The relationship between earnings, education, age, and cohort size for Canadian men during the period 1986-2001

By Giang Thi-Thuy Nguyen
Student number: 2962164

My major paper is presented to the
Department of Economics of the University of Ottawa
in partial fulfillment of the requirements of the M.A Degree

Supervisor: Professor Kathleen M. Day

ECO 7997

Ottawa, Ontario
September 2007

Abstract

Individual data for Canadian men from Censuses for the years 1986, 1991, 1996 and 2001 are used to re-examine Dooley's (1986) study of changes in the relationships between experience, education, cohort size, and earnings. This study adopted Dooley's (1986) model with some modifications. The difference between current Canadian male earnings patterns and those in the 1970s when the Baby Boom generation entered the labour force is used to find out if Canadians are overeducated. The regression analyses indicate that cohort size does not always have a negative effect on earnings. In contrast, it has a positive effect on earnings for individuals with a university education, and thus confirms that Canadians are not overeducated. The analyses also reveal the important role of experience in increasing earnings regardless of educational attainment levels.

I. Introduction

Education and earnings are topics that have received a lot of attention. In fact, everyone has to make a decision regarding how much to invest in education. The answer will depend on the return to education, which is reflected by earnings. Given the importance of earnings and education, I chose to study the relationships between earnings, education, experience, and cohort size of Canadian men during the period 1986-2001. For simplicity, the analyses are only performed on male earnings, because women have discontinuous work patterns which make it difficult to accurately measure the relationships between earnings, experience, education and cohort size. Moreover, the results from using only male observations are more comparable to previous studies, and thus it is easier to compare the current earnings pattern to those of the 1970s and 1980s.

During the late 1970s and early 1980s, there were a lot of empirical studies of the effect of the Baby Boom generation's entry into the labour force on earnings in the United States (U.S) and Canada. Dooley (1986) studied Canadian men's earnings during the period 1971-1981 in order to see the cohort size effect of the Baby Boom generation on earnings and to answer the question "Are Canadians overeducated?" As the Canadian demographic structure changed during the period 1986-2001, the labour force structure also changed since the Baby Boom generation did not have any effect on labour force entry. In fact, they were close to retirement age by the end of my study period. Therefore, there may be a shortage in labour supply, especially for highly educated workers. The labour force growth rate has recently decreased due to the retirement of the Baby Boom generation and the declining population growth rate (Human Resources and Social Development Canada, 2007).

Therefore, I chose to redo Dooley's (1986) study using more current data from the Canadian Census for the period 1986-2001 in order to see if his findings still hold. However, I will modify Dooley's (1986) model to complement the difference in data sources. In this paper, I will try to provide an overall picture of current Canadian male earnings. I will focus on the following issues: (1) Are Canadians over educated? (2) What are the effects of experience and cohort size on earnings?

This paper includes five sections. Following this introduction is section II which provides a literature review of some empirical studies in the U.S and Canada during the 1970s and 1980s. Section III will discuss how the data were obtained from the Canadian Census and how to prepare the data for regression analyses. Section IV is about the model and explains the reasons why I modified Dooley's (1986) model. Section V is a discussion of some descriptive tables and empirical regression analysis results. Finally, a summary of my findings in section VI will conclude this paper.

II. Literature Review

In this review, I will look at some previous studies which relate to relationships among earnings, education, cohort size, and age or experience. Most of them are empirical studies of the U.S or Canadian context which are quite similar. However, there are also some differences in country specific factors such as data set, educational policy, demographics, etc. Therefore, these studies will yield some similar as well as different results. I will examine the similarities and differences of these studies in their approaches, types of data, models, and results.

1. General approach

There were many studies of the effect of education, age, or demographic structure on earnings in the U.S or Canada during the 1970s and 1980s. First, Vaillancourt & Henriques's (1986) study can be considered a general example of studies of the relationship between education and earnings. This study estimated the return to education which can be considered as a numerical result of interaction among education, experience, labour demand, and so on. Vaillancourt & Henriques (1986) used the most basic form of the human capital model, which includes age and age squared only as determinants of earnings. In addition, the return to education can provide us with some guidelines at the individual level since it is such a valuable resource for the decision of how much schooling to obtain.

A second group of studies focuses on the effect of cohort size on earnings. These studies examined some events such as the Baby Boom Generation's entry into the labour force or some changes in policy which could cause changes in enrolment in education, labour force participation, and return to education, and thus consequently affect earnings. In most of the cases, empirical data for the U.S or Canada over the late 1970s and early 1980s period were used to perform OLS regressions or to create charts in order to confirm these effects. They also applied the human capital model in their studies. However, their models are more complicated then Vaillancourt & Henriques's (1986) model because they included more determinants of earnings beside age and age squared, such as cohort size, time trend, unemployment rate, etc. These real life events help us to test the theory of various effects on earnings. Freeman (1979), Welch (1979) and Berger (1985) looked at the effect of the Baby Boom generation's participation in the labour

force on earnings in the U.S., while Dooley (1986) examined this same event in Canada. Dooley & Gottschalk (1984) also examined the effects of the Baby Boom generation on the labour force and earnings; however, their study is slightly different from those of Freeman (1979), Welch (1979), Berger (1985) and Dooley (1986) since they tried to explain the earnings inequality due to labour force growth.

Although they addressed different issues rather than the cohort size effect on earnings, the study of Lemieux & Card (2001) is still indirectly related to cohort size. The main focus of Lemieux & Card's (2001) paper is ability bias. However, they also used a unique event, the "Veteran Rehabilitation Act" (VRA) in Canada, as a natural experiment to study the effect of a large scale educational program on earnings and inter-cohort patterns of earnings and education. The VRA is one of two changes in public policy, the other being the G.I Bill program in U.S, which received a lot of attention.[1] The results of this study show that not only social events but also changes in public policy can affect earnings through changing cohort sizes.

One similarity between this study and the previously mentioned studies is that they consider the male observations only. However, some differences from other studies are Lemieux & Card's (2001) only studied certain groups such as males with university or college education, who were benefited from the VRA, and they did not explicitly look at the effect of cohort size on earnings. In addition, the policy changes are harder to predict than demographic changes. In fact, their study is still related to the above studies

[1] The G.I. Bill was signed in June 22, 2006 by President Franklin D. Roosevelt with the formal name of Servicemen's Readjustment Act (United States Department of Veterans Affair, 2007). In Canada, the Veteran Rehabilitation Act was also signed in 1944 (Lemieux & Card, 2001). These acts "established a comprehensive program of benefits for returning WW II veterans to further their education and ease their transition to civilian life" (Lemieux & Card, 2001, p. 315).

of the cohort size effect on earnings because they look at the cohort of veterans after World War II in order to study the effect of the large scale educational program on earnings and inter-cohort patterns of earnings and education. Lemieux & Card (2001) also estimated the return to education which is the main focus in Vaillancourt & Henriques's (1986) paper.

Although these fore-mentioned studies used different approaches to explain the changes in wage patterns, they all examined wage behaviour during the period after World War II. In particular, Welch (1979), Berger (1985), Freeman (1979), and Dooley & Gottschalk (1984) all suggested that a large cohort size would have a negative effect on earnings within each cohort. Theory predicts that the cohort size has a negative effect on earnings within each cohort if holding other things constant, because of simple labour market equilibrium. If the cohort size is larger, the labour supply will increase, and thus earnings will decrease because we are holding all other things constant including labour demand. This prediction was supported by the empirical evidence of larger cohort sizes in the labour market due to the entry of the Baby Boom generation into the labour force during the 1970s and 1980s. This theoretical prediction may be the reason why many authors expect the cohort size to have a negative effect on earnings. However, this may not always be the case since Connelly (1986) notes that the cohort size effect of one group on another should be also taken into account. Connelly (1986) also noticed some dissimilarities in predictive magnitude of the cohort size effect on earnings and inconsistencies among these studies. She thought it was the result of "differences in underlying model of labour demand" (p. 544). Therefore, she developed a framework to

analyze how the structure of the production function affects predictions about the change in earnings due to birth cohort (Connelly 1986, p. 544).

Today, the effect of demographic changes on the workforce continues to receive attention. In the Canadian public sector, demographic analysis and forecasting are used for Human Resource planning (Canadian Public Service Agency, 2000). As a result, the Canadian Public Service Agency (2000) has created a website for demographic analysis of public service workforce. One of its aspects is the use of the human capital model to determine why employees have different income levels. Human capital theory originated from Adam Smith's discussion of equalizing differences in his book "The Wealth of Nations" in 1776 (Berndt, 1991). In this model, besides the main factors such as education and experience, earnings are also associated with gender, official language, post secondary education, bilingual status, and so on.

2. Types of data

In order to perform the analysis, different data sources were used to collect necessary information on earnings, education, age, experience, etc. Various types of statistical survey may contain these variables. The Current Population Surveys (CPS) are widely used in the U.S while the Surveys of Consumer Finances (SCF) and the Canadian Census are common for Canadian studies.

Freeman (1979), Welch (1979), and Dooley & Gottschalk (1984) used the CPS for the period during the late 1960s and early 1970s. Dooley (1986) used the SCF from 1971 to 1981 with two-year intervals. The above two types of survey are quite similar in terms of how they were conducted. They include people of working age only. The SCF before 1981 and the CPS include observations who are the heads of house holds only.

Since 1981, all individuals of working age have been included in the SCF. However, there are still some differences between the CPS and SCF data. The CPS, which is conducted on all individuals 15 years old or older in 50,000 households monthly, is considered a primary source of U.S labour force information (U.S. Census Bureau, 2006). On the other hand, the SCF is conducted annually, and before 1981 it only provides information about the head of households. Thus, there is no information on other members of the household or people who have never been married (Dooley 1986, p. 147). In my opinion, this SCF data before 1981 provides a better representation of earnings for people who are married than earnings in general. In addition, the SCF is only available until 1997, and thus we can not use the SCF to study current Canadian earnings after 1997 in an overall context.

Regardless of the differences and similarities of the SCF and CPS, Freeman (1979), Welch (1979), Dooley & Gottschalk (1984), and Dooley's (1986) samples are quite similar. First, they all studied male earnings only. As was explained by Dooley (1986), the reason for using male earnings only is "the frequently intermittent nature of the market work pattern of women" (p.145). Second, they considered individuals at working age with positive income, although the age range varies a little across studies. Welch (1979) chose the age group between 14 and 64. Dooley & Gottschalk (1984) followed the same strategy as Welch (1979) but restrict the age range from 16 to 64. Freeman (1979) chose the age group from 18 to 64 and Dooley (1986) studied the group from 20 to 64. Third, they excluded observations with self-employment or imputed earnings.

Besides the above mentioned surveys, the Census is also a useful data source which was used by Vaillancourt & Henriques (1986) to study the return to education in Canada and Lemieux & Card (2001) to study the effect of the VRA. These studies are similar to the above study since they consider male observation only. The Census is conducted every 5 years on all individuals, in contrast to the SCF, which contains only information on heads of households before 1981 (Computing in Humanities and Social Science, University of Toronto, 2005). Therefore, it is less biased than the SCF in providing an overall picture for Canada. However, a weakness of the Census data is not providing enough comparison between specific years since it is conducted at 5-year intervals. Therefore, it will be harder to discover the trend, financial peak, and financial trough during business cycles.

There is another important difference in how the data were used for regression analyses. Except for the studies by Vaillancourt & Henriques (1986) and Lemieux & Card (2001), the observations in the regression analyses of the above mentioned studies were aggregated across individuals in groups by education and age. Moreover, the group of studies which use these aggregate data also use the SCF or CPS. On the other hand, Vaillancourt & Henriques (1986) and Lemieux & Card (2001) use individual observations from the Census for their regressions.

3. Models

All studies, which used data for the 1960s and 1970s that I have discussed above, attempted to explain wage behaviour. Hence, all models of those studies were built for the purpose of explaining various effects on wages. Most of them were slight modifications of the human capital model of earnings, and they agreed on the main

independent variables for their OLS regression for earnings, which are age, education, experience. Besides, Welch (1979), Freeman (1979), Dooley (1986), Berger (1985), and Dooley & Gottschalk (1984) added size of cohort, unemployment rate and time trend in order to study the cohort size effect on earnings.

The effect of birth cohort size on earnings received a lot of attention during the 1970s and 1980s when the Baby Boom generation entered the labour market. Connelly (1986) developed a framework to analyze this effect under different assumptions about the labour demand function which contribute to differences among the fore-mentioned studies. As a result, Connelly (1986) examined the following two models: (1) imperfectly substitutable workers across age groups (used by Freeman 1979); (2) imperfectly substitutable workers across experience-schooling groups (used by Welch 1979, Berger 1985). First, model (1) will have the production function form Q = F (N(a), K) where a is age, K is capital, and N(a) is a vector of the number of workers at age a (Connelly 1986, p. 545). This shows that the labour demand function N(a) will only depend on age and labour is imperfectly substitutable across age. According to Connelly (1986), the labour market cohort is $n(i) = \lambda(i) * b(t-i) * p(i)$ where $b(t-i)$ is the number of people born $t-i$ years ago, $\lambda(i)$ is the labour force participation rate of people at age i, and $p(i)$ is the probability of surviving from birth to age i. Second, the production function form of the model (2) is Q = F (N(a,s), K) where a is age, s is school attainment, K is capital, and N(a,s) is a vector of the number of workers at age a and school attainment level s (Connelly 1986, p. 547). In this model, labour demand depends on age and schooling and thus labour is imperfectly substitutable across age groups and schooling groups. Besides, Connelly (1986) stated that Dooley & Gottschalk (1984) do not use either of the above labour

demand models but assume labour to be "a single homogenous factor" in the production function (p.544). After developing the above framework and looking at previous empirical results, she suggested that "the importance of the linkage between birth cohort and labour market cohort has been overlooked" (Connelly 1986, p. 560). In addition, she thought that the change in wage of one age group should take into account the effect of changes in number of workers in its own group and other age groups as well. Finally, Connelly (1986) suggested that the analysis of changes in school enrolment rates and changes in earnings should be combined.

Next, I will discuss the model of cohort size effect on earnings in the papers of Welch (1979) and Berger (1985). There are some important assumptions and restrictions in both of their models. First, experience is not reported in the CPS and thus Welch imputed it.[2] Secondly, he used the career phase, which contains learning and working phases, to view the cohort size effect. Since the new worker must divide his or her time between learning and working, their wage will be affected (Berger 1985, p. 563). However, this effect should diminish over time as workers take less time learning. Third, Welch (1979) ignored the substitution between schooling groups because he wanted to concentrate his analysis on earnings differences within the same educational attainment group. Berger (1985) maintained the previous assumptions but relaxed Welch's key assumptions that "speed of transition from learner to fully vested worker is exogenous and thus independent of cohort size" because of congestion, costly or poor education due to large cohort size (p. 563).

[2] The detail of experience imputation was not discussed in his paper about the cohort size effect on earning. However, the procedure to estimate the number of years in the job market is briefly explained (Welch, 1979).

Berger (1985) re-examined Welch's (1979) study using a similar data set but added the years 1976-1979 to the sample period 1967-1975. They both categorized workers by age and education level and thus implied a labour demand function in which workers are imperfectly substitutable across age and schooling. In terms of the model, Berger (1985) and Welch (1979) both use similar regression equations in which the dependent variable is weekly or annual earnings, the independent variables include cohort size, experience, experience squared, spline, interaction of spline and cohort size, unemployment rate, time trend, and nonwork exclusion rate.[3] However, Berger (1985) modified Welch's (1979) model in order to estimate separate equations for two groups: (1) spline variable is greater than 1; (2) spline variable is smaller than or equal to 1. Therefore, Berger's (1985) model relaxes the restriction in Welch's (1979) model that all spline variables should be estimated in one group. Berger (1985) tested this restriction to see if the analysis of cohort size effect should be performed using his or Welch's (1979) model. He argued that an increase in cohort size may lead to an increase or decrease in human capital investment (Berger 1985, p.563). This conflicts with Welch's (1979) view that his result implies optimal investment in human capital.

In addition, I will look at the model used by Freeman (1979) in which the implied labour demand function is the first model in Connelly's (1986) study discussed above. He

---

[3] Berger (1985) noted that experience is imputed using a "full density" approach which was developed Welch & Ground (1976). The spline variable is calculated using following formula in Berger (1985, p. 565):

$$spline = \left(1 - \frac{EXPERIENCE}{EXPERIENCE'}\right) \times D$$

where D = 1 if EXPERIENCE < EXPERIENCE'; D = 0 if EXPERIENCE > EXPERIENCE'
EXPERIENCE is imputed using "full density" approach, EXPERIENCE' is exogenous.
The spline variable represents the rapid growth in earnings at the early career phase (Welch 1979, p. S87). In addition to spline, experience squared and cohort size & spline interaction variables are included in the regression equation to allow for the effect of cohort size and experience vary overtime (Berger 1985, p. 564).

used the standard labour demand framework, in which relative labour supply by age depends on the level of input substitutability (Freeman 1979, p. 301). As an illustration, a production function with high input substitutability will lead to less wage fluctuation if there is a large change in labour supply. He also estimated workers substitution and complementarity elasticities using constant elasticity of substitution (CES) and translogarithmic (TL) production function forms (Freeman 1979, p. 302). When determining the effect on earnings, Freeman used three independent variables: cycle, relative number of workers by age, and time. There is no education related independent variable in this OLS regression.

I will now examine Dooley's (1986) study, on which I will primarily base the method and regression analysis for this paper. He followed the model of Welch (1979) quite closely, but with some modifications to match his Canadian data set. First, he examined the mean of earnings and changes in the mean of annual or weekly earnings by age and school attainment category in order to look for trends. Second, he used OLS regression to estimate the effect of various factors which consists of age, education, cohort-size, the unemployment rate, and time trend on earnings. His model is slightly different from Welch's model because age is used as a substitute for experience, and there is no spline variable or interaction of spline and cohort size variable (Dooley 1986, p. 151-152). Moreover, in order to fix the problem that SCF data only provides information about the head of households Dooley defined his dependent variable as the difference in log earnings between time periods in his OLS regression.

Other studies which explain other events such as the VRA in Canada, the G.I. Bill in the U.S or estimate the return to education also apply the human capital in their

analyses. The return to education is an important component of the educational investment decision. Vaillancourt & Henriques (1986) estimated the return to university schooling in Canada by looking at the future earnings of white men aged 18 or older who are facing the decision of whether to enter university or not (p. 449). Vaillancourt (1995), and Vaillancourt & Bourdeau-Primeau (2002) also addressed the same issues using more current data. In order to estimate that rate of return, Vaillancourt & Henriques (1986) examined the tax profile, the earnings profile, the cost of attending university, the participation and unemployment rate, and the discount rate. The age-earnings profile was calculated using the OLS regression of log earnings on age and age squared which is quite similar to but simpler than the models used by Freeman (1979), Welch (1979), or Dooley (1986) above. However, Vaillancourt & Henriques (1986) did not look at the cohort size effect.

Furthermore, the effect of large scale education on earnings as a result of the VRA event in Canada was studied using OLS regression and IV estimation (Lemieux & Card 2001, p. 329-330). The OLS model consists of the two following equations:

(1) $S_i = X_i \alpha + u_i$ (S is schooling, X is vector of age, experience, etc)

(2) $y_i = S_i \beta + X_i \gamma + v_i$ (y is earnings, S is schooling, X is vector of age, experience, etc)

However, this model faces a bias problem due to a correlation between u_i and v_i since people with higher wages tend to obtain more schooling (Lemieux & Card 2001, p. 331). Therefore, Lemieux & Card use the instrumental variables method to fix this bias by adding the covariate Z_i which affects schooling but not earnings in equation (1) & (2) above. The model will become:

(3) $S_i = X_i \pi_{sx} + Z_i \pi_{sz} + \eta_i$

(4) $y_i = X_i \pi_{yx} + Z_i \pi_{yz} + \varepsilon_i$

In this paper, Lemieux & Card (2001) use "potential eligibility for VAR benefits as an instrument for schooling outcome" (p. 331). They also report the results of OLS and IV estimates of the return to education using the above models.

4. Results

The results of the above mentioned studies have some common features but also some differences which may be due to differences in data sets, methods, etc. Most of these studies except those of Vaillancourt & Henriques (1986) and Lemieux & Card (2001) focus on the effect of cohort size on earnings which is theoretically predicted to be negative if we are holding all other things constant. Their empirical results concluded that a larger entering cohort leads to lower relative wages of new entrants compared to experienced workers, especially among the groups with higher education which is consistent with the previously mentioned theory. A lot of studies about the effect of the Baby Boom generations entering the U.S labour force during the 1970s also confirmed this result of the negative cohort size effect on earnings (Welch 1979, Freeman 1979, Dooley & Gottschalk 1984). According to Dooley (1986), Welch and Freeman arrived at the same conclusions about the effect of the Baby Boom generation's entry into the labour market in the U.S:

> First, during the late 1960s and the first half of the 1970s, there was a sharp decline in the ratio of the earnings of young U.S male workers to older ones and that this decline was more pronounced among university graduates than among workers with a high school education. Second, these phenomena were caused, at least in part, by the labour market entry of the baby boom and the especially rapid growth in the number of university graduates. Finally, it is agreed that young university graduates suffered more, relative to less educated youth, than would be warranted by a *simple* supply shift explanation. (Dooley 1986, p. 144)

Furthermore, Welch (1979) suggested that the wage loss of the Baby Boom generation due to that large entrance will diminish over time as they gain more experience.

Berger (1985) re-examined Welch's (1979) study of the effect of cohort size on earnings. In contrast to Welch, he concluded that this adverse effect on earnings will remain throughout the career of a large cohort. It may be caused by Berger's (1985) relaxation of Welch's (1979) model assumption about the speed of transition from learners to fully-vested workers as mentioned in subsection 3. Also according to Berger (1985), other studies with different identification assumptions about experience, cohort size, and period will possibly have different conclusions (p. 573).

Dooley (1986) performed a similar study on Canadian data during the period 1971-1981 which was obviously affected by the Baby Boom generation's participation in the labour force. He said these data reveal the wage compression across levels of education and ages. He also confirmed the decrease in relative earnings between people with university degrees and those with some secondary education (Dooley 1986, p. 156). Additionally, the estimated private rates of return to university schooling range are from 7% to 15% depending on different regions in Canada (Vaillancourt & Henriques 1986, p. 454). These rates decreased by 5% from 1971 to 1981 (Vaillancourt & Henriques 1986, p. 455). The study by Lemieux & Card (2001) of the effect of the VRA program in Canada also showed a strong effect of policy changes on earnings. They found the estimated return to schooling using the 1971 Canadian Census data is 15%. However, according to Lemieux & Card (2001), this rate of return to schooling is only 7% in the absence of VRA, which is quite close to Vaillancourt & Henriques's (1986) result.

In summary, besides the two most important determinants of earnings during the 1970s and early 1980s, which are experience and education, there are some other factors which can significantly affect earnings. These factors are cohort size, social events, and expected return on education. Changes in cohort size represent change in the labour supply, which will affect earnings through the labour market. Some special social events or changes in policy such as the VRA also affect earnings indirectly through its effect on the educational entrant rate. In most of the above studies that I have reviewed, the effect of demographic change on earnings was emphasized a lot. I think the effect of demographic change may be overestimated because the possibility that people will adjust their educational investment decision was not taken into account. However, individual educational investment decisions are normally based on the rate of return to education. In fact, not all individuals may make a rational decision because most people may not be able to estimate this return to education. Therefore, I think that is why Dooley (1986) concluded that the average Canadian is overeducated. Overeducated here means people over-invest in their education, and thus there is an over-supply of labour at high education levels. Consequently, their earnings are lower than the earnings they would receive if they chose the optimal investment in education. In my opinion, because Canadians during the 1970s and 1980s did not fully take into account the effect of the Baby Boom generation's entry into the labour force, they over-invested in education because of overestimating the return to education.

In the later sections, I will perform the analysis on the wage behaviour during the period 1986-2001 on Canadian data. I will then compare my results on more current Canadian demographics to these above results in order find out if there is any change and why. Due

to demographic change and the Baby Boom generation being close to retirement age, the cohort size is much smaller now. Theoretically, holding other things constant, cohort size is expected to have a negative effect on earnings. However, because current cohort sizes are different from those in the past and not all relevant factors - which affect earnings - are included in my model, I expect that cohort size may not have a negative effect on earnings for all levels of educational attainment in regression analyses. In other word, a larger cohort size might raise earnings. Experience, the unemployment rate, and the time trend are expected to have similar effects on earnings as in the above studies.

III. Data

The data used for the analysis in this paper are from the Canadian Censuses of 1986, 1991, 1996, and 2001, which cover a 15-year period. These are obtained from the microdata files in the Canadian Census Analyser.[4] According to the microdata note in the Canadian Census Analyser, these files consist of samples of 500,434 (in 1986), 809,654 (in 1991), 792,448 (in 1996), and 801,055 (in 2001) which is about 1-3% of the total population. From each Census microdata file, I obtained a sample which is only 10% of the total sample. Therefore, I have four samples of 24,227 (in 1986), 19,831 (in 1991), 18,026 (in 1996), and 18,895 (in 2001). These samples contain only applicable or available observations with certain characteristics which I will discuss later.[5] I will use the information about age, wage, highest level of education, and full-time or part-time

[4] Access through University of Ottawa library website: http://proxy.bib.uottawa.ca:2256/census/.
[5] Applicable or available observations are the observations that match the variable definition criteria and that Statistic Canada has information about. For example, the wage and salaries variable is only reported for individuals aged 15 or older. Therefore, it will not be applicable for people under 15 years old. Similarly, the highest level of schooling variable is also reported for the population of the age 15 or older. (this information is obtained from 2001 Census Public Use Microdata File: Individual File Users Documentation, Statistics Canada 2001)

work status to calculate all the variables needed for my analysis.[6] Further details on calculation and analysis will be discussed later in section IV.

The Canadian Census data are different from the Survey of Consumer Finances data that were used by Dooley (1986) in the following ways. First, the Census is conducted at 5-year intervals, while the SCF was conducted every year. This is a disadvantage of using the Census data since it is hard to discover any trend or provide a comparison between specific years. Second, before 1981, the SCF only provides information about the heads of households. This was considered a serious problem with the SCF since it has no observations that never married or lived with their parents (Dooley 1986, p. 147). On the other hand, the Census provides information on all individuals. Although since 1981 the SCF provides information on all individuals, it was terminated in 1997. Therefore, the use of the Canadian Census data can provide us with more current information about Canadian earnings.

On the other hand, I still follow very closely Dooley's (1986) method. As a result, I have put some limitations on the observations in my samples. First, my samples only contain individuals between the ages of 20 and 64 which is the same as Dooley (1986). In my opinion, this age range suits my analysis of people at working age. Since there was a change in the requirements for high school completion,[7] I chose the starting age of 20, instead of 18, to make sure that most people in my sample have completed high school unless they dropped out. Second, I will conduct an analysis of Canadian male earnings only. Today, men and women are more equal in working, and thus women's discontinuous work patterns are less severe than in the past. However, the women still

[6] The Shazam econometric software is used for statistical analysis and Excel is used for creating the tables.
[7] Students are now required to complete a 12 grade program instead of a 13 grade program to graduate from high school in Ontario.

tend to take more time off during their working life than men for maternity leave or other family issues. Therefore, male earnings can represent earnings in my analysis better since male age is a more accurate proxy for experience than female earnings. Moreover, I decide to use male earnings only in this analysis in order to make my results more comparable to those of the previous studies in the 1970s and 1980s. By comparing my results with those of the previous studies, I will be able to see how the earnings patterns have changed. Third, my samples only include observations with positive wages and salaries which exclude self-employment earnings. Hence, they have a positive number of weeks worked in the year previous to each Census year. The definition of the wages and salaries variable for both full-time and part-time workers in the Census does not include self-employment income or unpaid work. Although my samples include individuals with and without self-employment income which is not part of the earnings provided by the Census, there is no need for further manipulation of my obtained samples because I focus on the effect on earnings which excludes self-employment income. In addition, self-employment income is very small for full-time workers with positive income in my analysis. Therefore, I will ignore self-employment income in my analysis. Furthermore, most of my reports and analysis will be on full-time workers. The data set includes both full-time and part-time workers and hence, I will still have some reports which include both types of workers.

I will divide the observations in each sample into different age groups and educational attainment levels for the descriptive tables. The age groups are the same as in Dooley's (1986) study, which has 9 groups: 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and 60-64. The educational attainment levels are slightly different from

Dooley (1986). Based on the information provided by the Census on highest level of schooling, I will have six groups instead of four as in Dooley's (1986) study; they are less than high school, high school or trade certificate, some college without diploma, college with diploma, some university without degree, and university with degree. The university degree educational attainment level includes all individuals with bachelor's, master's, and PhD degrees. The Canadian Census provides the highest levels of education in fourteen categories which I think is too many and unnecessary for my analysis. Therefore, I merge them into six levels of highest educational attainment as mentioned before.[8]

In order to compare earnings between different Census years, I need to take into account the inflation rate. Therefore, all my figures on earnings will be in 2001 dollars. The inflation rates which will be used to convert all earnings to 2001 dollars are obtained using the inflation calculator tool, which is based on the monthly Consumer Price Index (CPI), to calculate the inflation rate (Bank of Canada, 2007).[9] In addition, I also retrieve monthly aggregate unemployment rates from 1986 to 2001, which will be used for regression analysis later, from the Labour Force Survey (this can be accessed through CANSIM II, University of Ottawa library).[10]

The samples that I am using for my analysis only represent from 0.2% to 0.3% of the total population.[11] Since each of the observations in the Census data represents many

---

[8] For more detail on highest level of schooling as reported in the Canadian Census, see appendix A.

[9] The Inflation Calculator Tool from Bank of Canada website calculates the average inflation rate from one year to another using the monthly Consumer Price Index (CPI).

[10] CANSIM (Canadian Socio-Economic Information Management System) is a database which provides time series data about various social and economic aspects in Canada. This database is provided by Statistics Canada and the University of Ottawa subscribes to it so that faculty, staffs, and students can access CANSIM II through the library website. By accessing table 2820087 of the Labour Force Survey estimate (LFS), I obtained the unemployment rate series V2062815 for people of age 15 or over of both sexes.

[11] The Census public use files contain a sample from 2% to 3% of Canadian population. My data set only takes 10% of these samples. Therefore, they only represent about 0.2% - 0.3% of the total population.

other observations which are not part of those samples, a weight corresponding to each variable should be used to generate estimations for the whole population (Statistics Canada, 2001). However, my samples are only 10% of Census data. Therefore, I will use a re-based weight variable which is computed as follow:

Re-based weight = weight variable in Census * (my sample size/total Census size)

Census is not a stratified sample and the weight variable only varies a little from 30 to 33. Before 1986, there is no weight variable available in the Census and the weights are assumed to be the same. Therefore, the weight variable should not have significant effect on empirical analyses. However, I still use the weight variable to construct data for the table of empirical results and variables of the regression analysis as it was suggested in the Census. The purpose of using weight variable is to confirm if weights have any effect on my results. I will further discuss the use of these weights in the next section about the model.

IV. Model

My regression equation is based on the following first equation from Dooley (1986, p.151) with some modifications.

$$\mathbf{Ln\ E_{it}} = \beta_0 + \beta_1 \mathbf{Age_i} + \beta_2 \mathbf{Age_i^2} + \beta_3 \mathbf{U_t} + \beta_4 \mathbf{t} + \beta_5 \mathbf{Age_i{*}t} + \beta_6 \mathbf{lnCS_{it}} + \beta_7 \mathbf{lnCS_{it}{*}Age_i} + \beta_8 \mathbf{PTM_{it}} + \varepsilon_{it} \qquad \mathbf{(1)},$$

where

Ln E_{it} = mean of log earnings (annual or weekly) of age group ***i*** in year ***t***
Age_i, Age_i^2 = level of age and age squared,
U_t = aggregate unemployment rate in year t,
$lnCS_{it}$ = log of the number of observations in age group *i* in year *t*,
t = time trend,
PTM_{it} = proportion of age group *i* that usually worked part weeks in year *t*

(used only for weekly earnings),

ε_{it} = error terms assumed $N(0, \sigma^2/n_{it})$ where n_{it} is the number of observations in age group *i* in year *t*. (Dooley 1986, p. 151)

First, I will look at the specification of the above equation. Age and age squared are substitutes for experience and experience squared. These substitutes are more appropriate for men than for women since men normally have more continuous working patterns than women. Age squared is put into the equation in order to capture the rapid earnings growth in early career, especially for individuals with more schooling (Welch 1979, p. 87). The unemployment rate presents how well the economy is doing in general and thus, its presentation in the equation shows the effect of the economy on earnings. The log of cohort size variable shows the impact of cohort size on earnings. The time trend variable takes into account the trend effect and the PTM variable takes into account part-time or full-time status. Finally, Age*t and lnCS*Age are variables which allow for some nonlinearity in the equation. The purpose of my analysis is to find out the relationship between education, earnings, age and cohort size. Therefore, the log of earnings, age, age squared, log of cohort size variables are necessary to be included in my model.

Second, I will discuss how to construct the variables for the above regression equation. Unlike Dooley (1986), I consider each year as one age group in order to increase the number of observations for the OLS regressions instead of dividing the age range from 20 to 64 into 9 groups of 5 years each as Dooley (1986). As a result, I have 45 age groups for each educational attainment level. The data from all of the Census years are pooled together for the OLS regressions. The cohort size equals the sum of all re-based weights which correspond to the observations in each of the educational attainment

levels and the age groups in each of the samples. Because of using the re-based weights, this cohort size is a better estimation of the number of people in the specific age group of the whole Canadian population in each of the Census years 1991, 1996 and 2001. This cohort size also varies with each of the educational attainment levels, age and the Census year. In addition, the log of earnings is also weighted using this re-based weight variable. Time t is the year 1986, 1991, 1996 and 2001 when the Census was conducted. The unemployment rate corresponding to each time period t is the average monthly unemployment rate for year t.

However, the weight variable is not provided in the 1986 Census. Therefore, I will focus on the regression analyses performed on the 1991, 1996, and 2001 Censuses only. I still run the regression analysis on all four Censuses of 1986, 1991, 1996, and 2001. The regression on four Censuses will be performed using unweighted data. The purpose of this is to see the impact of the weight variables on my results. The only difference between unweighted data and weighted data is how the cohort size is calculated. In the case of unweighted data, the cohort size is the number of observations by age and educational attainment levels in each sample from the 1986, 1991, 1996 and 2001 Censuses.

Third, the variable PTM_{it} is used by Dooley in the case of weekly earnings only while the Canadian Census provides us with annual earnings only. Therefore, I will omit this variable from my equation. I first run the regressions for each of the six educational attainment levels with Dooley's (1986) equation above except for not including PTM_{it} in the equation. Unfortunately, my results using Dooley's (1986) model show some evidence of multicollinearity. Except for the regressions for the educational levels

"college without diploma" and "university without degree", these models have good explanatory power with R-squared about or higher than 0.8; but the coefficients do not yield the expected signs.[12] The age variable, which is a substitute for experience, does not have the positive expected sign and is not significant either. The log of cohort size does not yield the expected sign either. Therefore, I ran tests for multicollinearity, the results of which are reported in appendix B. I found that all of the auxiliary R-squared for regressions of each explanatory variable on all others are above 0.9 which is very high (see table 10). The highest condition indexes are over 100 for the OLS regressions of all educational attainment levels (see table 9). These high condition indexes and high auxiliary R-squareds indicate a severe multicollinearity problem.

In order to correct this multicollinearity problem, I considered dropping some of the independent variables from this model. The purpose of the interaction terms Age*T and lnCS*AGE in the regression equation is to allow for some nonlinearity. In my opinion, I think they may not contribute much to the explanatory power of the model. Therefore, I decided to drop the interaction terms Age*T and lnCS*AGE for the purpose of correcting the multicollinearity problem. My regression equation will be as follows:

$$\mathbf{Ln\ E_{it}} = \beta_0 + \beta_1 \mathbf{Age_i} + \beta_2 \mathbf{Age_i^2} + \beta_3 \mathbf{U_t} + \beta_4 \mathbf{t} + \beta_5 \mathbf{lnCS_{it}} + \varepsilon_{it} \qquad \mathbf{(2)},$$

where the remaining variables are defined as before. I will estimate this equation and Dooley's (1986) above equation separately for each level of education. I will later compare the results of the regressions using Dooley's (1986) equation to those using the modified equation to see the effect of dropping Age*T and lnCS*AGE later in the paper.

In the above equation, Dooley (1986) allows for some heteroskedaticity in the error terms associated with the cohort size. Therefore, I used the Goldfeld-Quandt test to

[12] See table 4 for these results.

test for a specific type of heteroskedasticity associated with the cohort size. In addition, I also performed the White test, which is the most general test for heteroskedasticity. Both of these tests have the null hypothesis of homoskedasticity. Their results indicate that the P-value is close to zero for both tests and for all levels of educational attainment.[13] In this case of an obvious heterokedasticity problem, I think weighted least squares is the most appropriate estimation method. Therefore, I use a weight which is equal to one over the cohort size to correct the heteroskedasticity problem.[14]

V. Discussion of the descriptive and empirical regression analysis results

1. Descriptive tables

In my opinion, not only regression analyses can tell a story but also descriptive tables. Before doing regression analyses, I will look for some preliminary results of the relationship between earnings, education, cohort size, and age in the descriptive tables. In addition, some possible patterns or distributions are often shown in the descriptive tables but not the regression analysis.

I prepared three tables 1, 2, and 3 which are similar to the first three tables in Dooley's (1986) paper, but based on data from the period 1986-2001. Table 1 presents the mean of annual earnings for the 1986, 1991, 1996 and 2001 Censuses by age group and educational attainment level in 2001 dollars. All the means of annual earnings except

[13] Table 11 provides the test statistics and p-value of two tests for heterokedasticity.
[14] In Dooley's (1986) study, his cohort size and earnings are compute using the population weight provided in the SCF. He allowed for the heteroskedasticsity which associates with the cohort size. His cohort size and earnings are compute using the population weight provided in the SCF. However, he did not test for hetoreskedasticity or try to correct it.

that for the 1986 Census are weighted, using the weight variables provided in each Census.

The mean of annual earnings for each of the educational attainment levels shows a similar pattern. The mean of earnings increases with age until the ages of 45 to 54 and then decreases. This pattern can be seen more clearly by looking at the figure 1, which consists of graphs based on the data in table 1. In addition, this pattern is consistent in all Censuses except for the educational attainment level "some university without degree". Instead, there is some fluctuation in the means of earnings for individuals with some university without degree in all Census years. This is not a very surprising result because there are many reasons for leaving university without getting a degree, such as getting a good job offer, taking time off, etc. It is possible for some individuals at this level of education to have very high earnings while others have low earnings. Therefore, it is possible to have a fluctuating earnings pattern.

Moreover, the peak of earnings, which is between the ages of 45 and 54, is quite predictable. People normally accumulate experience during their working life and have sufficient experience by the age of 45 to 54. As a result, they are able to work more effectively with more experience and earn higher wages. It is interesting that the mean of earnings decreases from age 55 to 64 despite the fact that this group has the most experience. One possibility is that people from the age of 55 to 64 may prefer more

TABLE 1: Mean of annual earnings for full-time workers in Canada (all figures in 2001 dollars)

2001 Census

	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Less than high school	18,929	26,341	30,363	32,677	36,592	36,796	37,568	36,639	34,736
High school degree or trade certificate	20,238	28,657	34,047	37,436	44,018	42,792	44,601	44,737	42,454
Some college without diploma	19,924	28,170	32,200	40,380	41,795	43,685	42,443	38,801	36,727
College with diploma	21,890	32,342	41,177	43,426	47,046	49,231	51,315	48,119	40,357
Some university without degree	13,257	31,060	35,752	41,936	53,552	49,726	49,929	58,576	56,017
University with degree	18,523	36,885	49,737	57,968	64,464	64,942	66,398	66,559	63,183

1996

	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Less than high school	16,248	25,366	28,753	32,341	34,884	36,231	37,669	34,112	35,705
High school degree or trade certificate	18,107	27,493	34,514	38,336	39,353	43,625	43,846	43,653	42,640
Some college without diploma	16,444	28,639	33,381	41,168	42,705	41,713	44,178	44,341	37,754
College with diploma	18,145	31,454	37,516	42,966	45,695	46,979	48,740	46,149	40,505
Some university without degree	12,482	28,159	32,956	43,642	46,800	50,788	54,857	54,371	30,925
University with degree	14,620	30,724	45,201	53,531	56,299	62,985	65,727	62,948	66,471

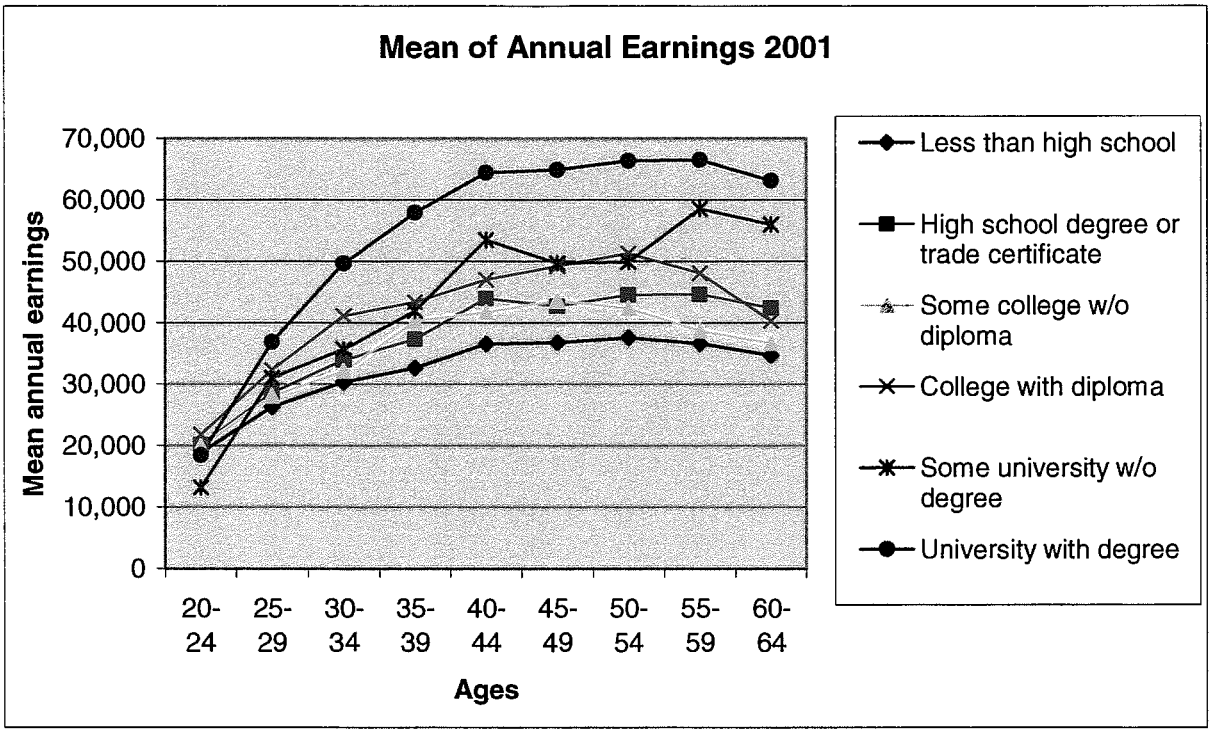
1991

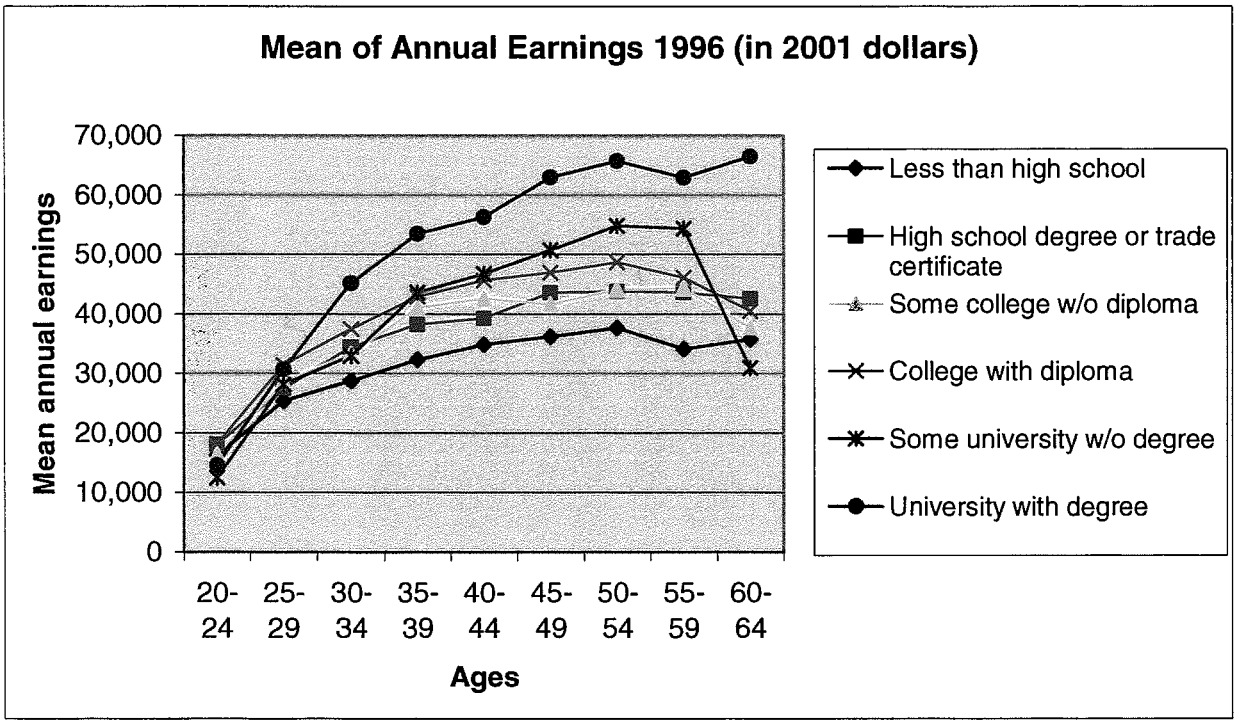
	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Less than high school	18,348	25,806	29,719	32,882	36,149	37,352	36,816	35,749	33,166
High school degree or trade certificate	19,976	28,997	36,507	39,775	41,973	45,217	45,046	43,178	40,615
Some college without diploma	18,119	27,664	35,009	39,540	43,080	41,764	48,106	41,831	40,187
College with diploma	21,431	33,328	39,229	43,700	47,240	46,018	48,314	42,682	39,999
Some university without degree	12,586	30,339	38,109	45,127	46,444	56,525	53,870	49,532	41,820
University with degree	15,166	34,170	47,331	53,179	60,224	67,808	72,109	68,221	70,582

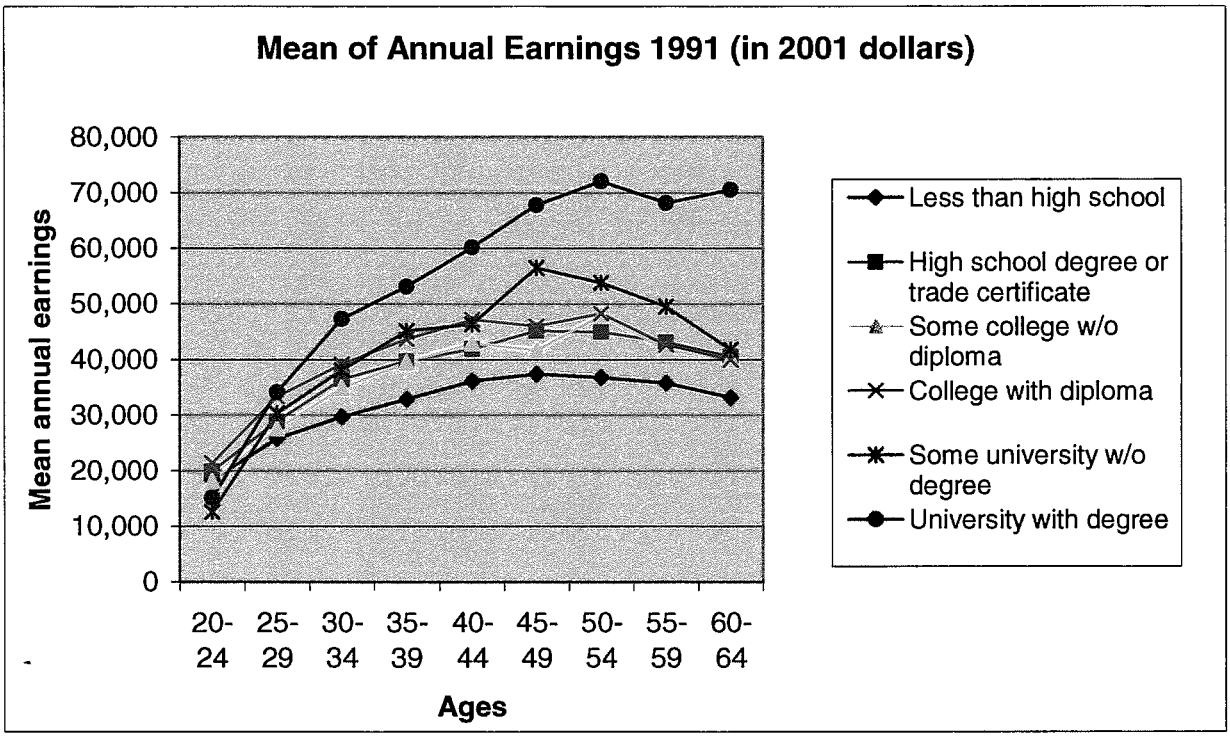
1986

	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Less than high school	18,365	28,042	33,002	34,826	38,276	40,792	38,322	36,576	34,462
High school degree or trade certificate	20,499	30,539	39,837	42,635	45,211	45,519	47,156	43,789	39,318
Some college without diploma	17,199	30,962	37,975	42,936	50,480	44,550	44,423	45,137	40,621
College with diploma	18,900	33,903	39,417	44,756	48,047	46,031	47,962	47,090	45,630
Some university without degree	13,619	29,909	42,231	43,798	56,606	54,894	59,562	51,334	44,514
University with degree	14,786	35,504	47,411	58,980	60,612	62,349	64,081	66,439	59,238

Figure 1: Annual Earnings by Education Levels in 2001 dollars






Mean of Annual Earnings 1991 (in 2001 dollars)
Mean annual earnings
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
Ages
Less than high school
High school degree or trade certificate
Some college w/o diploma
College with diploma
Some university w/o degree
University with degree

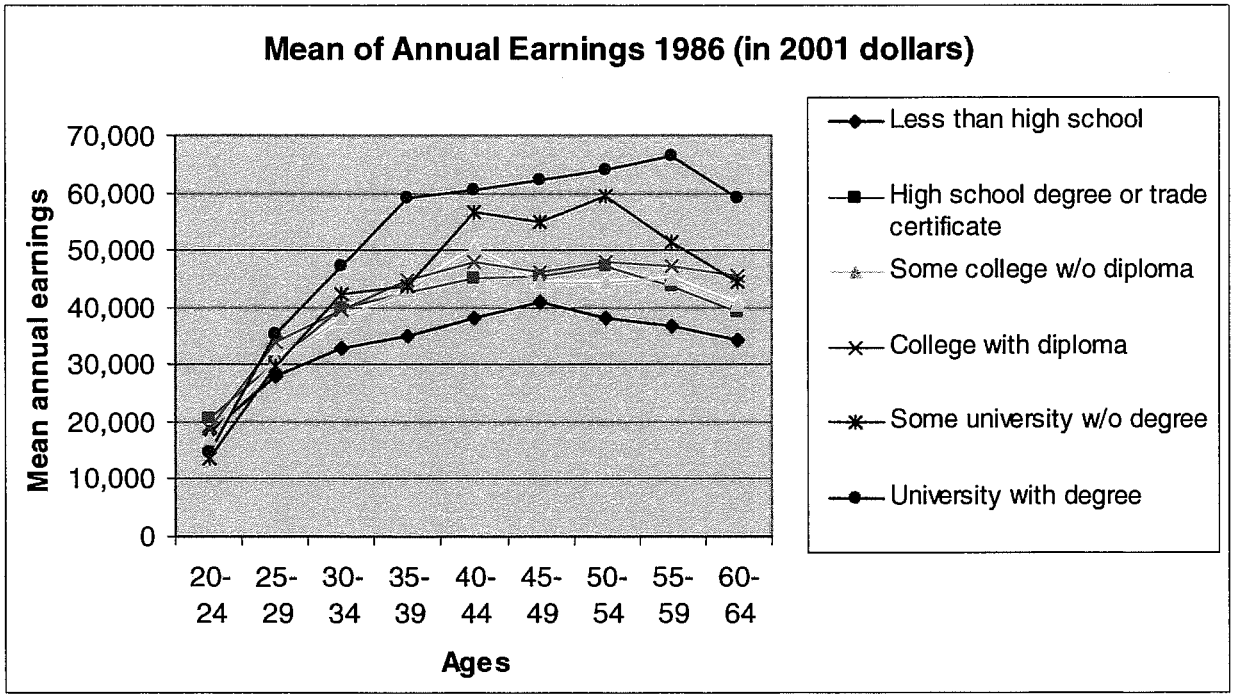

Mean of Annual Earnings 1986 (in 2001 dollars)
Mean annual earnings
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
Ages
Less than high school
High school degree or trade certificate
Some college w/o diploma
College with diploma
Some university w/o degree
University with degree

TABLE 2: Change in mean of log annual earnings of full-time workers

2001/1996									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	0.18	0.04	0.19	0.02	-0.01	0.06	0.03	0.07	0.03
High school degree or trade certificate	0.09	0.03	-0.04	-0.04	0.10	-0.06	-0.03	-0.05	-0.05
Some college without diploma	0.38	0.02	0.04	-0.15	-0.20	-0.08	-0.04	-0.17	-0.01
College with diploma	0.23	0.10	0.07	0.01	0.02	0.07	0.09	0.00	-0.02
Some university without degree	-0.03	0.14	0.00	-0.13	0.28	-0.14	-0.26	0.13	0.52
University with degree	0.30	0.19	0.13	0.07	0.07	0.00	-0.04	-0.01	-0.15

1996/1991									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	-0.17	0.01	-0.17	-0.09	-0.02	-0.10	0.00	-0.13	0.01
High school degree or trade certificate	-0.09	-0.12	-0.07	-0.05	-0.07	0.04	-0.06	0.02	0.15
Some college without diploma	-0.21	-0.01	-0.06	0.04	0.02	-0.05	-0.16	-0.13	-0.27
College with diploma	-0.26	-0.06	-0.05	-0.04	-0.07	-0.05	-0.05	0.07	-0.06
Some university without degree	0.00	-0.16	-0.18	0.07	-0.10	-0.19	0.09	-0.12	0.02
University with degree	-0.09	-0.17	-0.10	-0.07	-0.08	-0.09	-0.12	-0.10	-0.01

2001/1991									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	0.01	0.05	0.02	-0.07	-0.03	-0.05	0.03	-0.06	0.04
High school degree or trade certificate	0.00	-0.09	-0.11	-0.09	0.03	-0.02	-0.09	-0.03	0.09
Some college without diploma	0.17	0.01	-0.02	-0.10	-0.18	-0.13	-0.20	-0.30	-0.28
College with diploma	-0.03	0.04	0.02	-0.03	-0.05	0.01	0.04	0.08	-0.08
Some university without degree	-0.03	-0.02	-0.18	-0.06	0.18	-0.33	-0.17	0.01	0.54
University with degree	0.21	0.02	0.03	0.00	-0.01	-0.10	-0.16	-0.11	-0.16

TABLE 3: Age distribution within educational categories of full-time workers in Canada

2001									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	9.53%	9.25%	9.56%	13.42%	15.59%	14.55%	12.31%	10.52%	5.26%
High school degree or trade certificate	10.93%	8.92%	9.85%	14.65%	17.01%	15.58%	12.35%	7.83%	2.89%
Some college without diploma	17.43%	14.78%	14.91%	13.95%	13.51%	11.83%	6.49%	5.61%	1.49%
College with diploma	8.33%	12.15%	13.67%	15.93%	16.96%	13.48%	10.57%	6.16%	2.76%
Some university without degree	21.22%	13.93%	11.38%	11.50%	10.71%	13.32%	10.16%	6.07%	1.72%
University with degree	4.23%	13.20%	15.49%	16.29%	14.93%	13.66%	12.32%	7.29%	2.58%
All levels of education	9.16%	11.48%	12.58%	14.95%	15.66%	14.02%	11.43%	7.61%	3.11%
1996									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	7.78%	9.08%	12.98%	15.79%	13.11%	12.63%	12.10%	10.54%	6.00%
High school degree or trade certificate	9.92%	11.56%	16.05%	15.99%	15.41%	12.97%	8.83%	6.36%	2.92%
Some college without diploma	16.85%	16.35%	16.65%	15.25%	11.94%	10.33%	6.62%	3.51%	2.51%
College with diploma	7.20%	11.66%	19.25%	17.71%	14.47%	12.67%	9.08%	5.31%	2.66%
Some university without degree	25.37%	13.62%	11.38%	13.99%	11.19%	12.31%	7.28%	3.73%	1.12%
University with degree	4.67%	13.56%	16.61%	16.20%	15.31%	15.38%	10.47%	5.22%	2.56%
All levels of education	8.39%	11.88%	16.11%	16.28%	14.29%	13.26%	9.87%	6.53%	0.65%
1991									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	8.51%	12.21%	14.09%	12.41%	11.60%	11.81%	10.90%	10.77%	7.70%
High school degree or trade certificate	12.36%	16.21%	16.21%	15.01%	12.41%	10.16%	8.21%	6.02%	3.41%
Some college without diploma	19.29%	18.51%	16.70%	15.14%	11.42%	8.39%	4.93%	3.20%	2.42%
College with diploma	7.88%	16.67%	19.52%	17.07%	13.87%	9.65%	6.64%	5.43%	3.25%
Some university without degree	0.03%	14.42%	12.34%	14.53%	12.57%	8.19%	4.15%	3.92%	2.19%
University with degree	5.36%	13.81%	17.59%	17.97%	17.59%	12.63%	7.32%	5.16%	2.58%
All levels of education	10.05%	14.75%	16.44%	15.34%	13.55%	10.84%	8.04%	6.74%	0.56%
1986									
	20-24	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**
Less than high school	13.14%	13.68%	11.43%	10.64%	10.46%	11.11%	11.76%	10.70%	7.10%
High school degree or trade certificate	17.04%	16.52%	15.27%	13.55%	10.89%	8.55%	7.50%	6.83%	3.86%
Some college without diploma	20.85%	21.65%	16.73%	12.48%	8.90%	7.70%	4.52%	3.85%	3.32%
College with diploma	11.80%	19.05%	18.09%	16.10%	11.80%	8.32%	6.67%	4.50%	3.68%
Some university without degree	1.60%	14.31%	16.88%	16.51%	9.72%	7.71%	3.12%	2.57%	2.39%
University with degree	7.39%	14.83%	19.24%	19.88%	14.37%	9.78%	6.84%	4.45%	3.21%
All levels of education	13.63%	15.96%	15.47%	14.37%	11.40%	9.49%	8.21%	6.80%	0.63%

(The shaded columns represent groups that belong to the Baby Boom generation. Rows may not sum to 100% due to rounding errors)

leisure, and thus work less. However, the mean of earnings for the age of 50 to 64 does not always decrease in all of the Census years for people with a university degree. This suggests that there may be a higher demand for workers with university degree in some years which depends on the economic boom and bust.

In each level of schooling - university, college, and high school - the mean of earnings for people with a degree, diploma or certificate is always higher than that for people without a degree, diploma or certificate in every age group. This result is quite obvious because an individual with a certificate, diploma, or degree can always get a better job than if he or she does not have a certificate, diploma, or degree. Table 1 also shows that the people with higher educational attainment such as some university start at lower annual earnings at a young age, but their earnings grow much faster later. Thus, their average earnings over the life time will be higher than those of people with lower than university educational attainment. This may be another indicator that experience is a more important determinant of earnings for more highly educated people because as people with higher education accumulate more experience, their earnings increase faster. I will discuss this issue in more detail in the next subsections using age-earnings profile.

Table 2 provides changes in means of the log annual earnings for the 1991, 1996, and 2001 Censuses only which cover the period 1991-2001, because the 1986 Census is not comparable due to the unweighted earnings. Since change in the log of x is an estimate of the relative change in x, table 2 also provides the estimated percentage changes in means of annual earnings. This table does not show any clear relationship between age and earnings. In general, most groups experience a decrease in real earnings from 1991 to 2001. Most of this decrease in real earnings happened during the 1991-1996

period. It is clear that 1991-1996 was a recession period while 1996-2001 was a recovery period. The change in earnings during the period 1991-2001 is quite different from the period 1971-1981 in Dooley's (1986) study. While Dooley (1986) found growth in real earnings over 1971-1981, especially during 1971-1977, it seems to be a few cases where there is a growth in earnings in my samples during 1991-2001. For instance, the mean annual earnings increase by 17% for individuals with some college but no diploma at the age 20-24, 8% for individuals with a college diploma at the age 55-59, 18% for individuals with some university but no degree at the age 40-44, and 21% for individuals with a university degree at the age 20-24 and 54% for individuals with some university but no degree at the age 60-64. In particular, most of the decreases in earnings are marked for 1996/1991, especially for the youngest workers and university educated workers. In brief, the differences between the changes in means of the log annual earnings between my sample period and that of Dooley (1986) signals that my study using more current data will not yield the same conclusions with Dooley (1986).

There has been a trend in business process outsourcing (BPO) in the developed countries including Canada. According to Hijzen (2007), outsourcing is considered to be "driven by technological development in telecommunication and transportation" which allow us to shift works to other countries (p.190). Many companies choose BPO in order to stay competitive by taking advantage of cheap labour cost in developing countries. The top destinations for BPO industry are India, Canada, Philippines, China, Mexico, etc (Forey & Lockwood, 2007). It is interesting that Canada is one of the top destinations for BPO. In fact, the U.S does a lot of outsourcing to Canada due to similarities in language, culture, time zone, business practice, and education system. However, Canada also does a

lot of outsourcing to other Asian countries. The net effect of outsourcing in Canada is beyond the scope of this paper. The BPO activities include various back office functions, support services, and call centers for many industries such as IT, banking, insurance, etc (Forey & Lockwood 2007, p. 309).

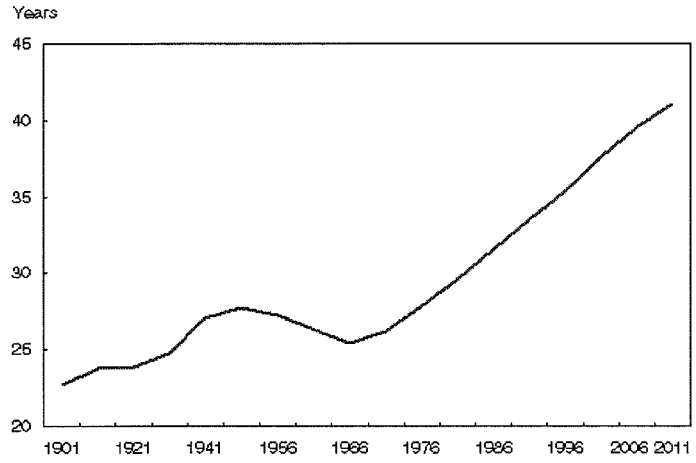
One of many concerns with BPO is job loss in many developed countries including Canada. Therefore, in my opinion, the outsourcing activities in 1990s contributed to job losses, and thus helped to cause the decrease in earnings for young and university educated workers in Canada in 1991-1996. However, outsourcing is only one of many factors that contribute to job losses and the earnings declining, because it may due to recession as well. Outsourcing appears to have hurt certain sectors such as IT and certain groups of workers such as low skilled workers. A study by Braun & Scheffel (2007) found evidence that "low skilled workers experience a decline in the union wage premium when working in industries with high outsourcing intensities" in Germany (p. 12). Since outsourcing appears to affect certain sectors and my analysis does not include a sector indicator, I decide to ignore the outsourcing issue for simplification purpose.

Finally, table 3 provides the age distribution within each of the educational attainment levels. The Baby Boom generation in Canada was defined to be born between the years 1946 and 1960 (Martel & Malenfant, 2007). As a result, they will be between the ages of 36-55 in 2001, 31-50 in 1996, 26-45 in 1991 and 21-40 in 1986. The shares of age groups falling into the above age ranges are the highest and this Baby Boom generation seizes about 60% of the total population in each Census year according to table 3. Among the groups belonging to the Baby Boom generation, the age group 25-29 has the highest share of 15.96% in 1986, the age group 30-34 has the highest share of

16.44% in 1991, the age group 35-39 has the highest share of 16.28% in 1996, and the age group 40-44 has the highest share of 15.66% in 2001. I have noticed that the peak share of the total Canadian population is moving from one age group in a Census year to the next age group in the next Census year because the Census is conducted in 5-year intervals and the age groups are also in 5-year intervals.

Table 3 also illustrates the aging of the Canadian population. This can also be seen more clearly in figure 2, which shows that the median age has been increasing over 1901-2006, except for the period when the Baby Boomers were born during 1946-1965. The median age is predicted to keep increasing and will be over 40 in 2011.

Figure 2: Median age 1901-2011 in Canada.



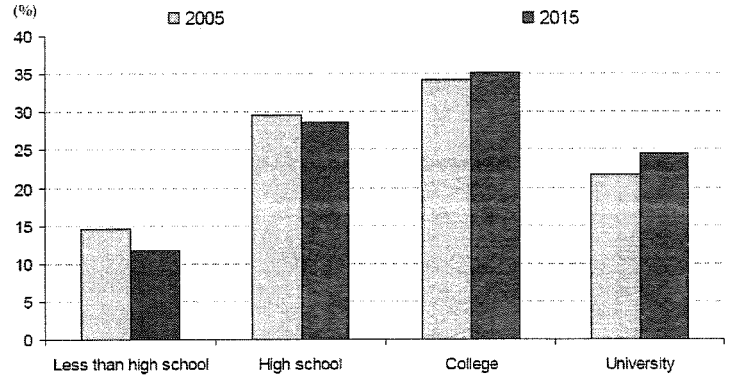
Sources: Statistic Canada, Age and sex profile. Retrieved July 22, 2007 from http://www12.statcan.ca/english/census01/Products/Analytic/companion/age/canada.cfm.

As we move from one Census year to the next, the percentage of population in each of the age groups within each of the educational attainment levels will be almost equal to that of the next age group within the same educational attainment level in the next Census year. For example, there are 16.61% of people at the age 30-34 in the university with degree educational attainment level in the 1996 Census. The corresponding figure in the 2001 Census is 16.29% of people at the age 35-39 in the

university with degree educational attainment level. From 1986 to 2001, within the university with degree educational attainment level, the age group which has the highest percentage share is getting older, from 30-34 in 1986, 1991 to 35-39 in 1996, 40-44 in 2001.

The age distribution within the educational level of university with degree has a similar pattern to the annual earnings. The percentage share increases with age until 30-34, 35-39 or 40-44 depending on the Census year, and then decreases. This pattern and the change in the age of the group with the peak share suggests that the age distribution is shifting toward older age and there will be a lack of workers with university degree in the next 15 to 20 years due to retirement. In my opinion, the growth rate of the labour force with university degree will need to be higher than that of other educational attainment levels in order to meet the demand for highly educated workers. This is also consistent with a recent Human Resources and Social Development Canada (HRSCD) report on the labour force. According to HRSDC (2007), the labour force growth rate was 1.7% between 1996 and 2005, and it is estimated to be 1.3% for 2006-2010 and 0.7% for 2011-2015. HRSDC (2007) also said this decrease in the growth rate of the Canadian labour force is caused by lower population growth and a declining rate of labour force participation as the Baby Boom generation retires. As a result, HRSDC (2007) projected that the share of the labour force with a college or university education will grow fastest (see figure 3).

Figure 3: Share of labour force by educational attainment, actual and projected.



Source: Human Resources and Social Development Canada (2007). Looking ahead: A 10-year Outlook for the Canadian Labour Market 2006-2015. Retrieved July 12, 2007 from http://www.hrsdc.gc.ca/en/publications_resources/research/categories/labour_market_e/sp_615_10_06/LA06-Supply-29Jan07.pdf.

2. Regression analysis

All my regression analyses will skip observations which have a cohort size equal to zero.[15] This zero cohort size is caused by my small sample, since it is only 10% of the Census microdata file, or about 0.1% to 0.3% of the total Canadian population. Moreover, my samples have some specific characteristics which were discussed in section III. In tables 4 and 5, I will provide the analyses for full-time workers while I will provide the analyses for both full-time and part-time workers in table 8. Table 6 and 7 are the analyses performed on unweighted data. Next, I will discuss the tables of regression results which used different models and data.

Table 4 provides estimates of Dooley's (1986) equation (1) using my data from the 1991, 1996 and 2001 Censuses. I also correct for heteroskedasticity using weighted least squares with the weight that was described in section IV. Most of the regressions except those for college without diploma and university without degree have explanatory power since the adjusted R-squared are about or higher than 0.8. The test of overall

[15] The numbers of observations range from 132 to 136 for the regression analyses performed on 1991, 1996 and 2001 Censuses and 176 to 179 for the regression analyses performed on 1986, 1991, 1996, and 2001 Censuses.

significance shows that all the regressions in table 4 have explanatory power as well. Although experience may be a more important determinant of earnings for the jobs which require higher education, it is expected to have a positive effect on earnings regardless of the type of job such as full-time or part-time, or requiring more or less education. Therefore, the coefficient associated with age, which is a substitute of experience, is expected to be positive for all of the educational attainment levels. Table 4 shows that these coefficients are negative for the educational attainment levels of high school with certificate, college with diploma, and university without degree. The t-ratios corresponding to the age variables show that they are not significant either for the educational attainment levels of high school without certificate, college without diploma and university with degree.

As I discussed in the literature review, theory predicts that cohort size has a negative effect on earnings. This prediction is the same as results in the fore-mentioned studies which were conducted in the 1970s and 1980s. Therefore, I also expect the cohort size to have a negative effect on earnings at all educational attainment levels. However, the coefficients associated with the log of the cohort size variable are positive for high school without certificate, and college with or without a diploma levels. This reversal of signs of the coefficient of log of cohort size coefficient may due to the inclusion of the interaction variables, which lead to collinearity. I will later look at the model without interaction terms to see if the log of cohort size coefficient yields negative sign.

The coefficients corresponding to the age and log of cohort size variables indicate that Dooley's (1986) model may not explain current Canadian earnings best. Therefore, I will next look at the results of the regression analysis using the modified model in table 5.

TABLE 4: Regression estimates of determinants of mean of log annual earnings using the 1991, 1996 and 2001 Census

(using Dooley's (1986) model and data for full-time workers)

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without diploma	with diploma	without degree	with degree
Constant	-6.70090	65.41800	-97.26600	26.36900	182.21000	-23.82300
	(-0.3569)	(1.939)	(-1.852)	(1.169)	(2.155)	(-0.8352)
Age	0.69129	-1.11840	3.16470	-0.27684	-3.65160	0.82409
	(1.747)	(2.265)	(3.195)	(-0.6039)	(-2.220)	(1.283)
Age squared	-0.00125	-0.00143	-0.00229	-0.00195	-0.00110	-0.00230
	(-11.65)	(-8.741)	(-8.09)	(-10.92)	(-2.37)	(-6.219)
Unemployment rate (U)	-3.06400	-0.66984	0.60040	-3.16660	-9.30520	-2.42830
	(-3.54)	(-0.219)	-0.30770	(-2.658)	(-2.591)	(-1.501)
Time (t)	0.00684	-0.02886	0.04763	-0.01331	0.08674	0.01206
	(0.7574)	(-1.749)	(1.818)	(-1.173)	(-2.060)	(0.8434)
Log of cohort size (lnCS)	0.07424	-0.04590	0.89341	0.70475	-0.10468	-0.00023
	(0.5239)	(-0.2809)	(-2.86)	(7.668)	(-0.2879)	(-0.6993)
Age*t	-0.00026	0.00063	-0.00139	0.00031	0.00188	0.56295
	(-1.37)	(2.577)	(-2.832)	(1.334)	(2.296)	(7.705)
lnCS*Age	-0.00449	-0.00014	-0.01966	-0.01588	0.00356	-0.01414
	(-1.694)	(-0.05227)	(-3.892)	(-10.35)	(0.5556)	(-7.653)
R^2	0.87070	0.79880	0.5625	0.8797	0.3949	0.934
adjusted R^2	0.86350	0.78760	0.5382	0.8730	0.3608	0.9303
Test of overall significance						
F-statistic	121.231	71.474	23.142	131.663	11.562	254.744
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	134	134	134	134	132	134

t-ratio in parentheses

TABLE 5: Regression estimates of determinants of mean of log annual earnings using the 1991, 1996 and 2001 Census

(for full-time workers using modified model)

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without diploma	with diploma	without degree	with degree
Constant	19.73700	-3.44210	50.73400	0.53239	-6.88270	34.45800
	(2.961)	(-0.1716)	(-4.006)	(0.05682)	(-0.3388)	(2.887)
Age	0.12122	0.12878	0.13430	0.18061	0.13951	0.14380
	(13.62)	(10.28)	(7.095)	(9.556)	(5.039)	(4.791)
Age squared	-0.00126	-0.00132	-0.00157	-0.00193	-0.00129	-0.00121
	(-11.68)	(-8.166)	(-6.714)	(-7.996)	(-3.923)	(-3.301)
Unemployment rate	-3.17500	0.86271	0.89052	-1.34910	-9.34060	-3.00490
	(-3.657)	(0.2812)	(0.4256)	(-0.8505)	(-2.568)	(-1.503)
Time (t)	-0.00529	0.00542	-0.02030	0.00361	0.00698	-0.01457
	(-1.667)	(0.5495)	(-3.205)	(0.7615)	(0.6912)	(-2.432)
Log of cohort size (lnCS)	-0.14600	-0.00398	-0.30251	-0.11320	0.09013	0.16390
	(-2.480)	(-0.07331)	(-5.442)	(-1.791)	(1.14)	(2.39)
R^2	0.8673	0.7863	0.4877	0.7774	0.36890	0.8967
adjusted R^2	0.8621	0.7780	0.4677	0.7687	0.34390	0.8927
Test of overall significance						
F-statistic	167.247	94.197	24.369	89.422	14.730	222.299
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	134	134	134	134	132	134

t-ratio in parentheses

Figure 4:

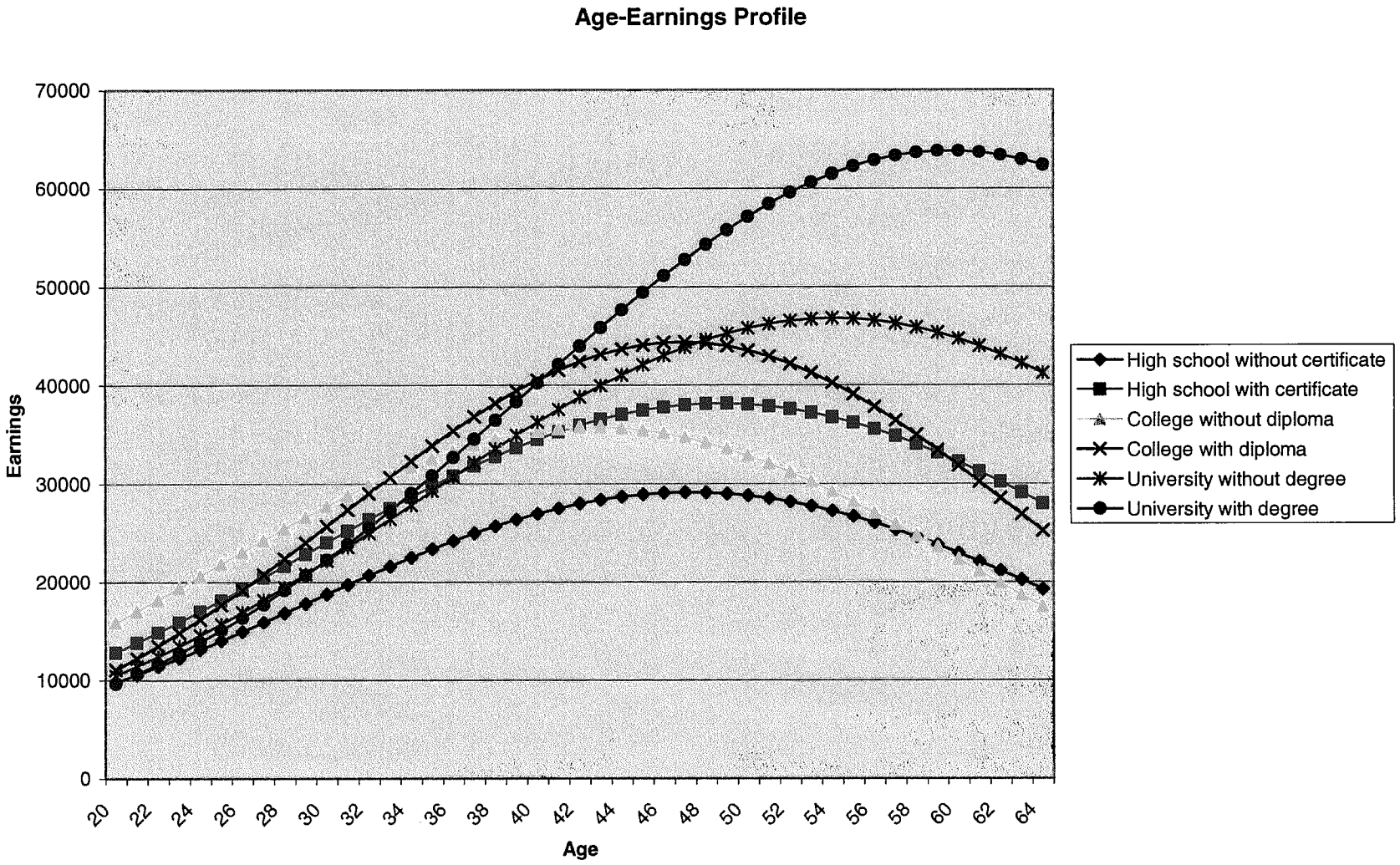

Age-Earnings Profile
Earnings
70000
60000
50000
40000
30000
20000
10000
0
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64
Age
High school without certificate
High school with certificate
College without diploma
College with diploma
University without degree
University with degree

TABLE 6: Regression estimates of determinants of mean of log annual earnings using unweighted data from the 1991, 1996, &2001 Census

(for full-time workers using modified model)

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without diploma	with diploma	without degree	with degree
Constant	61.22400	64.78800	23.14500	52.28200	81.65900	43.59200
	(4.306)	(3.139)	(0.6584)	(2.691)	(2.349)	(1.075)
Age	0.10638	0.12638	0.12602	0.07956	0.17076	0.17220
	(15.56)	(10.95)	(7.795)	(5.724)	(11.25)	(4.981)
Age squared	-0.00108	-0.00130	-0.00141	-0.00063	-0.00190	-0.00156
	(-13.14)	(-8.699)	(-7.292)	(-3.515)	(-11.04)	(-3.680)
Unemployment rate	-8.49810	-9.96940	5.44660	-5.62410	-15.91200	-2.80110
	(-3.560)	(-3.146)	(1.004)	(-1.832)	(-2.993)	(-0.4406)
Time (t)	-0.02638	-0.02835	-0.00769	-0.02253	-0.03631	-0.01875
	(-3.747)	(-2.778)	(-0.4427)	(-2.346)	(-2.113)	(-0.9347)
Log of cohort size (lnCS)	-0.02635	0.03832	-0.20142	0.33031	-0.34334	0.07891
	(-0.5144)	(0.6763)	(-3.858)	(5.659)	(-6.831)	(1.002)
R^2	0.8947	0.7623	0.4938	0.8615	0.6701	0.8604
adjusted R^2	0.8906	0.7530	0.4740	0.8561	0.6571	0.8550
Test of overall significance						
F-statistic	217.611	82.085	24.970	159.256	51.197	157.804
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	134	134	134	134	132	134

t-ratio in parentheses

TABLE 7: Regression estimates of determinants of mean of log annual earnings using unweighted data from the 1986, 1991, 1996, &2001 Census

(for full-time workers using modified model)

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without diploma	with diploma	without degree	with degree
Constant	31.62500	45.18800	1.17070	52.76100	-0.91437	3.66580
	(5.644)	(4.852)	(0.1061)	(5.46)	(-0.06994)	(0.289)
Age	0.10927	0.11929	0.11244	0.11131	0.17686	0.17717
	(18.73)	(13.45)	(8.078)	(10.1)	(12.16)	(8.069)
Age squared	-0.00112	-0.00122	-0.00118	-0.00103	-0.00183	-0.00163
	(-15.82)	(-10.66)	(-7.322)	(-7.243)	(-11.23)	(-6.022)
Unemployment rate	-4.67440	-7.20940	8.77640	-6.45170	-4.90960	3.36740
	(-2.969)	(-3.553)	(3.234)	(-3.3)	(-1.58)	(1.14)
Time (t)	-0.01180	0.01861	0.00309	-0.02277	0.00402	0.00090
	(-4.234)	(-4.009)	(0.5669)	(-4.741)	(0.6184)	(0.1445)
Log of cohort size (lnCS)	-0.00716	0.04442	-0.11045	0.21819	-0.06494	0.10372
	(-0.1543)	(0.8931)	(-2.25)	(4.517)	(-1.527)	(1.885)
R^2	0.8624	0.7190	0.4294	0.7984	0.5328	0.8723
adjusted R^2	0.8584	0.7109	0.4130	0.7926	0.5191	0.8686
Test of overall significance						
F-statistic	216.842	88.527	26.043	137.048	38.774	236.325
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	179	179	179	179	176	179

t-ratio in parentheses

TABLE 8 : Regression estimates of determinants of mean of log annual earnings using the 1991, 1996 and 2001 Census

(for both full-time and part-time workers using modified model)

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without diploma	with diploma	without degree	with degree
Constant	97.6440	3.6307	41.9240	62.0470	220.8100	110.9600
	(5.154)	(0.1491)	(1.202)	(2.396)	(3.094)	(3.165)
Age	0.1393	0.1611	0.1333	0.1933	0.1945	0.2043
	(16.90)	(11.72)	(7.667)	(11.87)	(6.17)	(7.681)
Age squared	-0.0015	-0.0017	-0.0016	-0.0021	-0.0021	-0.0020
	(-14.78)	(-9.659)	(-7.519)	(-9.903)	(-5.7)	(-5.847)
Unemployment rate	-11.6170	0.5181	-4.7485	-7.7428	-41.1950	-10.4730
	(-4.431)	(0.139)	(0.8665)	(-1.82)	(-3.711)	(-1.931)
Time (t)	-0.0441	0.0014	-0.0161	-0.0272	-0.1049	0.0529
	(-4.769)	(0.116)	(-0.9327)	(-2.134)	(-2.993)	(-3.043)
Log of cohort size (lnCS)	-0.1520	0.0288	-0.3047	-0.0868	-0.1875	0.1166
	(-2.370)	(0.4199)	(-5.516)	(-1.353)	(-2.123)	(1.442)
R^2	0.8632	0.7510	0.5374	0.7861	0.33170	0.8805
adjusted R^2	0.8578	0.7412	0.5194	0.7778	0.30540	0.8758
Test of overall significance						
F-statistic	161.467	77.200	29.742	94.094	12.606	188.585
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	134	134	134	134	133	134

t-ratio in parentheses

This table provides the estimates of the modified model using the 1991, 1996, and 2001 Census.

First, if we compare the R-square and adjusted R-square for each of the educational attainment levels, we will see that they decrease only a little from table 4 to table 5. This implies that the Age*t and lnCS*Age terms in Dooley's (1986) model do not contribute much to the explanatory power of the model. It also means that my modified model has explanatory power which is consistent with the test of significance results in table 5.

Second, the modified model appears to explain current earnings better because its variables seem to yield the expected signs. It appears that the variables yield expected sign without the interaction terms. Its regression results fit my primary expectation that the experience will have a positive effect on earnings for all levels of educational attainment. In table 5, the coefficients corresponding to the age variables are positive and those corresponding to age squared are negative for all of the educational attainment levels. The age and age squared variables are significant at the 5% of significance levels in all six regressions. Table 5 shows that keeping other things constant, if age increases by 1, the log of earnings will approximately increase by 12.12% for high school without certificate, 12.9% for high school with certificate, 13.43% for college without degree, 18.06% for college with degree, 13.95% for university without degree, and 14.38% for university with degree or higher. However, these are only approximations because age-squared also changes as age is changing. Therefore, I decided to analyze the age-earnings profile to see the effect of age, which is a substitute of experience, on earnings taking age squared into account.

The age-earnings profile for each educational attainment level is computed using the method introduced by Berndt (1991, p. 199-200). I fix all other variables in the equation (2) at the sample mean. I allow age to have a value from 20 to 60 and compute the log of earnings using the equation (2) with the coefficients equal to the estimations in table 5, then take the exponent of the log of earnings value to get earnings. These results are used to graph the age-earnings profile for each of the educational attainment levels in figure 4. This figure shows that earnings are concave functions of experience. The steeper slope of age-earnings curves suggests earnings grow faster with experience which in this case is represented by age. Alternatively, the steeper the slope of the age-earnings curve of an educational attainment level, the more important experience is as a determinant of earnings for people at that educational attainment levels. In general, the slope is steeper for higher educational attainment levels. Based on the figure 4, we can see that people with higher education start with a lower wage in their youth but their earnings grow much faster and keep growing over a longer period. Consequently, the peak of the curve is higher and more to the right for higher educational attainment levels. If we compare age-earnings profile of the university with degree level with other levels, this relationship is shown especially clearly. Therefore, experience is a more important determinant of earnings for highly educated workers, especially for those with a university degree, because they normally have jobs that require a great deal of knowledge and experience.

Third, the unemployment rate and time trend represent impact of the business cycle and how the Canadian economy can affect earnings in general. I assume that the unemployment rates are the same across sectors and educational attainment levels in each year because the unemployment rate is not available by sector and education level

categories. A higher unemployment rate indicates that the economy is not doing well overall and thus will lower earnings. Most of the coefficients corresponding to the unemployment rate variables in table 5 have the expected negative sign, except in the regressions for the high school with certificate and college without diploma levels. However, only two of the six coefficients, which are those in the regressions for high school without certificate and university without degree, are significant at the 5% significance level. These coefficients only provide raw estimations of the Canadian economy impact on earnings because the unemployment rates differs across sectors and there may be economical booms in some sectors while there are economical busts in other. The coefficient of the time trend variable does not have the same sign over all education levels. It is only significant at the 5% significance level in the regressions for college without diploma and university with degree levels. The negative coefficient of the time trend conflicts with the hypothesis that there is a shortage of university educated workers. However, this shortage will be more significant in the future when all of the Baby Boom generation retires. In my study period, most of them were still working and the demand for highly educated workers depends on the boom and bust of the Canadian economy.

Finally, let me discuss the log of cohort size in this modified model. After analyzing tables 1, 2, and 3, I think that cohort size may not always have a negative effect on earnings for all of the educational attainment levels. This expectation is different from Dooley's (1986) results that larger cohort size will lower earnings. The effect of cohort size on earnings should be negative if holding all other relevant determinants of earnings constant. Therefore, my different expectation of the impact of cohort size on earnings

may be explained by the following factors. First, the demographic structure is changing over time. In fact, the demographic structure of my study period 1986-2001 is different from Dooley's (1986) period 1971-1981. In my period 1986-2001 of analysis, a significant change in the labour force is the retirement of the Baby Boom generation instead of their entry into the labour force during 1971-1981. Consequently, there is a lack of labour supply especially for jobs which require college or university education. Therefore, I expect that the coefficient corresponding to the log of cohort size to be non-negative for the university education level. Second, not all of the relevant determinants of labour demand are included in my model. As fore-mentioned in section V, I have dropped some variables which represent nonlinearity in the model. Earnings are also determined by the boom and bust of various sectors in Canadian economy. I also performed the regression analyses for each of the educational attainment levels without considering the possible interaction between educational levels. According to Berndt (1991), individuals' abilities are likely to correlate to schooling and thus the omission of ability may lead to biased estimation (p. 166). In fact, my model did not include individual ability as a determinant of earnings. Third, the unemployment rate may vary with age, education level, or sectors.

Three of the coefficients of the log of cohort size are significant at the 5% significance level: those in the regressions for high school without certificate, college without diploma, and university with degree educational attainment levels. They are negative except for the university with or without degree levels. In comparison to table 4, the results regarding the log of cohort size are quite different. However, they confirm my

prediction that cohort size will not have negative effect on earnings for the university with degree level.

Table 5 indicates that if the log of cohort size increases by 1, the log of earnings will increase by 0.16390 for the university with degree educational attainment level. Alternatively, the elasticity of earnings with respect to the cohort size is 0.16390, which is quite small. This small elasticity means the shortage of university educated workers is currently small. This is because the majority of the Baby Boom generation is still working. In addition, the t-ratio corresponding to the log of cohort size for the university with degree educational attainment level indicates that it is significant. Although the coefficient of the log of cohort size in my model is not significant in most of the regressions, it confirms the predictions based on table 1, 2, and 3 that there is a lack of highly educated workers, especially with a university education. On the other hand, the coefficients corresponding to the log of cohort size are negative in four regressions for the high school and college levels, with or without degree. Another possible cause of the non-negative log of cohort size coefficient is the increase in demand for highly educated workers, which is not captured by this variable.

In comparing the regression analysis results using Dooley's (1986) model in table 4 to those using the modified model in table 5, the age variable performs better in the modified model since its coefficients have the expected positive sign and they are significant at the 5% significance level for all educational levels. The log of cohort size also performs better since it explains increase in demand of workers with university education. The coefficients of the unemployment rate variable are almost the same in both table 4 and 5 in terms of their signs and significance.

In summary, my estimated modified model has a different conclusion from Dooley's (1986) that larger cohort size will lower earnings. In fact, my results show this conclusion is only right for college and high school level. Cohort size actually has positive effect on earnings for people with university education. This confirms my prediction of the lack of highly educated workers and indicates that Canadians are not overeducated.

I also perform regression analyses on the unweighted data from the 1991, 1996, and 2001 Censuses (see the results in table 6). The data on earnings are not weighted using the weight variable provided in the Census because there is no weight variable in the 1986 Census. The cohort size is the number of observations in each sample. The purpose of doing this analysis is to see the effect of the weight variable on my estimation. Since the weight variable ranges from 30 to 33, its effect should not be significant. However, the results in table 6 show that the weight variable affects the performance of the log of cohort size variables for which the sign and significance changes from table 5 to table 6. This is surprising result which is different from my expectation that the weight would not affect the empirical analysis because it only varies a little across all observations. However, a little variation in the weight may cause a significant change in the cohort size since it is sum of the all the weights corresponding to each observation in a group by age and education. Therefore, the regression analysis results of the log of cohort size are different from table 5 to table 6. On the other hand, the results for other coefficients do not change much. These results in table 6 are similar in table 7 which provides the estimations of my modified model using the unweighted data from all four

Censuses. I can basically draw the same conclusion about the effect of experience on earnings in the modified model, but the cohort size effect on earnings is different.

Finally, table 8 provides the estimation of the modified model using the 1991, 1996, and 2001 Census. Unlike tables 4, 5, 6, and 7, which provide the regression analyses for full-time workers only, table 8 uses both full-time and part-time workers in my samples for the regression analysis. The process of preparing the data for these regressions was similar to those for table 4 and 5, but includes both full-time and part-time workers' earnings in calculating the mean of annual earnings by educational attainment level and age. In addition, the part-time wage used in the analysis does not include self-employment income according the Census definition of the income variable.

In table 8, the coefficients corresponding to the experience variable are positive and significant at the 5% significance level in all six regressions, and the magnitudes are close to those in table 5. It indicates that experience has a positive effect on earnings regardless of part-time or full-time jobs. In addition, the coefficients corresponding to the log of cohort size variable are significant in the three regressions for high school, college, university without certificate, diploma or degree and not significant in three regressions for high school, college, university with certificate, diploma or degree at the 5% significance level. In the regression for university with degree or higher level, the coefficient of the log of cohort size is positive, which is the same as the result in table 5.

VI. Conclusion

My paper has provided an updated picture of Canadian earnings. It also reveals the changes in the relationships between earnings, education, and experience of Canadian

men using four samples of data from the Canadian Census during the period 1986-2001. By comparing my results with the results in the literature review, I can see the differences between current earnings patterns and those of the 1970s and the 1980s.

Based on the empirical results and the analyses in section V, I can conclude the following points. First, higher education leads to higher average lifetime earnings. Second, experience always has a positive impact on earnings regardless of the educational attainment level and whether the job is full-time or part-time. However, the effect of experience on earnings is larger for jobs which require higher education. Third, cohort size does not always lower earnings. The cohort size has a positive effect on earnings for the university level and a negative effect on earnings for the high school and college levels. The positive impact of cohort size on earnings indicates a lack of labour supply for the jobs which require university education or too much workers with incomplete qualification for these jobs. The combination of cohort size and experience impacts on earnings leads me to another important result: Canadians are not overeducated. These results are different from Dooley's (1986) conclusions that there is a decline in relative earnings between people with a university degree and people with less than a university education and cohort size has a negative effect on earnings (p. 156). Consequently, he concluded that Canadians are overeducated. I instead study the age-earnings profile, descriptive table and regression analyses based on the human capital model and come to the opposite result regarding the overeducated issue. Fourth, the weight variable in the Census which is used to generate population estimates affects how cohort size is measured, and thus affects the regression analysis results.

In this paper, I mainly look at labour supply in order to focus on the impact of cohort size and experience on earnings. In fact, labour demand also has some effects on earnings and the labour force structure. Jobs which have higher labour demand normally have higher relative earnings. As a result, people will take this into account when making decisions about their education and job. Differences in labour supply and demand across sectors also affect earnings and flow of the labour from one sector to another. Therefore, I think labour supply, individuals' decision on educational investment, decisions on the type of jobs to pursue, the expectations on earnings, and differences across various sectors should be considered in analyzing the pattern of earnings.

References

Bank of Canada (2007). *Rates and Statistic: Inflation Calculator*. Retrieved June 9, 2007 from http://www.bankofcanada.ca/en/rates/index.html.

Berndt, E. R. (1991). The Practice of Econometrics: Classic and Contemporary (p. 150 – 223). Massachusetts, United States of America. Addison-Wesley Pub. Co., c1991.

Berger, Mark C. (1985). The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence. *The Journal of Political Economy*, Vol. 93, No. 3. (Jun., 1985), pp. 561-573.

Berger, Mark C. (1989). Demographic Cycles, Cohort Size, and Earnings. *Demography*, Vol. 26, No. 2. (May, 1989), pp. 311-321.

Braun, S. & Scheffel, J. (2007). Does International ***Outsourcing*** Depress Union Wages? Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, SFB 649.

Canadian Public Service Agency (2000). *Public Service Demographics and Surveys*. Retrieved May 29, 2007 from http://www.psagency-agencefp.gc.ca/hr-rh/psds-dfps/index_e.asp.

Computing in Humanities and Social Science, University of Toronto (2005). *Canadian Census Analyser: Microdata Note*. Retrieved June 1, 2007, from http://proxy.bib.uottawa.ca:2409/census/document/micronotes.html.

Connelly R. (1986). A Framework for Analyzing the Impact of Cohort Size on Education and Labor Earning. *The Journal of Human Resources*, Vol. 21, No. 4. (Autumn, 1986), pp. 543-562.

Dooley, Martin D. (1986). The Overeducated Canadian? Changes in the Relationship among Earnings, Education, and Age for Canadian Men: 1971-81. *The Canadian Journal of Economics / Revue canadienne d'Economique*, Vol. 19, No. 1. (Feb., 1986), pp. 142-159.

Dooley, Martin D. & Gottschalk P. (1984), Earnings Inequality among Males in the United States: Trends and the Effect of Labor Force Growth, *The Journal of Political Economy*, Vol. 92, No. 1. (Feb., 1984), pp. 59-89.

Dooley, D., Prause, J., & Ham-Rowbottom, Kathleen A. (2000). Underemployment and Depression: Longitudinal Relationships. *Journal of Health and Social Behavior*, Vol. 41, No. 4. (Dec., 2000), pp. 421-436.

Freeman, Richard B. (1979). The Effect of Demographic Factors on Age-Earnings Profiles. *The Journal of Human Resources*, Vol. 14, No. 3. (Summer, 1979), pp. 289-318.

Forey, G. & Lockwood, J. (2007). ''I'd love to put someone in jail for this'': An initial investigation of English in the business processing outsourcing (BPO) industry. Science Direct: English for Specific Purposes 26 (2007), pp. 308–326.

Hijzen, A. (2007). International Outsourcing, Technological Change, and Wage Inequality. *Review of International Economics*, 15(1), pp. 188–205.

Human Resources and Social Development Canada (2007). Looking ahead: A 10-year Outlook for the Canadian Labour Market 2006-2015. Retrieved July 12, 2007 from http://www.hrsdc.gc.ca/en/publications_resources/research/categories/labour_market_e/sp_615_10_06/LA06-Supply-29Jan07.pdf.

Lemieux, T. & Card, D. (2001). Education, Earnings, and the 'Canadian G.I. Bill'. *The Canadian Journal of Economics / Revue canadienne d'Economique*, Vol. 34, No. 2. (May, 2001), pp. 313-344.

Martel, L. & Malenfant, E. C. (2007). Canada: Baby Boomers remain the largest generation. Retrieve July 26, 2007 from http://www.thematuremarket.com/SeniorStrategic/canada_baby_boomers-9005-5.html.

Statistics Canada (2001). *2001 Census Public Use Microdata File: Individual File Users Documentation*. Retrieved June 8, 2007 from http://proxy.bib.uottawa.ca:2199/census/document/microdef2001.html.

Statistics Canada, Age and sex profile. Retrieved July 22, 2007 from http://www12.statcan.ca/english/census01/Products/Analytic/companion/age/canada.cfm.

U.S Census Bureau (2006). *Current Population Survey (CPS)*. Retrieved June 1, 2007, from http://www.census.gov/cps/.

United States Department of Veterans Affair (2007). *GI-Bill History*. Retrieved May 29, 2007, from http://www.gibill.va.gov/GI_Bill_Info/history.htm.

University of Regina: Data Library Services (1999). Weight Variable in General Social Survey (GSS): how should they be used?. Retrieved June 9, 2007 from http://uregina.ca/datalibrary/weight.html.

Vaillancourt, F. (1995). The Private and Total Returns to Education in Canada, 1985. *Canadian Journal of Economics*, vol. 28, no. 3, August 1995, pp. 532-54.

Vaillancourt, F. & Bourdeau-Primeau, S. (2002). The Returns to University Education in Canada, 1990 and 1995. Renovating the ivory tower: Canadian universities and the knowledge economy. 2002, pp. 215-40.

Vaillancourt, F. & Henriques, I. (1986). The Returns to University Schooling in Canada. *Canadian Public Policy / Analyse de Politiques*, Vol. 12, No. 3. (Sep., 1986), pp. 449-458.

Welch, F. (1979). Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. *The Journal of Political Economy*, Vol. 87, No. 5, Part 2: Education and Income Distribution. (Oct., 1979), pp. S65-S97.

Welch, F. & Ground, W. (1976). An Experience Imputation or an Imputation Experience. Mimeographed. RAND Corp. 1976.

Appendix A

This table provides the highest level of education for population 15 years of age or higher. There are 14 levels in total.

Code	Description	Counts	Includes
99	Not applicable	155,094	Persons less than 15 years of age
1	Less than Grade 5	14,056	
2	Grades 5 to 8	49,390	
3	Grades 9 to 13	138,894	
4	Secondary (high) school graduation certificate	91,053	
5	Trades certificate or diploma	22,605	
	College:		
6	Without trades or college certificate or diploma	41,694	
7	With trades certificate or diploma	43,085	
8	With college certificate or diploma	78,335	
	University:		
9	Without certificate, diploma or degree	28,393	
10	With university or college certificate or diploma	38,627	
11	With bachelor or first professional degree	68,615	
12	With certificate or diploma above bachelor level	10,346	
13	With master's degree(s)	17,323	
14	With earned doctorate	3,545	

Source: This appendix is obtained from Statistic Canada (2001). 2001 Cencus Public Use Microdata File: Individual File Users Documentation (p. 113). Retrieved June 8, 2007 from http://proxy.bib.uottawa.ca:2199/census/document/microdef2001.html.

My definition of highest levels of education based on the above table

Highest level of educations	Corresponding code in the Census
High school without certificate	1, 2, 3
High school with certificate	4, 5
College without diploma	6
College with diploma	7, 8
University without degree	9
University with degree	10, 11, 12, 13, 14

Appendix B

TABLE 9: Condition indexes

	Education Attainment Levels					
	Highschool		College		University	
	without certificate	with certificate	without degree	with degree	without degree	with degree
highest CI	5401200	1851	1863	2038	1863	2152

TABLE 10: Auxiliary R-squared

	Education Attainment Levels					
	Highschool		College		University	
Independent variables	without certificate	with certificate	without degree	with degree	without degree	with degree
AGE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
AGE^2	0.9932	0.9951	0.9939	0.9972	0.9946	0.9980
U	0.1189	0.1297	0.0905	0.0560	0.2838	0.0379
Time (t)	0.9386	0.9376	0.9156	0.9170	0.9203	0.9227
AGE*t	1.0000	1.0000	1.0000	1.0000	1.0000	0.9141
lnCS	0.9470	0.9782	0.0989	0.9492	0.9817	0.9141
lnCS * AGE	0.9993	0.9796	0.9972	0.9965	0.9957	0.9953

These results in table 9 and 10 were obtained from the tests performed on regressions in table 4

Table 11: Tests for heteroskedasticity

	Education Attainment Levels						
	Highschool		College		University		
Independent variables	without certificate	with certificate	without degree	with degree	without degree	with degree	
Goldfeld-Quant test							
test statistic	607.9000	3.0690	2.8310	6.2070	194.5000	7.6870	
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
White test (with X, squares of X varibales, all cross product of independent variables)							
based on R^2							
test statistic	—	—	—	—	—	—	(*)
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
based on SSR							
test statistic	—	—	—	—	—	—	(**)
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

These results in table 11 were obtained from the tests performed on regressions in table 5

Note: (*), (**) the test statistics for White test are so large that their exact figures are not provided in Shazam

X is a vector of independent variables