2005.

Robert F Lucas. Contributions to economics journals by the canadian economics profession, 1981-90. *Canadian Journal of Economics*, pages 945–960, 1995.

Kellie L Maske, Garey C Durden, and Patricia E Gaynor. Determinants of scholarly productivity among male and female economists. *Economic inquiry*, 41(4):555–564, 2003.

John M McDowell and Janet Kiholm Smith. The effect of gender-sorting on propensity to coauthor: Implications for academic promotion. *Economic Inquiry*, 30(1):68–82, 1992.

Robert McMillan. Erratum to competition, incentives, and public school productivity. *Journal of Public Economics*, 89(5):1133–1154, 2005.

Marshall H Medoff. Collaboration and the quality of economics research. *Labour Economics*, 10(5):597–608, 2003.

Kevin Milligan, Enrico Moretti, and Philip Oreopoulos. Does education improve citizenship? evidence from the united states and the united kingdom. *Journal of public Economics*, 88(9):1667–1695, 2004.

Daniel L Millimet and Vasudha Rangaprasad. Strategic competition amongst public schools. *Regional Science and Urban Economics*, 37(2):199–219, 2007.

Richard E Mueller. Public-and private-sector wage differentials in canada revisited. *Industrial Relations: A Journal of Economy and Society*, 39(3):375–400, 2000.

Yongmei Ni. The impact of charter schools on the efficiency of traditional public schools: Evidence from michigan. *Economics of Education Review*, 28(5):571–584, 2009.

Joëlle Noailly, Sunčica Vujić, and Ali Aouragh. The effects of competition on the quality of primary schools in the netherlands. *Environment and Planning A*, 44(9):2153–2170, 2012.

Philip Oreopoulos and Kjell G Salvanes. Priceless: The nonpecuniary benefits of schooling. *The journal of economic perspectives*, 25(1):159–184, 2011.

James Otto, Douglas A Sanford Jr, and Douglas N Ross. Does ratemyprofessor. com really rate my professor? *Assessment & Evaluation in Higher Education*, 33(4):355–368, 2008.

Jin Heum Park. Estimation of sheepskin effects using the old and the new measures of educational attainment in the current population survey. *Economics Letters*, 62(2):237–240, 1999.

Michael J Piette and Kevin L Ross. An analysis of the determinants of co-authorship in economics. *The Journal of Economic Education*, 23(3):277–283, 1992.

David Prescott and Bo Wandschneider. Public/private sector wage differentials in canada-evidence from the 1991 and 1982 surveys of consumer finance. *Applied Economics*, 31 (6):723–731, 1999.

Russell W Rumberger and Scott L Thomas. The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review*, 12(1):1–19, 1993.

Heather Sarsons. Gender differences in recognition for group work. *Harvard University*, 3, 2015.

Tim R Sass. Charter schools and student achievement in florida. *Education*, 1(1):91–122, 2006.

Raymond D Sauer. Estimates of the returns to quality and coauthorship in economic academia. *The Journal of Political Economy*, pages 855–866, 1988.

Tammy Schirle. The gender wage gap in the canadian provinces, 1997-2014. *Canadian Public Policy*, 41(4):309–319, 2015.

Anindya Sen, Hideki Ariizumi, and Natasha DeSousa. Evaluating the relationship between pay and research productivity: Panel data evidence from ontario universities. *Canadian Public Policy*, 40(1):1–14, 2014.

Daniel M Shapiro and Morton Stelcner. Canadian public-private sector earnings differentials, 1970–1980. *Industrial Relations: A Journal of Economy and Society*, 28(1):72–81, 1989.

Wayne Simpson and JC Herbert Emery. Canadian economics in decline: Implications for canada's economics journals. *Canadian Public Policy*, 38(4):445–470, 2012.

Seymour Spilerman and Tormod Lunde. Features of educational attainment and job promotion prospects. *American Journal of Sociology*, 97(3):689–720, 1991.

Samuel R Staley and John P Blair. Institutions, quality competition and public service provision: The case of public education. *Constitutional Political Economy*, 6(1):21–33, 1995.

Benjamin Tal and Emanuella Enenajor. Degrees of success: The payoff to higher education in canada. *CIBC World Markets*, 2013.

Raaj Tiagi. Public sector wage premium in canada: evidence from labour force survey. *Labour*, 24(4):456–473, 2010.

Douglas A Webber. The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour Economics*, 28: 14–23, 2014.

Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Nelson Education, 2015.

Ron Zimmer and Richard Buddin. Is charter school competition in california improving the performance of traditional public schools? *Public Administration Review*, 69(5): 831–845, 2009.

Appendix A

APPENDICES

Appendices of Chapter 1

The distribution of employees by degree and government organization

Table A.1: The distribution of employees by degree and government organization

		degree				
Sector	Measure	College	BA	BA + Diploma Double BA	Master	PhD
Crown agencies	percent	410	2106	14	859	223
	number of obs	11.35%	58.31%	0.39%	23.78%	6.17%
Hospitals & Boards of Public Health	percent	320	1661	233	2515	963
	number of obs	5.62%	29.18%	4.09%	44.18%	16.92%
Hydro	percent	1714	3630	435	1285	209
	number of obs	23.57%	49.91%	5.98%	17.67%	2.87%
Ministries	percent	387	2943	353	1827	454
	number of obs	6.49%	49.35%	5.92%	30.63%	7.61%
Municipalities and Services	percent	1744	3778	622	1837	127
	number of obs	21.51%	46.60%	7.67%	22.66%	1.57%
School Boards	percent	78	1050	1179	1997	117
	number of obs	1.76%	23.75%	26.67%	45.17%	2.65%

Source: authors' own calculation.

Ontario Public Sectors

Table A.2: Ontario Public Sectors

Sectors	Organizations (sub-sectors)
Ministries	Aboriginal Affairs, Agriculture, Food and Rural Affairs, Attorney General, Children and Youth Services, Citizenship, Immigration and International Trade, Community and Social Services, Community Safety and Correctional Services, Economic Development, Employment and Infrastructure, Education, Energy, Environment and Climate Change, Finance, Government and Consumer Services, Francophone Affairs, Health and Long-Term Care, Intergovernmental Affairs, Labor, Municipal Affairs and Housing, Natural Resources and Forestry, Northern Development and Mines, Pan/Parapan American Games, Secretariat, Research and Innovation, Seniors' Secretariat, Tourism, Culture and Sport, Training, Colleges and Universities, Transportation, Treasury Board Secretariat.
Legislative Assembly and Offices	French Language Services Commissioner, Information & Privacy Commissioner Environmental Commissioner, Integrity Commissioner, Legislative Assembly, Office of the Auditor General, Lieutenant Governor, Ombudsman Ontario, Office of the Chief Electoral Officer, Provincial Advocate.
Judiciary	Ontario Court of Justice
Crown Agencies	Advertising Review Board, Consent & Capacity Board, Health Board Secretariat, Child & Family Services Review Board, Alcohol & Gaming Commission of Ontario , AgriCorp, Education Quality & Accountability Office, Grievance Settlement Board eHealth Ontario, Cancer Care Ontario, Fire Marshal's Public Fire Safety Council, Criminal Injuries Compensation Board, Deposit Insurance Corporation of Ontario Financial Services Commission, Health Force Ontario Marketing & Recruitment, Higher Education Quality Council of Ontario, Human Rights Legal Support Centre, Landlord & Tenant Board, Legal Aid Ontario, Liquor Control Board of Ontario, Niagara Parks Commission, Metropolitan Toronto Convention Centre Corporation, Metrolinx, Local Health Integration Network, Office of the Worker Adviser, The Ontario French Language Educational Communications Authority, Office of the Employer Adviser, Office of the Conflict of Interest Commissioner Office of the Independent Police Review Director, Ontario Clean Water Agency, Ontario Agency for Health Protection & Promotion, Ontario Energy Board, Ontario Educational Communications Authority, Ontario Highway Transport Board, Ontario Heritage Trust ,etc.
Hydro One and Ontario Power Generation	Hydro One, Hydro One Brampton Networks Inc., Ontario Power Generation.
Municipalities and services	City of Barrie, City of Brampton, City of Cambridge, City of Greater Sudbury, City of Dryden, etc.
Hospitals and Boards of Public Health	Alexandra Hospital, Almonte General Hospital, Atikokan General Hospital, Baycrest Centre for Geriatric Care, Brant Community Healthcare System, Board of Health for the Northwestern Health Unit, Bridgepoint Hospital, Bluewater Health, Board of Health for Simcoe Muskoka District Health Unit, etc.
School Boards	Algoma District School Board, Avon Maitland District School Board, Brant Haldimand Norfolk Catholic District School Board, etc.
Colleges	Algonquin College, Centennial College, Canadore College, Conestoga College, Confederation College, Fanshawe College, etc.
Universities	Algoma University, Lakehead University, Laurentian University of Sudbury, McMaster University, Brock University, etc.

Data source: Ministry of finance website

Descriptive statistics of job rank

Table A.3: Descriptive statistics of job rank

job rank	NOC code	Title
Senior manager	0011-0016	**Legislators:** cabinet minister, city councilor, First Nations band chief, governor general, lieutenant-governor, mayor, premier, premier minister, school board trustee, senator **Senior govt manager:** assistant deputy minister, chairperson, Human Rights, Commission, chief statistician-government services, deputy minister, city administrator, director general, executive director, high commissioner **Senior managers in other organizations in public sectors:** CEO, CFO, general manager, president, vice-president, executive director, corporate controller
Middle manager	0111-0912	all the occupations labeled with a four-digit NOC code 0111 to 0912 **managers in public administration:** government manager, government director, clerk of the committee-Legislative Assembly, etc. **university and college:** dean-college, dean-university, dean-faculty of science, registrar, manager-business school, student activities dean, etc. **school boards**: administrator- board of education, chief superintendent director of school, school principle, vice principle, superintendent **hospital:** manager, director of nursing, chief of medical staff, etc. **managers in other organizations in Ontario public sector**
others		others

Source: Statistics Canada.

Two sample t test: with VS without LinkedIn, by sector

Table A.4: Two sample t test: with VS without LinkedIn, by sector

	average nominal salary				
Sector	With	Without	Diff	t	p-value
Crown agencies	141,166.2	136,221	4,945.12	5.46	0.0000
	(755.27)	(499.43)			
Hospitals and Boards of Public Health	147,945	143,349.2	4,595.80	4.54	0.0000
	(901.62)	(418.15)			
Hydro	138,630.8	125,771.7	12,859.06	23.49	0.0000
	(456.28)	(324.47)			
Ministries	125,790.4	124,541.2	1,249.25	3.26	0.0011
	(358.92)	(149.20)			
Municipalities and Services	122,399.1	115,353.4	7,045.69	36.51	0.0000
	(233.93)	(63.98)			
School Boards	117,340.9	113,837	3,503.78	15.07	0.0000
	(262.92)	(72.21)			

Source: authors' own calculation. Standard errors are in parenthesis. diff = mean(with) - mean(without). H_0: diff = 0

Data collecting and cleaning process

The Sunshine List is drawn directly from the official database of the Ministry of Finance website on a year-to-year basis and the raw data contains around 652,804 records over 1996 to 2013. The Sunshine List data presented at the Ministry of Finance website is generally divided into 10 sectors: college, crown agencies, hospitals, hydro one and Ontario power generation, judiciary, legislative assembly, ministries, municipalities, school boards and university. For each record, the following attributes are given: employer, surname, given name, position, salary paid, and taxable benefits. We directly collect the data from the Ministry of Finance website.

Once the samples were extracted, it was first necessary to organize it in a consistent way since there are several changes occurred during the disclosure process from year to year. First of all, some years the employer is presented in full name and some years in short name. For example, "George Brown College" is reported as "George Brown College of Applied Arts and Technology" in one year and then "George Brown" in another year. To maintain the consistency, we change all employer name in a full name format. For example, as long as there are characters as "George Brown" are identified, we replace it with "George Brown College of Applied Arts and Technology". Secondly, there is some inconsistency of surnames/given names. Some years they reported their professional designations, such as "Dr." and "Hon.", before given name or at the end of surname. We manually detect this issue and delete the professional designations. Thirdly, the position name is also experienced some inconsistency over years. Some years the position name is presented in full name and some years in an abbreviated version. Similar as the way we deal with the inconsistency of employer name, we match and organize all position in the full name format. For instance, "Ass't. Deputy Minister" is detected and replaced by "Assistant Deputy Minister".

As a second step, we collect information about the education and experience of those government employees by extracting their information from LinkedIn website manually. We keep only those individuals for whom we can clearly identify their education background and careers from their LinkedIn profile. The data on each individual's education background consist of the following information: the time period and the field of study of each rewarded degree above high school (information on HS and below is not available for most individuals in our sample), the institution name rewarding each degree. We then categorize the educational background into three main different dimensions of educational achievement to examine the effect of education on earnings and career advancement: years of schooling, different educational credentials as well as field of study.

The actual labor market experience data is also extract from LinkedIn website which contains the name of employer, the position as well as the beginning and ending dates of each position. The measure of general experience Exp_i is calculated from the starting point of each individual's first job posted on LinkedIn profile.

We use National Occupational Classification (NOC) 2011 as the proxy for job rank in our analysis, we construct the following three different levels of the status rank hierarchy: senior manager, middle manager and others. As mentioned previously, the raw data we download from Ministry of Finance website include the position name to which he or she is having at each year from 2005 to 2013. We match the position names in our data to each hierarchy level by the following procedure: we first create a collection of job titles for each hierarchy level, we then do a grid search for each position name in our sample to find out which hierarchy level it belongs to. For those position names which are matched to "Other" category, it could because of the true match or it would have been in "senior manager" or "middle manager" category, but it ended up in "Other" category because some errors existed in our raw data, we ensure the accuracy by manually check them one by one.

In our analysis, Gender is mainly identified from an individual's LinkedIn profile image (about 45% of individuals). However, for those who are not uploading their profile pictures or it is hard to tell the gender from the profile image, we try to identify their gender through some free online tools such as GenderChecker, GenderGuess, etc. Those website tools have a large database of names which is derived from certain countries Census data, an user simply input a first name, their search engines will let you know the gender of that name. There are also specific gender-checking websites for Asian names and Indian Names. They are still a small fraction of names are hard to tell gender from, especially Asian names. We dropped those individuals [1].

To make the readers more clear about each variable we are using and how we collect it, more detailed information of the data collecting and organizing procedure is presented in the following Table:

[1] Just a few.

Table A.5: Detailed data collecting and organizing procedures

Variable	Source	Procedure
Salary	Ministry of Finance	1. extract individual name, salary, organization name, positions of those top-earners in Ontario public sector on a year-by-year basis. 2. merge the yearly data by individual name using Excel. 3. use the province-specific consumer price index to convert the salary data into constant 1992 dollars.
Individual characteristics	LinkedIn	1. search his or her names and the organization to which he or she is affiliated to on LinkedIn website. 2. with the purpose of increasing the precision of our searching results, double check the career category on LinkedIn profile to make sure he or she is the right person we are looking for. 3. download his or her educational information, career and profile images into excel. 4. do 1 to 3 for each individual listed in the salary data
A. Education	LinkedIn	1. organize each individual's educational background by university name, credential title, field of study, starting and ending date of each degree. 2. calculate years of schooling using $S_i = 12 + \sum_{j=1}^{n} s_j$. 3. categorize degrees into 5 dummies: 1=college, 2=BA, 3=MA, 4=PhD, 5=BA+diploma. 4. match the field of study into each category.
B. Experience	LinkedIn	1. find out the starting year of his or her first job. 2. calculate exp using Exp1 = Year - starting year of his or her first job. 3. calculate anther measure for exp using: Exp2 = Year - the year an individual obtained his/her highest degree . (where Year is the time variable in our data: 2005 - 2013.)
C. Rank dummy	Ministry of Finance	1. create a collection of job titles for each hierarchy. 2. do a grid search for each position name in our sample to find out which hierarchy level it belongs to. 3. ensure the accuracy by manually check those matched to "Other" category.
D.Gender	LinkedIn	1. identify gender from an individual's LinkedIn profile image. 2. if cannot be identified, use GenderChecker and other tools. 3. if still cannot be identified, dropped.

Source: Author's own calculation.

Appendices of Chapter 2

"Top-10" and "Top-21" journal publications

"Top-10" journal publications	**"Top-21" journals publications**
Journal of Political Economy	Journal of Political Economy
Econometrica	Econometrica
Quarterly Journal of Economics	Quarterly Journal of Economics
American Economic Review	American Economic Review
Review of Economic Studies	Review of Economic Studies
Journal of Monetary Economics	Journal of Monetary Economics
Journal of Economic Theory	Journal of Economic Theory
Journal of Econometrics	Journal of Econometrics
Journal of Economic Perspectives	Journal of Economic Perspectives
	Economic Journal
	Journal of Finance
	Journal of Financial Economics
	European Economic Review
	Journal of International Economics
	Journal of Public Economics
	Review of Economics and Statistics
	International Economic Review
	Journal of Money, Credit, and Banking
	Review of Financial Studies
	Journal of Economics
	Journal of Business and Economics Statistics.
	Canadian Journal of Economics

Source: Sen et al. (2014).

JEL subject code

A	General Economics, Handbooks and Teaching
B	History of Economic Thought, Methodology, and Heterodox Approaches
C	Mathematical and Quantitative Methods
D	Microeconomics
E	Macroeconomics and Monetary Economics
F	International Economics
G	Financial Economics
H	Public Economics
I	Health, Education, and Welfare
J	Labor and Demographic Economics
K	Law and Economics
L	Industrial Organization
M	Business Administration and Business Economics; Marketing; Accounting
N	Economic History
O	Economic Development, Technological Change, and Growth
P	Agricultural and Natural Resource Economics; Environmental and Ecological Economics
R	Urban, Rural, and Regional Economics
Y	Miscellaneous Categories
Z	Other Special Topics

Some other regression results

Table A.6: OLS estimates of the effects of co-authored publications in the "TOP-21" and "TOP-60" ranked journals with lagged dependent variable

	publications in top 21 journals		publications in top 60 journals	
	(1)	(2)	(3)	(4)
Proportion of co-authored publication	0.7116***	0.6319***	0.7041***	0.7054***
	(0.1856)	(0.1833)	(0.1559)	(0.1590)
Lagged publications	0.3747***	0.3217***	0.5867***	0.5560***
	(0.0993)	(0.0942)	(0.0925)	(0.1104)
Male	-0.5073	-0.5474	0.2846	0.2774
	(0.3459)	(0.3227)	(0.2561)	(0.2682)
Associate professor	0.2985	0.3457*	0.1013	0.0747
	(0.1877)	(0.1870)	(0.2844)	(0.3193)
Full professor	0.2660	0.2422	-0.1971	-0.2080
	(0.2321)	(0.2387)	(0.2864)	(0.3467)
Experience	−0.0740*	−0.0805**	-0.0327	-0.0372
	(0.0401)	(0.0386)	(0.0451)	(0.0493)
Experience squared	0.0018**	0.0019***	0.00091	0.0010
	(0.00091)	(0.00080)	(0.00099)	(0.0010)
PhD from US	-0.1023	-0.2402	0.2200	0.1440
	(0.2106)	(0.2238)	(0.2681)	(0.3112)
PhD from Canada	0.2350	0.1488	0.2090	0.2067
	(0.2585)	(0.2762)	(0.2524)	(0.3410)
Average rating on teaching	-0.1111	-0.1044	-0.1086	-0.0829
	(0.0694)	(0.6542)	(0.0670)	(0.0687)
SSHRC	-0.1574	-0.0769	−0.3128*	-0.2377
	(0.2211)	(0.2058)	(0.1728)	(0.2046)
University characteristics				
Medical schools	0.4567		0.5308	
	(0.5187)		(0.3732)	
Comprehensive universities	-0.5148		-0.4458	
	(0.3986)		(0.3393)	
Union	0.0702		0.2217	
	(0.1858)		(0.2403)	
Nocap	0.3995		-0.5351	
	(0.2367)		(0.4391)	
Year fixed effect	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes
R-squared	0.7034	0.7589	0.7663	0.7883
Observations	58	58		

Note: Dependent variable in all columns is the natural log of the total number of citations adjusted by year. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively.

Table A.7: The effect of different types of co-authorship on productivity: using another definition of types of co-authorship

	publications in top 10 journals		publications in top 21 journals		publications in other journals	
	(1)	(2)	(3)	(4)	(5)	(6)
With university colleagues	0.1301	0.1301	0.4161	0.2615	0.5228	0.7067**
	(0.2459)	(0.2459)	(0.2850)	(0.2423)	(0.3946)	(0.3595)
With other Canadian colleagues	0.8993	0.8993	0.1053	0.3531	-0.2536	-0.1944
	(0.6342)	(0.6342)	(0.3683)	(0.3215)	(0.4990)	(0.4827)
With US authors	0.9180**	0.9180**	0.6073*	0.4724	0.3469	0.2617
	(0.4441)	(0.4441)	(0.3796)	(0.4402)	(0.5403)	(0.5744)
Coauthors are from us and Canada			-0.8520	−1.7274**	−1.7580***	−1.3751*
			(1.2179)	(0.8120)	(0.7637)	(0.7941)
Individual characteristics						
Male	0.5675	0.5675	0.4618	0.4646	0.6768	0.4115
	(0.7594)	(0.7594)	(0.5556)	(0.4083)	(0.4026)	(0.3719)
Associate professor	-0.1678	-0.1678	0.0743	0.2276	-0.1538	-0.1807
	(0.7884)	(0.7884)	(0.5083)	(0.5121)	(0.3739)	(0.3799)
Full professor	-1.0859	-1.0859	0.0883	-0.1149	0.9804***	0.9854**
	(0.8522)	(0.8522)	(0.4965)	(0.5543)	(0.4389)	(0.4036)
Experience	0.3962***	0.3962***	0.3101***	0.2221**	0.1420*	0.1037*
	(0.1489)	(0.1489)	(0.1068)	(0.1044)	(0.0778)	(0.0720)
Experience squared	−0.0077***	−0.0077***	−0.0045**	−0.0041**	−0.0025*	−0.0019*
	(0.0027)	(0.0027)	(0.0017)	(0.0017)	(0.0013)	(0.0010)
PhD from US	-0.8710	-0.8710	0.1355	0.1861	-0.7664	-1.0020
	(1.4126)	(1.4126)	(0.6156)	(0.5901)	(0.5261)	(0.5715)
PhD from Canada	-1.5881	-1.5881	-0.2502	-0.5292	-0.2193	-0.3738
	(1.4412)	(1.4412)	(0.5360)	(0.5619)	(0.6388)	(0.5826)
Average rating on teaching	-0.3430	-0.3430	-0.2127	-0.2055	-0.2228	-0.2557
	(0.3140)	(0.3140)	(0.1431)	(0.1350)	(0.1191)	(0.1158)
SSHRC	−2.9707*	−2.9707*	−0.9154*	−0.9538***	0.6070	0.6483
	(1.7357)	(1.7357)	(0.5640)	(0.5173)	(0.4979)	(0.5633)
University characteristics						
Medical schools	1.2018		1.2335		1.2207	
	(0.6581)		(0.6882)		(0.7668)	
Comprehensive universities	0.0953		-0.1674		0.1641	
	(0.6346)		(0.4393)		(0.7454)	
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes	No	Yes
R-squared	0.4985	0.4985	0.3460	0.5281	0.1458	0.1926
Observations	159	159	105	105	1,517	1,517

Note: Dependent variable in all columns is the total number of corresponding publications aggregated over 3 year period. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. Proportion of other type of co-authored articles is the omitted category.

Table A.8: Tobit estimates (Marginal Effects) based on individual level data (1996-2012) and with year-university fixed effects, using number of publications as measure for research productivity

	publications in top 10 journals		publications in top 21 journals		all publications		AER	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Publications (top journals)	0.0273*	0.0126	0.0289**	0.0163	0.0081***	0.0081***	0.0021*	0.0021*
	(0.0146)	(0.0014)	(0.0132)	(0.0125)	(0.0028)	(0.0028)	(0.0011)	(0.0011)
Non-top journals	0.0066**	0.0077***	0.0061**	0.0073***				
	(0.0029)	(0.0027)	(0.0029)	(0.0027)				
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes	No	Yes	No	Yes
Observations	5374	5374	5374	5374	5374	5374	5374	5374

Note: Dependent variable is the natural log of salary. Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. We also include teaching and service characteristic, a dummy for whether individual i obtained his/her highest degree form US or Canadian institution, sshhrc dummy, rank dummy, gender, experience and its squared, etc in each specification.

Table A.9: Tobit estimates (Marginal Effects) based on individual level data (1996-2012), using citation counts as measure for research productivity

	without co-authorship adjustment		with co-authorship adjustment	
	(1)	(2)	(3)	(4)
Citations($ln(citation+1)$)	0.0054***	0.0053***	0.0063***	0.0063***
	(0.0014)	(0.0014)	(0.0015)	(0.0015)
Year fixed effect	Yes	Yes	Yes	Yes
University fixed effect	No	Yes	No	Yes
Observations	5374	5374	5374	5374

Note: Standard errors are in parenthesis and clustered at the individual level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. We also include teaching and service characteristic, a dummy for whether individual i obtained his/her highest degree form US or Canadian institution, sshhrc dummy, rank dummy, gender, experience and its squared, etc in each specification.

Appendices of Chapter 3

OLS estimates of school competition on school performance, log-log model

Table A.10: OLS estimates based on school level EQAO data (2004 - 2013)

	OSSLT passed PE		OSSLT passed FTE		Grade 9 level 3 math AC		Grade 9 level 3 math AP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OLS with year fixed effect								
Number of all nearby school	-0.0264^{*}		-0.0096		0.0477^{***}		-0.0606^{***}	
	(0.0171)		(0.0132)		(0.0158)		(0.0242)	
Number of nearby public school		-0.0189		-0.00021		-0.0128		-0.0302
		(0.0207)		(0.0149)		(0.0196)		(0.0286)
Number of nearby Catholic school		-0.0035		-0.0147		-0.0235		0.0040
		(0.0197)		(0.0164)		(0.0201)		(0.0344)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.3199	0.3311	0.5995	0.5699	0.1677	0.1797	0.1770	0.1736
Observations	3,076	3,076	3,076	3,076	3,076	3,076	3,076	3,076
School Board Fixed effect								
Number of all nearby school	-0.0224		-0.0078		-0.0230		-0.0140	
	(0.0169)		(0.0123)		(0.0157)		(0.0240)	
Number of nearby public school		0.0022		0.0018		0.0238		0.0361
		(0.0246)		(0.0172)		(0.0200)		(0.0304)
Number of nearby Catholic school		-0.0084		-0.0214		-0.0238		0.0092
		(0.0224)		(0.0172)		(0.0220)		(0.0370)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.3449	0.3517	0.6033	0.5744	0.2111	0.2263	0.2553	0.2445
Observations	3,076	3,076	3,076	3,076	3,076	3,076	3,076	3,076
Restrict data to Peel, Halton, Toronto, York and Ottawa								
Number of all nearby school	-0.0326		0.0026		0.0071		0.0223	
	(0.0308)		(0.0216)		(0.0231)		(0.0360)	
Number of nearby public school		0.0317		0.0344		0.0433		0.0608
		(0.0312)		(0.0222)		(0.0237)		(0.0401)
Number of nearby Catholic school		-0.0325		-0.0358		-0.0514		-0.0294
		(0.0280)		(0.0228)		(0.0264)		(0.0465)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.4071	0.4150	0.5427	0.5291	0.2607	0.2638	0.2154	0.2171
Observations	1664	1664	1664	1664	1664	1664	1664	1664
Only TDSB								
Number of all nearby school	0.1801^{***}		0.1467^{**}		0.1253^{**}		0.1293^{*}	
	(0.0730)		(0.0622)		(0.0607)		(0.0654)	
Number of nearby public school		0.1623^{***}		0.1250^{***}		0.1217^{***}		0.1200^{***}
		(0.0535)		(0.0430)		(0.0500)		(0.0449)
Number of nearby Catholic school		0.0335		0.0175		0.0102		0.0321
		(0.0432)		(0.0451)		(0.0432)		(0.0900)
School board fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.4382	0.4532	0.5010	0.4939	0.2862	0.2901	0.2032	0.2078
Observations	734	734	734	734	734	734	734	734

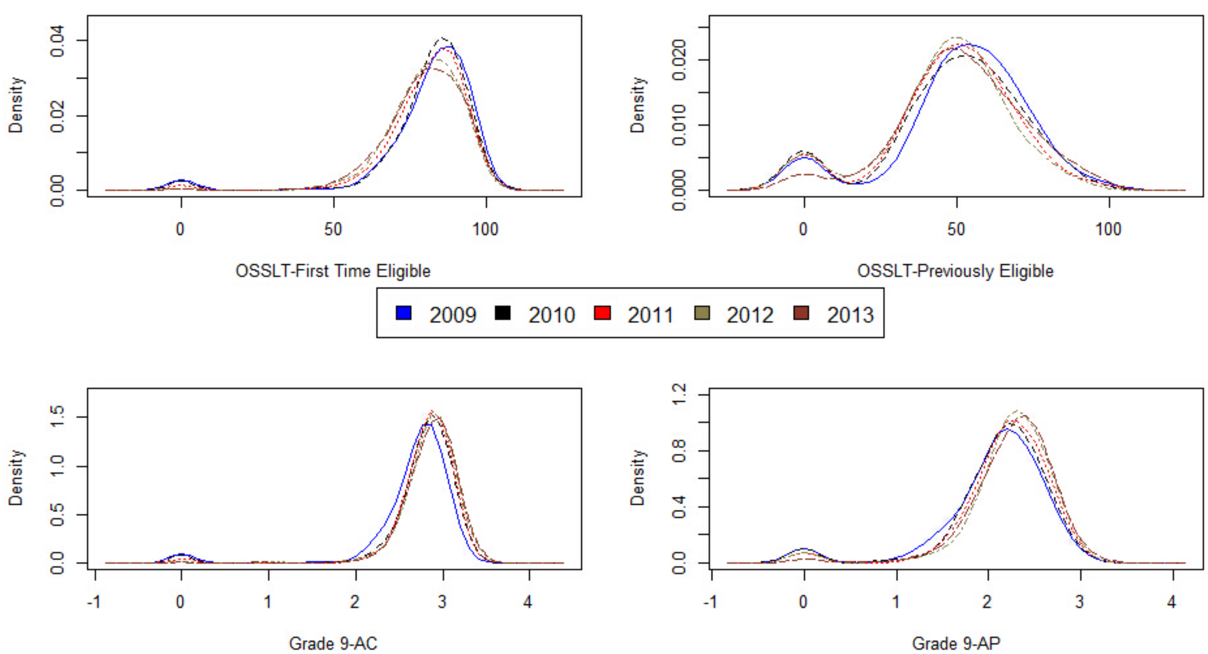
Note: Standard errors are in parenthesis and clustered at the school level. ***, **, * indicate significant level at 1 percent, 5 percent, and 10 percent, respectively. All regressions also include the neighborhood and school characteristics.

Another data-set collected from the Fraser Institute

Table A.11: Summary statistics of school achievement data

Variables	Observations	Mean	SD	Min	Max
OSSLT passed PE	1668	0.50	0.201	0	1
OSSLT passed FTE	1668	0.81	0.135	0	1
Grade 9 level 3 math-academic	1668	2.79	0.44	0	3.5
Grade 9 level 3 math-applied	1668	2.16	0.54	0	3.3

Figure A.1: kernel-density plot of school achievement data from 2009 to 2013, by different EQAO assessment



The codes of producing Figure 3.6

```
mysetdiff<-function (x, y, multiple=FALSE)
         {
         x <- as.vector(x)
         y <- as.vector(y)
         if (length(x) || length(y)) {
         if (!multiple) {
         unique( x[match(x, y, 0L) == 0L])
         }
         else  x[match(x, y, 0L) == 0L]
         }
         else x
         }
mydistnace<-function(x)
          {
          z<-matrix(,10,4)
          s<-vector()
          dis<-list()
          for (i in 1:10) {
          z[i,1]<-abs(x[i]-0)
          z[i,2]<-abs(x[i]-3)
          z[i,3]<-abs(x[i]-6)
          z[i,4]<-abs(x[i]-9)
          s[i]<-min(cbind(z[i,1],z[i,2],z[i,3],z[i,4]))
          dis[[i]]<-mysetdiff(z[i,],s[i],mult=T)
          }
          return(dis)
          }
x<-c(0:9)
distance<-simplify2array(lapply(mydistnace(x), mean))
plot(x, distance, type="b",xlab="schools",col=1:length(x),xaxt = "n",
     ylab="average distance" )
axis(1, at=c(0,3,6,9), labels=letters[1:4])
mtext(c("A"),side=1,at=1,col="blue")
arrows(1, 5, x1 = 1, y1 = 0, length = 5, angle = 10, code = 2,
       col = "blue",lty = 3, lwd = par("lwd"))
```

Average number of nearby schools and average school performance, by school board

Figure A.2: Average number of nearby schools and average school performance, by school board