The Effect of Education on Regional Economic Performance in Canada

By Min Fang

Student number: 4371290

Major paper 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 2008

The Effect of Education on Regional Economic Performance in Canada

Abstract

Using data collected from the 2001 Canadian Census Profile in E-STAT, the effect of education on regional economic performance is examined at the Canadian census division level. The positive effect of the educational attainment of the census division population on average employment income is quantified in a linear equation. The results of the analysis show that a higher level of educational attainment will generate a greater return to education. The increment in education increases average employment income for all provinces, thus affecting regional disparities in Canada. Disparities across provinces and across census divisions both become smaller.

I. Introduction

As human capital theory has developed, more concern has focused on the role of education in the economy. Education is considered to be an important investment in human capital, which could benefit an educated individual over a whole lifetime. More highly educated people are more productive than the less educated ones when they enter the labour market. However, many studies of education mainly focus on its effect on the earnings of individuals. Total years of schooling and the highest level of schooling are the two variables most often used in econometric models to represent the individual's education level. For example, Sanmartin's (2001) study shows that each additional year of education implies an increase in earnings of 6.7%.

Because of the scarcity of data on regional economic performance, studies of the effect of education at the regional level are much fewer in number than those at the individual level. However there are still many researchers who have devoted their efforts to this field. Alonso-Rodriguez (2001) studies the effect of education on the differences between high and nonhigh income countries. Examinations of disparities between regional economies have been undertaken by many researchers in different countries such as Spain, Mexico, and the U.S.[1]

Since there are few papers that focus on the effect of education on regional economic performance from a Canada perspective, I have chosen to study the relationship between the highest level of schooling and average employment income in

[1] See Galindo Martin and Alvarez Herranz (2004), Arellano and Fullerton (2005), and Sloboda (1999).

Canada. Data on about 288 census divisions in ten provinces and three territories are collected from the 2001 Canadian Census Profiles. This data source provides information on not only average income for each census division, but also on the educational attainment of the population of the census division, making it an excellent data source for this study.

My paper is organized as follows. Section II reviews previous studies on education and its effect on both the earnings of individuals and regional economic growth. Section III presents the data and the variables chosen to explain regional disparities in average employment income. In section IV, I will use OLS to estimate the model, do some diagnostic tests and discuss the regression results. Then in Section V, the descriptive statistics of the education level variables are discussed and the regression output is used to simulate the impact of a higher level of educational achievement in the 10 provinces of Canada. Finally, Section VI concludes the paper.

II. Literature Review

Prior to my paper, many researchers have made great efforts to evaluate the return to education. The effects of education on individuals have been carefully studied for a long time. People benefit from the investment in human capital and perform more productively if they have a high level of educational attainment. In recent decades, there has been increasing concern about the effect of education on regional economies. Researchers have used different data and models to examine disparities in per capita income performance caused by regional differences in the educational attainment of regional populations.

One of the most important benefits of education is that individuals can be more productive and earn a higher income after participating in the labour market, which is the key to understanding the link between investment in education and regional average income differences. Thus regional economic performance is affected by the educational attainment level of the region's population. Therefore reviewing the effect of education on individuals' earnings is necessary before examining the impact of education at the regional level. In the following, I will examine effect of education on individuals and on the regional economy.

1. The effect of education on individuals' earnings

As early as the 1960s, Becker (1964) emphasized the role of education in human capital development. Investment in human beings, like investment in other fixed assets such as buildings and equipment, can generate a stream of future benefits. A higher level of educational achievement improves human capital, provides higher productivity and results in income improvements.

Many empirical studies have tested the predictions of Becker's human capital model. For example, Tilak (1989) classifies the household into three groups (poorest 40 percent, middle 40 percent, and richest 20 percent) and collects data on 30 developing and developed countries. He examines the relationship between the income shares of different household groups and the rate of return to primary and secondary education. He finds that the higher income groups in the society benefit from a higher rate of return to any level of education, and rapid expansion of the educational system may reduce income inequalities. Card (1999), using pooled micro data from the 1994, 1995, and 1996 March Current Population Surveys in the U.S. (including 102718 men and 95360 women), examines the log-linear human capital earning function developed by Mincer (1974). His results support the idea that education has a positive effect on individuals' earnings for both men and women. Furthermore, he suggests that the conventional human capital earnings function is alive, but it is too parsimonious to fully characterize the joint distribution of earnings, age and schooling. The positive effect of education on earnings is also mentioned in Lemieux (2001).

In recent decades, more researchers have tried to quantify the effect of an additional year of education on earnings. As mentioned in Sanmartin (2001, 33), "The model used by most studies is the one developed by Mincer (1974), who proposes a linear relation between the logarithm of earnings and the number of years of education. Heckman and Polachek (1974) offer empirical evidence in favour of the log-linear functional form against alternative models." But this linear form assumes that the rate of return is the same for all education levels. Instead of assuming the return to education is constant, Sanmartin (2001) uses dummy variables for each year of education instead of a specific functional form. Her research shows that the effect of education on earnings is not statistically significant before the completion of secondary education and from then on the relationship can be considered linear. Each additional year of education implies an increase in earnings of 6.7% in Spain.

Jorgenson and Fraumeni's (1993) study shows that more educated or better trained people are more productive than less educated or poorly trained people in the U.S. Both education and training are investments in human capital. Their paper estimates the value of increments in lifetime incomes due to increases in educational attainment by comparing the incomes of individuals of the same age and sex with different levels of education. The impact of higher educational attainment is not only shown in earning power in the labour market but also revealed in the value of nonmarket activities such as parenting or enjoyment of leisure.

Knowing that returns to education increased during the 1980s, and hence less-skilled workers lost ground relative to more-skilled ones, Enchautegui (1993) uses micro data to study what happened to the employment and wages of Puerto Rican men in the northeast of the U.S. She finds that they too benefited from educational upgrading, increasing returns to education, and wage growth during the 1980s. When combined with the other papers reviewed, her paper confirms that wages increase with educational attainment, implying that regional economic performance should depend on its human capital stock. The empirical evidence for this hypothesis is reviewed in the next subsection.

2. The effect of education on the regional economy

Modern economic growth theory emphasizes the effects of human capital on the regional economy. In addition to human capital, other variables such as population density, public investment, and income distribution are also introduced into the analysis at a national or regional level. But regional studies of the effect of education are still less common than those at the individual level because of the scarcity of data on education at the regional level.

Different regional levels are chosen to be the target geographic region in analyses of the education effect. Alonso-Rodriguez (2001) studies the effect of education on a national level. Using the World Bank's income categories, the relationship between high and nonhigh income countries and a set of indicators of economic growth is studied in his paper. Using a logistic regression model, he shows that expenditure on education and the net primary enrollment ratio for girls and boys are the most relevant factors helping low income economies to move in the direction of high income countries.

In Canton (2007), data on educational attainment for 95 countries are taken from the Organization for Economic Co-operation and Development (OECD) website. His research shows that "econometric estimations of the social returns to education using high-quality data on educational attainment and a macro Mincer relationship between human capital and years of schooling yield the plausible results" (p. 465). The results indicate that if the average education level of the labour force increases by one year, labour productivity will increase by 7%–10% in the short run and 11%–15% in the long run.

Some researchers have examined the effect of education on economic performance in different regions in one country. Galindo Martin and Alvarez Herranz (2004) use data from the Spanish Statistical Institute to examine whether a better education process and social capital can improve the economic growth process in 19 Spanish Regions, and whether it can explain the existing differences in per capita GDP between those regions. Purely education based indicators of the composition of the finished studies level are used to measure human capital in their model. They find that the externalities from a better education could enhance the economic growth process in different regional areas, and show that human capital is an important factor that can improve the economic growth process in regions. Therefore, economic policies should focus on improving education levels.

Arellano and Fullerton's (2005) study employs Mexican census data for the year 2000 for all 31 states and the Federal District of Mexico City to quantify regional income performance. They use the proportions of the population 15 years or older who have completed different levels of education as explanatory variables to examine the relationship between per capita income and education. The state population density per square kilometer is also an independent variable included in their model. They utilize per capita Gross State Product (GSP) as the dependent variable and find that "for each additional percentage point of the population who attends at least some primary school, real per capita GSP rises by slightly more than 29 pesos. For each additional percentage point increase in the number of persons who successfully complete elementary school, real per capita GSP rises by approximately 97 pesos. For each additional percentage point gain in the number of persons who study beyond the sixth grade, real per capita GSP in Mexico grows by more than 504 pesos" (p. 236). Their results confirm the strong links between education and incomes across Mexico. Then, at the end of their paper, they use the estimated coefficients to simulate the potential gains associated with improved educational attainment.

In addition to the papers mentioned above, Domazlicky et al. (1996), Sloboda (1999) and Fullerton (2001) attempt to quantify the impact of educational attainment on per capita personal income at a lower level - the county level - in the United States. Many different model specifications have been used in these studies.

Domazlicky et al. (1996) uses an econometric model to measure the cost of high school noncompletion in southeast Missouri counties. His sample includes 115 counties in the state of Missouri. Per capita personal income is collected from the 1994 edition of the Regional Economic Information System. Data on educational attainment are from the 1990 Census of Population. The dependent variable used in his paper is the per capita personal income of a county. Three measures of educational attainment for adults over the age of 25 are used in his analysis: the percent who dropped out of high school; the percent who acquired some college, but did not complete their degree; and the percent who have completed a bachelor's degree or a professional degree. Three other variables are also included in the model: the female labour force participation rate, the percentage of the population that is less than 18 years, and the percentage of the population that is older than 64 years. In addition, he also used two dummy variables to indicate the counties in two different metropolitan areas. His study shows that the percent that has completed a bachelor's degree has a significant positive effect on per capita personal income. Furthermore, counties which are located in metropolitan areas exhibit higher per capita personal income than the others.

Sloboda's (1999) paper uses a linear regression model and a nonparametric regression model to examine the impact of education on regions. The sample for his analysis is comprised of 102 counties in Illinois. All data for the regressors came from the 1990 Census of the Population, Social and Economic Characteristics Report. In this paper, the specification of the linear regression model follows the one used in Domazlicky et al. (1996). The dependent variable and explanatory variables are all the same, except that he uses one dummy variable to represent counties in the Chicago area instead of the two metropolitan dummy variables in Domazlicky et al. (1996). The results of his regression suggest that the coefficients of the three education level variables are significant at the five percent level, but the other four explanatory variables do not have a significant effect. However, educational attainment does affect regional economic performance. In the last part of his paper, the nonparametric model results indicate that the parametric specification provides strong explanatory power.

Fullerton's (2001) study collects cross-section data for all 254 Texas counties from the 1990 Census and from the Regional Economic Information System. The percentages of adults 25 or older who have completed schooling at several different education levels are used to measure the educational attainment levels of the population. In order to emphasize the location of different counties, they includes a "border" variable which indicates that a county is located along the border between Texas and Mexico in addition to the educational attainment level variables in his econometric model. Two variables to capture the age structure of the population (as in Domazlicky et al. (1996)) and three variables that characterize the language skills of the population are also included in the model. Their study shows that "border county per capita income falls below other state regions by about $3,200" (p. 7) and "reduction of the dropout rate to a level commensurate with the rest of the state would have potentially increased income per border resident by more than $2,600 in 1990" (p. 8). This implies both border counties and other regions will benefit by reducing secondary school dropout rates.

Coulombe and Tremblay have devoted great efforts to studying the effect of education on the growth of provincial economies in Canada. In Coulombe and Tremblay (2001, p. 176), they show that an important part of the convergence of per capita income across the Canadian provinces can be explained by the dynamic accumulation of human capital. The indicators of human capital used in their paper are derived from Statistics Canada Census data on educational achievement. Their empirical analysis also suggests that the "best proxy for the aggregate human capital stock comes from indicators based on the population of both sexes and males, 15 years and over or 25 years and over, with a university degree."

In Coulombe and Tremblay (2007), they used pooled time-series (1951–2001) cross-sectional (the 10 provinces) empirical models (TSCS) to estimate the mean effect of human capital on aggregate provincial per capita income in Canada. Two measures of human capital are used in their analysis: the first is university achievement, measured as the percentage of the working-age population holding a university degree; the second one is an indicator of skills based on literacy test scores. The empirical analysis provides support for the view that the disparities in human capital can explain a large proportion of the per capita income disparities across provinces in Canada. Both university achievement and skill levels exert a positive and significant effect on per capita income levels and on relative provincial income.

Although Coulombe and Tremblay (2007) look at the effect of human capital on regional disparities at the provincial level, there is little research on the relationship between per capita income performance and the highest level of schooling at a more disaggregated level in Canada. Since the U.S. and Canada have many characteristics in common, in the next section, I will use a model specification and estimation method similar to those used in Domazlicky et al. (1996), Sloboda (1999), and Fullerton (2001) for the U.S. to estimate a linear equation which shows the relationship between educational attainment levels and regional average income in Canada.

III. Data and Model Specification

1. Data

The data used in this paper are Canadian aggregate data for geographic regions drawn from the 2001 Census conducted by Statistics Canada. All the data were retrieved from Statistics Canada's E-STAT educational resource.[2]

In Canada, there are only ten provinces and three territories. Thus if this level of aggregation were used, the sample would only contain thirteen observations, which is too small to do an interesting analysis. Therefore I chose a lower regional level to be my target geographic region - the census division. According to the definitions in Statistics Canada (2004, p. 225), "census division" (CD) is "the general term applying to geographic areas established by provincial law, which are intermediate geographic areas between the census subdivision and the province." The census division includes "divisions, counties, regional districts, regional municipalities and seven other types of geographic areas made up of groups of census subdivisions." In Prince Edward Island (PE) there are three census divisions, while in Nova Scotia (NS) there are eighteen and in New Brunswick (NB) there are fifteen. Quebec (QC), Ontario (ON) and British Columbia (BC), which are the three biggest provinces in Canada, have 99, 49 and 28 census divisions respectively in their provincial regions. There are six regions in total in the three territories (Yukon Territory, Northwest Territory and Nunavut).

In Newfoundland and Labrador (NF), Manitoba (MB), Saskatchewan (SK) and Alberta (AB), administrative units such as counties do not exist. Therefore, census divisions have been created by Statistics Canada in co-operation with these provinces. There are ten, twenty-three, eighteen and nineteen census divisions in these provinces

[2] The data can be seem form the E-STAT website: http://estat.statcan.ca /cgi-win/cnsmcgi.pgm .

respectively. In total, there are 288 census divisions in Canada, i.e., 288 counties. Thus I have 288 observations in my sample for my econometric analysis.

For each census division, information on a total of 1709 characteristics are available in the 2001 Census of Population profiles. I use only the information on census division name, province code, land area, total population, population 15 years and over, population 65 years and over, population by knowledge of official languages (English, French, bilingual and no official language), population 20 years and over by highest level of schooling (less than grade 9, grade 9 to 13, trade certificate, college and university), and average employment income of the population 15 years and over.[3] However, not all of these variables can be used directly in the model. In the next subsection, the construction of more new explanatory variables using these variables will be discussed.

2. Model Specification

When faced with the problem of how to evaluate regional economic performance, the World Bank favors Gross National Income (GNI) for comparing the relative size of economies and uses it to classify countries in low, middle and high-income categories (World Development Indicators 2007). According to the World Bank, GNI is a good indicator of economic capacity. But at the same time, GNI alone cannot measure average social welfare. So GNI per capita is also used by the World Bank to classify economies and evaluate regional average economic performance. However, the regional equivalent

[3] See Appendix A for a list of the variables obtained from the 2001 Census Profiles.

of Gross National Income is not available for census divisions. Total income at this level is also not available. But employment income is the component of total income that is the most closely related to education levels, and I can find data on average employment income at the census division level. So the average employment income for the population 15 years and over with employment income is used as the dependent variable in this paper.

Before actually estimating the model, new explanatory variables related to education, population density and other characteristics have to be constructed. As in both Fullerton (2001) and Arellano and Fullerton (2005), the percentages of the population that have attained given education levels are used as independent variables. In constructing these variables, the total population 20 years and over is used as the denominator.[4] For each level of education, the population for whom this is the highest level of schooling is used as the numerator. Therefore, the first educational variable (PLES9) is the percentage of the population 20 years or older whose highest level of schooling is less than grade 9. Additional educational variables are constructed in a similar fashion: the percentage of the population that completed grade nine but did not complete high school (PNOHIGH), the percentage of the population whose highest educational qualification is a high school graduation diploma (PHIGH), the percentage of the population whose highest qualification is a trades certificate or diploma (PTRAD),

[4] In the 2001 Census Profiles, the educational attainment levels are reported for the total population 20 years and over by highest level of schooling based on a 20% sample. See Appendix A for details.

the percentage of the population who attended college but did not obtain a certificate or diploma (PNOCOLL), the percentage of the population whose highest level of educational attainment is a college certificate or diploma (PCOLL), the percentage of the population that attended university but did not receive a degree (PNODEG) and the percentage of the population whose highest level of attainment is a university degree (PDEG). To avoid perfect multicollinearity, I do not include PLES9 in my regression model.[5] These educational variables are the most important explanatory variables in the model, because I want to examine the relationship between them and the census division-level average employment income.

Population density is another important factor which could affect average income. It is widely recognized that urbanization effects will cause metropolitan economies to exhibit higher than average levels of productivity within national and regional economies. Fullerton's (2001) study uses population density (population per square kilometer) as an explanatory variable in his model; the higher is the population density the more urbanized is the region. However, other studies use different variables to measure population density and/or urbanization. Arellano and Fullerton (2005) define an urban area as one whose population is larger than 599999. According to the definition in Statistics Canada (2004, p. 262), "the general concept of an urban area is that of an area containing a dense concentration of population. Statistics Canada defines an urban area

[5] Since the educational attainment variables sum to 100, one must be omitted to avoid perfect multicollinearity.

as an area which has attained a population concentration of at least 1,000, and a population density of at least 400 per square kilometer." Territories lying outside urban areas are considered to be rural areas. Taken together, urban and rural areas cover all of Canada.

After using the total population in each census division and the corresponding land area to compute population density, I construct three dummy variables to represent urban areas of different size. URBAN400 is equal to 1, if the area's population density is greater than or equal to 400 people per square kilometer and population is larger than 1000. URBAN200 represents a population density greater than or equal to 200 but less than 400, while URBAN100 represents a population density greater than or equal to 100 but less than 200.

Geographical location can also have a huge influence on average income. The empirical evidence supports the view that geographic attributes influence population levels and educational attainment. Iyigun's (2005) study shows that favorable geographic characteristics, such as latitude and terrain, should lead to higher population densities and have the potential to affect many household choices, such as investment in education. Later, these choices will influence an economy's future economic growth. Due to the importance of geographical location, I created some variables to characterize the location of each census division.

First, I construct ten dummy variables for the provinces. The dummy variable NF equals 1 if the census division is located in Newfoundland (province code equals 10).[6] PE, NS, NB, QC, ON, MB, SK, AB and BC are dummy variables that represent the other nine provinces respectively. Areas outside these ten provinces are in one of the three territories. Since all three territories are located in the Canadian north and together contain only six census divisions, I do not distinguish between them in my empirical analysis.

Second, I construct one dummy variable (ADJUS) which indicates if the census division is in the border area adjacent to the U.S. As mentioned earlier, Fullerton (2001) includes a Mexican border variable in his model. In Canada, census divisions located in the border area adjacent to the U.S. are closer to a major market, so they are expected to have a higher average employment income than others. In my model, ADJUS equals 1 if a census division is adjacent to the U.S. The information about area location is collected from the reference map provided by Statistics Canada.[7] This map carefully shows boundaries, names and codes for all census divisions within each province and territory.

The percentage of the population 65 and over is also included in the model as an independent variable. There is an obvious relationship between the proportion of retirement-age adults and average employment income. Most of these people drop out of the labour market. Although they may still have a fixed company pension, this is not

[6] According to the variable definitions in the Census Dictionary, the province codes are as follows. 10 Newfoundland and Labrador(NF), 11 Prince Edward Island (PE), 12 Nova Scotia (NS), 13 New Brunswick (NB), 24 Quebec (QC), 35 Ontario (ON), 46 Manitoba (MB), 47 Saskatchewan (SK), 48 Alberta (AB), 59 British Columbia (BC), 60 Yukon Territory (YK), 61 Northwest Territories (NT), 62 Nunavut (NU).

[7] The census reference maps are available at: http://geodepot.statcan.ca.proxy.bib.uottawa.ca/Diss/Maps/ReferenceMaps/n_cd_e.cfm

included in employment income. Hence the more people older than 65, the less productive the regional population is likely to be, and the lower average employment income will be.

A variable which could measure each census division's language characteristics is also included. English and French are the two official languages of Canada. The ability to speak at least one official language is expected to increase average employment income. The proportion of the population which has knowledge of the official languages (including English, French and bilingual) is called PLANG.[8] The percentage of the population which does not have official language skills is called PNOL. Since PNOL and PLANG sum up to one, I just use PNOL in the model to avoid perfect multicollinearity.

The descriptive statistics for the sample are shown in Table 1. The mean of average employment income is $31757, while the maximum average income is $45835 and the minimum income is $15974. On average, 17.42% of the population 20 years and over has completed grade nine but did not complete high school. This is the largest educational category. Only 6.36% of the population attended college but did not obtain a certificate or diploma, which is the smallest educational category. The percentages of the population whose highest level of attainment is a college diploma or a university degree are the second and third highest of the educational variables. Five percent of the census

[8] Initially, I wanted to use three variables (PENG, PFRE AND PBIL) to represent the percentages of the population who speak English, French or are bilingual respectively. But after doing this, I discovered that there is a strong negative relationship between PENG and PFRE, which implies a severe multicollinearity problem exists. So I chose to use PLANG to indicate the proportion of the population with official language skills. The coefficient of correlation between PENG and PFRE is -0.943.

divisions, actually 14 census divisions, are large urban areas, while 4% of the census divisions are mid-sized urban areas (12 regions). Another 4% are small urban areas (11 regions). The remaining regions are considered to be rural areas. Twenty-two percent of the census divisions are adjacent to the U.S. Furthermore, 12.01% of the population in Canada are people who are 65 years and over and 1.49% of the population can not speak either official language. More descriptive statistics about the education variables at the provincial level will be discussed in section V.

After defining all the explanatory variables, the model is constructed as follows:

$$y_i = X'_i\alpha + Z'_i\beta + P'_i\lambda + R'_i\gamma + \varepsilon_i ,$$

where

y_i = average employment income (in dollars) of the population 15 years and over with employment income in area *i*;

X_i = a vector that contains the education level variables PNOHIGH, PHIGH, PTRAD, PNOCOLL, PCOLL, PNODEG and PDEG in area *i*;

Z_i = a vector that contains the population density variables URBAN100, URBAN200 and URBAN400 for area *i*;

P_i = a vector that contains the dummy variables for the ten provinces;

R_i = a vector that contains other explanatory variables (ADJUS, PAGE65 and PNOL); and

ε_i = a random error term.

The Greek letters α, β, λ, and γ represent vectors of coefficients. A complete list of the variables of the model and their definitions can be found in Table 2. ε_i is assumed to be normally distributed, with a mean of zero and a constant variance for all *i*.

My major interest is to examine how the variables in the vector *X* influence average employment income. In addition, I would like to see whether the other variables have positive or negative effects on the dependent variable. The model will be estimated using OLS, but some diagnostic tests will be carried out. In the next section the results of these tests and the parameter estimates will be discussed.

IV. Empirical Results

1. Diagnostic Tests

Some diagnostic tests are applied to my regression equation to check whether the OLS assumptions are valid.[9] The results of these tests are listed in Table 3.

First, I check the normality of the random error term. The derivation of the distributions of the t and F statistics depends on the assumption that the random term is normally distributed. If the distribution of the error term is unknown, then the distributions of t-statistics and F-statistics are unknown, and t and F tests of linear hypotheses about the coefficients will no longer be valid. If I don't verify this at the beginning, later hypothesis tests might not be valid.

Here I use two methods to test for normality. The first is the Jarque-Bera (JB) test. The observed value of the JB statistic is 113.095 with a p-value equal to zero. Thus at the five percent level of significance, I must reject the null hypothesis that the error terms are normally distributed. The second one is the Chi-squared Goodness of Fit Test. In this

[9] The econometric software package SHAZAM 10 is used for my analysis.

case I obtain a chi-square value of 58.840 with a p-value equal to zero. Both tests thus indicate that the errors are not normally distributed in my sample. Theoretically, the action in the face of nonnormal errors is to use an asymptotic hypothesis test instead of the t or F test. But in practice, in most cases using critical values from the t and F distributions in small samples will result a more conservative test (Green 2008, section 4.9 and 5.4). Thus I will still use t and F statistics to test hypotheses about the coefficients.

Second, simple correlation coefficients are used to test for multicollinearity. All of the correlation coefficients between the independent variables are much less than 0.8. This result implies that the potential problem of multicollinearity is low. The auxiliary R^2s are also checked, since multicollinearity can involve more than two variables. Some of these auxiliary R^2s in Table 4 are larger than 0.8, which is an indication of a linear relationship between the explanatory variables. Since these two tests generate different results and multicollinearity can be viewed as a data problem, it does not guarantee that there will exist a problem with statistical inference. Therefore, I do nothing about multicollinearity.

Third, I test for heteroskedasticity. Heteroskedasticity is very common in cross-section data like that used here. I use Koenker's (1981) modified version of the Breusch-Pagan-Godfrey (BPG) test to detect heteroskedasticity, because it is more powerful in samples with nonnormal errors. The observed value of the statistic is 48.964 with a p-value equal to 0.001, which implies that heteroskedasticity does exist in my model at the five percent significance level.

Finally, I check for specification errors in my model. I use Ramsey's RESET test and the modified FRESETS and FRESETL tests for this purpose.[10] Each of these three tests provides three variants; nine test statistics are shown in Table 3. Only the first variant of the FRESETS test shows there is no specification error in the model at the five percent significance level. The other eight variants imply that specification errors exist in the model. These errors may be due to omitted variables. However, I have already included all the most relevant explanatory variables in my model, so it is not clear what other variables should be included.

2. OLS Results

As mentioned in the previous subsection, the Ordinary Least Squares (OLS) estimation technique is used to estimate the model. The estimated standard errors have been corrected for heteroskedasticity using White's heteroskedasticity consistent variance-covariance matrix. Table 4 provides the estimated coefficients and other regression results. As shown in the table, the equation has reasonable explanatory power since both R^2 and the adjusted R^2 are greater than 0.8, which are quite high values for cross-section data. Together the explanatory variables explain more than eighty percent of the variation in average employment income. The F-statistic for the test of overall

[10] See DeBenedictis and Giles (1998) for an explanation of the FRESETS and FRESETL tests.

significance is 71.920, which also indicates the regression has explanatory power at the five percent significance level.

My major concern is to test the hypothesis that average employment income at the regional level is affected by the educational attainment of the region's population. Therefore I begin by carrying out a joint test of the significance of the coefficients of the seven education variables. The results are shown in Table 5. The F-statistic for this joint test is 27.899 with a p-value equal to 0.000, which suggests that not all these seven coefficients are zero. In other words, educational attainment does have an effect on average earnings.

Next individual tests of the coefficients of each education variable are carried out. At the five percent significance level, the coefficients of the variables PNOHIGH, PHIGH, PTRAD, PCOLL and PDEG are statistically significant. This implies that the percentage of the population that did not complete high school, the percentage of the population which completed at most high school, has at most a trades certificate, at most a college diploma, or completed a university degree all have an important effect on regional average income. But the coefficients of the variables PNOCOLL and PNODEG are not statistically significant. That is, there is no difference in the effect on average income of the percentage of adults who did not complete grade nine (the omitted category), the percentage of adults who attended college but did not obtain a college certificate and the percentage of adults who attended university but did not receive a degree. Domazlicky et al.'s (1996) study shows a similar result: the percent who acquired some college, but did not complete their degree does not significantly affect per capita personal income at the county level in southeast Missouri.

In the studies examining the effect of education on individuals' earnings, Card (1999), Lemieux (2001) and Sanmartin (2001) suggest that each additional year of education implies an increase in earnings. Since the percentage of the population 20 years or older whose highest level of schooling is less than grade nine is the variable excluded from the model, the estimated coefficients for the other educational attainment explanatory variables are expected to have positive signs. Consistent with this expectation, the coefficients are all greater than zero. The coefficient of the variable PNOHIGH is 213.91, which implies that if the percentage of adults who did not complete high school increases by one percentage point, keeping the other variables unchanged, the average employment income in that area will increase by $213.91.

The previous empirical results for other countries also suggest that a higher level of educational attainment will generate a greater return to education; thus the coefficient of a higher level of schooling is expected to be larger than its predecessor. That means gradually higher increases in average employment income. The results in Table 4 confirm this expectation (the coefficients of PNOCOLL and PNODEG are not considered here because of the lack of statistical significance). It is obvious that the return to a one percentage point increase in the percentage of the population with a university degree is largest, at $577.21. The increments in regional average employment income caused by each additional percentage point increase in the percentage that did not complete high school, the percentage that completed high school and the percentage that have a trade certificate are $213.91, $335.44 and $432.69 respectively. Similar results are found in Fullerton (2001) and Arellano and Fullerton (2005). The surprising find here is that the coefficient of the percentage of adults who have a college certificate is only 170.1, which is even lower than that for not completing high school.[11] This may suggest that the importance of high school and a trade certificate is greater than that of a college diploma. Another surprising result is that the coefficients of PNOCOLL and PNODEG are not significant, since these people did at least complete high school. However, overall the regression results support the idea that educational attainment affects regional average income.

The regression results also show that living in an urban area is associated with higher average employment income. In the joint test of the coefficients of URBAN400, URBAN200 and URBAN100, the F-statistic is 5.341 with a p-value equal to 0.001. This result indicates that the population density variables have an effect on average employment income at the five percent significance level. Large urban census divisions whose population density is 400 people per square kilometer or greater exhibit an average employment income that is more than $2000 above that of areas with a population density less than 100. Mid-sized urban areas whose population density is less than 400 but greater than or equal to 200 have an average employment income that is

[11] I double checked the data on education in E-STAT and made sure there was no error in selecting the data. The process used to construct the explanatory variables was also checked.

$2400 higher than that of areas with a population density less than 100. The average employment income difference between large urban areas and mid-sized urban areas is only $400, which is not too much. At the same time, there is no distinct difference between areas with a population density less than 100 and areas with a population density less than 200 but greater than or equal to 100, at the five percent level of significance. Therefore, the more urbanized is the region, the higher is average employment income, and if the population density is greater than or equal to 200 per square kilometer, the urban effect is significant. These results confirm the findings in Fullerton (2001) and Arellano (2005).

The location variables appear to have little effect on average employment income. The reference "province" of location of census divisions consists of the three territories. The coefficients of the other provinces represent the difference in average employment income between each province and the territories, holding all else constant. Note that in Table 4, most of these variables have negative coefficients. This means that if the census division is located outside the territories, it tends to have a lower average employment income. It is hard to explain this result. Maybe in the territories, the atrocious environment means many people do not like to live in that area. However, resource-based industries like coal mines need skilled labour, so high salaries may be required to attract people to work there. Since no detailed data or relevant studies can shed much light on this issue, no more discussion will focus on it.

Fortunately, the individual tests of the coefficients of the provincial variables show that none of the coefficients except the coefficient of PE are statistically significant at the five percent significance level. Only one more coefficient (the coefficient of NF) is significant at the ten percent significance level. Since almost no provincial differences are left, I do not need to worry too much about the sign and the magnitude of these coefficients. However, the F statistic for a joint test of significance of the coefficients of the provincial dummies is 9.53, which is consistent with the finding that at least two of the provincial dummies have significant coefficients.

On the other hand, the coefficient of the variable ADJUS (which identifies census divisions located in the border area adjacent to the U.S.) is -357.34, which suggests that border areas have a lower average employment income than non-border areas. But this coefficient is also not statistically significant at the five percent significance level.

The parameter estimate for the percentage of the population 65 years and over, PAGE65, is -398.58. It is significant at the five percent significance level. Each additional percentage point increase in PAGE65 will cause average employment income to decrease by $398.58. This result is consistent with the expectation that the variable PAGE65 will have a negative effect on average employment income. Sloboda (1999) finds the same result as well.

Throughout all labour markets in Canada, English and French are the most important communication skills which have a potential effect on personal income and thus regional average income levels. But unexpectedly, the estimated coefficient of the percentage of the population which does not have official language skills has a positive sign. That means if more people in an area do not have official language skills, that area will have a higher average income. This is very odd. However, the coefficient is not statistically significant and the language characteristic is not of major interest, so no more discussion will focus on it.

In summary, the regression results suggest that the education level variables have a significant effect on average employment income at the census division level. Higher levels of educational attainment will generate a greater return to education. But there is no difference in the effect on average income of the percentage of adults who did not complete grade nine, the percentage of adults who attended college but did not obtain a college certificate and the percentage of adults who attended university but did not receive a degree. The next section will examine the effect on average employment income at the provincial level of changes in the educational attainment of the population.

V. Simulation Results

In this section, the model and estimated coefficients are used to simulate the possible impact of an increase in an educational attainment level. All of the changes explored in the simulations constitute improvements in the region's stock of human capital. According to the results in the previous section, the coefficients of PNOCOLL and PNODEG are not statistically significant, and PNOHIGH represents the level of not having completed high school, so these three levels of education are not the subject of simulations. The other four variables represent the percentage of the population who successfully completed the high school (PHIGH), or have a trade certificate (PTRAD), or have a college certificate or diploma (PCOLL) or have a university degree. The potential gains caused by an increment in one of these four levels are computed in this section.

Two different types of simulation are carried out. In the first, one education level variable (PHIGH/PTRAD/PCOLL/PDEG) is increased in each census division by five percentage points while holding all the other education variables fixed. Since the estimated coefficients are positive for all these variables, all census divisions will experience an increase in their regional average employment income in this case. The second type of simulation consists of increasing the proportion at a given education level to the national average level. In census divisions in which the proportion of the population at that level of education is already above the national average, no change is made. Since some census divisions' education levels are lower than average, while others are higher, the outcome of this simulation is not certain. For both types of simulation, the results are presented at the provincial level, and each education level variable is discussed in turn.

Note that because the education level variables are defined as percentages of the population, it is impossible for one to increase without another one falling. For all the simulations, it is implicitly assumed that the increase in the proportion of the population at the specified level of educational attainment is accompanied by a decrease in the proportion of the population that did not complete grade nine. The latter variable is the educational category that was excluded from the estimating equation.

Since the dependent variable is the average employment income of the population 15 years and over, all provincial and national means of employment income are weighted by the population 15 years and over of each census division. In carrying out the simulations, the predicted average employment income of each census division is first calculated using the estimated coefficients and the actual educational level. This value is then multiplied by the population 15 years and over to obtain the predicted total employment income of each census division. Then the predicted total employment income of each province is obtained by summing total employment income across census divisions. The predicted average employment income of each province is computed as the predicted total employment income of the province divided by the population 15 years and over in the corresponding province. These calculations are then repeated with one level of educational attainment set at a new value. The changes in both the predicted total income and predicted average income of each province are measured as the difference between the predicted values before and after the change.

The percentage of the population that completed high school (PHIGH) is the first to be considered. Descriptive statistics for this variable are shown in Table 6-1. The provincial and national means in the table are weighted by the population 20 years and over of each census division (20% sample), since this is the population measure used to construct the education variables.

The table shows that in Canada, 13.91% of the population 20 years and over successfully completed high school, but did not attain a higher level of education. The census division with the highest percentage at this education level, 22.26%, is located in Quebec. The census division with the lowest percentage at this level of education, 3.75%, lies in one of the territories. Among the ten provinces, Quebec has the highest average percentage, 16.99%, for whom high school is the highest level of education, while Newfoundland has the lowest, only 9.08%. In total, there are 139 census divisions whose percentage at the high school level of education is lower than the national average level. All the census divisions located in Newfoundland, Prince Edward Island, Nova Scotia, Manitoba and Saskatchewan are below the national average, while most census divisions in New Brunswick, Quebec and Ontario are above the national average.

Next, how the provinces are affected by the two simulations is examined in Table 6-2 and Table 6-3. Table 6-2 uses the regression coefficients to calculate the impact on average employment income of raising each census division's high school educational achievement by five percentage points. In other words, the education level variable PHIGH is increased by five, while all the other explanatory variables remain fixed. Since the model used to simulate the change is linear, the increase in average employment income depends only on the coefficient of the education level variable PHIGH. The increase is thus the same for all census divisions and all provinces. This increase in average employment income is $1677.21. However, because the populations of the provinces are different, the increments in total employment income are different. Ontario benefits most from this change, experiencing an increase of $10,599 million. Prince Edward Island has the lowest population and thus benefits the least, with an increase of only $130 million.

Since the increase in employment income per person is the same for all regions, it might seem that there is no need to do simulations in this manner. However, because the initial average employment income is different for each province, the increase in average income may reduce provincial disparities in the level of average employment income. The effect on regional disparities at the provincial level will be discussed later in this section.

Table 6-3 instead raises below-average census divisions to the national average level. The census divisions whose high school education percentage (PHIGH) is lower than the national average level experience an increase in PHIGH to the national average level, while the other regions hold their levels unchanged. Since 94 out of 99 regions in Quebec have a higher high school education level than average, the increase in average employment income caused by this simulation is smallest in Quebec – only $14. Newfoundland has the lowest percentage of the population for whom high school graduation is the highest level of educational attainment, so benefits most in this simulation. Average employment income per person increases by $1615 in Newfoundland. The increases in Nova Scotia and Manitoba are also more than $1000. In terms of total employment income, New Brunswick gains $41 million, the smallest increase of all the provinces, while the largest gain is $1528 million for Ontario. Because the increment in total employment income is strongly affected by the size of the population 15 years and over in each province, Ontario, which has the largest population, is the most heavily influenced of all the provinces in both kinds of simulation. In contrast, average employment income is influenced primarily by the level of educational attainment and not by the population size.

The percentage of the population for whom a trade certificate is the highest level of education attained (PTRAD) is the second education level variable used in the simulations. Table 7-1 shows some descriptive statistics for this variable, while Tables 7-2 and 7-3 display the impact on average employment income and total employment income caused by the two simulations. The national average value of the population 20 years and over with a trade certificate is 11.78%. Newfoundland has the highest percentage of the population at the trade certificate level. Only three provinces (Quebec, Ontario and Manitoba) have a smaller percentage at this educational attainment level than the national average. In total, in 62 out of 288 census divisions the percentage of the population with a trade certificate is lower than the national average. This means most census divisions' values of PTRAD are higher than the average, while the other regions have extremely low percentages of the population with at most a trade certificate.

In the first simulation, the increase in average employment income is $2163.46 for all provinces. In the second case, the increments in average employment income differ across provinces. Because all census divisions located in Newfoundland, Prince Edward Island, Nova Scotia, Saskatchewan and Alberta are above the national average, PTRAD does not need to change in this case for these regions. Both the increase in average employment income and the increase in total employment income are zero for these five provinces. Ontario gains the most in terms of both average employment income and total employment income, with increases of $868 and $5488 million respectively. Quebec and British Columbia rank second and third in terms of increases in both average and total employment income.

Next, the percentage of the population which attended college and obtained a certificate or diploma (PCOLL), but did not obtain a university degree, is considered. Table 8-1, Table 8-2 and Table 8-3 in turn present the descriptive statistics for this variable and the two sets of simulation results. Looking first at Table 8-1, on average 16.18% of the population 20 years and over in Canada successfully completed college and received a certificate or diploma. In Ontario, the percentage of the population whose highest level of educational attainment is a college diploma is 17.09%, the highest in the country. Prince Edward Island, Alberta and British Columbia also have higher college attainment levels than the national average. Newfoundland has the lowest level, 11.90%. About two-thirds of the census divisions, or 191 out of 288, have college graduate percentages below the national average.

In the first set of simulation results, the increase in average employment income is $850 for all provinces. In the second set of simulation results, the largest increase in average employment income, $703, occurs in Newfoundland. The smallest increase is $38 in British Columbia. The smallest three increments in average employment income appear in British Columbia, Alberta and Ontario, the three provinces which have the highest percentages of college graduates.

With respect to total employment income, the results for the second simulation are quite different from those for average employment income. Quebec gains the most, more than $867 million. However, Prince Edward Island gains less than $12 million, which is the smallest increase. Since the population of Prince Edward Island is so small, even through the increase in average employment income in PE is larger than that in BC, Prince Edward Island still gains less than all the other provinces in terms of total employment income. The population difference can also explain why Quebec gains more than Newfoundland.

The last education variable that is examined is the percentage of the population whose highest level of attainment is a university degree (PDEG). Table 9-1, Table 9-2 and Table 9-3 provide some descriptive statistics for PDEG and the results of the simulations respectively. The national average across census division is 16.86% of the population with a university degree. Only Ontario and British Columbia lie above the national average. The number of universities located in these two provinces may provide an explanation for this fact. Only 10.52% of the population in Newfoundland has completed university, which is again the lowest percentage in the country. In total, 257 out of 288 census divisions in Canada lie below the national average for this education level. In most provinces, fewer than four census divisions lie above the national average for this education level.

As shown in Table 9-2, under the first simulation, the increase in average employment income is $2886 for all provinces. In the second simulation, where PDEG is raised to the national average level, Newfoundland benefits the most, since it has the lowest percentage of university graduates. Average employment income increases by more than $3550 in Newfoundland. Ontario gains the least ($1297) in this case. Manitoba, Alberta and British Columbia also experience increments of less than $2000. In terms of total employment income, the effect on Prince Edward Island is the largest, while the effect on Ontario is the smallest.

Taking all the results discussed above together, a five percentage point increase in PDEG (the percentage of the population 15 years and over which attended university and obtained a degree) has a greater effect on total employment income in Canada, $47223 million, than a five percentage point increase in any of the other education variables. This is because this variable has the largest coefficient of the four variables considered. When each educational attainment share is raised to the national average, the increase in total employment income in Canada is dependent not only on the estimated coefficients, but also on differences between the census divisions' actual education level and the national average level. The impact caused by a change in PEDG is again the largest ($27762 million). The smallest impact ($2453 million) is caused by a change in PCOLL (the percentage of the population that whose highest level of educational attainment is a college certificate or diploma).

Because provincial income disparities have long been of concern in Canada, it is also interesting to look at the effect of each simulation on regional disparities in average employment income. To this end, two measures of regional dispersion are computed, the standard deviation and the coefficient of variation (the standard deviation divided by the mean). Table 10 displays the results. The means and standard deviations shown in this table are weighted by the population 15 years and over with employment income. The standard deviation measures absolute dispersion, while the coefficient of variation measures dispersion relative to the mean. Although it is possible for one to go up while the other goes down, this didn't happen in my case.

It is obvious that the mean of average employment income increases as a result of each simulation, although when there is an across-the-board increase in the percentage of the population at one of the levels of educational attainment the standard deviations do not change. Nonetheless, the coefficients of variation become smaller in all four cases in the first simulation due to the increase in the mean of average employment income, which means that relative provincial disparities decrease.

In simulations where census divisions below the national average are raised to the national average percentage at a given level of educational attainment, both the weighted standard deviation and the coefficient variation across provinces decrease except when PTRAD, the percentage of the population whose highest level of education is a trade certificate, increases. As shown in Table 10, the standard deviation across provinces becomes bigger when PTRAD is raised to the national average in census divisions in which it was below the average level. I think the explanation is that in this case, in five of the ten provinces there are no census divisions with PTRAD below the national average, so incomes in these five provinces remained unchanged, while incomes in the other five increase. Furthermore these five provinces (except for Alberta) also have relatively low levels of average employment incomes before this change. Therefore, the simulation in this case increases disparities across provinces in Canada.

Thus in general, reducing disparities in educational attainment reduces both absolute and relative dispersion of average employment income across provinces. Increasing the percentage of each census division's population with a university degree to the national average level has the biggest impact on disparities. This result is consistent with those of Coulombe and Tremblay (2001, 2007). Thus overall, one can say that differences in educational attainment levels have an impact on provincial disparities in Canada.

Measures of regional disparities across census divisions within Canada are also presented, in Table 11. Again the mean and standard deviation are weighted by the population 15 years and over. Disparities across census divisions become smaller as well, except in the case where PTRAD is raised to the national average level. As in the case of provincial disparities, raising the percentage of the population with a university degree has the biggest impact on disparities across census divisions in Canada.

Overall, the results in this section show that raising the educational attainment level increases average employment income and total employment income for all provinces. Changes at the university level have the largest effect on employment income, regardless of the type of simulation. The resulting increases in average employment income have an impact on regional disparities in Canada. Both the disparities across provinces and the disparities across census divisions within Canada become smaller in seven of the eight simulations carried out.

VI. Conclusion

In this paper, the effect of education on regional economic performance is examined at the Canadian census division level. Information on education, population, land area and location for 288 census divisions in ten provinces and three territories are collected from the 2001 Canadian Census Profile in E-STAT. Then the relationship between the educational attainment of the census division population and average employment income is quantified in a linear equation. As in many previous studies, the highest level of schooling is used in this paper to represent the education level. To accurately represent the region's stock of human capital, the education level explanatory variables are defined as the percentage of the population at each level of schooling, such as did not complete high school, completed high school, has a trade certificate, started but did not complete college, completed college with a diploma, started but did not complete university and completed university with a degree. Other relevant variables are included in the model, such as population density, location, language, etc.

Based on my analysis, some findings are consistent with the previous research. The coefficient of each level of schooling is larger than that of its predecessor (the coefficient of PCOLL, the percentage of the population which has a college diploma, is the only exception); this means that a higher level of educational attainment will generate a greater return to education. The return to a one percentage point increase in the percentage of the population with a university degree is largest, at $577.21. The effect on average employment income is not significantly different for adults who did not complete grade nine, adults who attended college but did not obtain a college certificate or diploma, and adults who attended university but did not receive a degree.

The included population density variables were found to have an effect on average employment income as well: the more urbanized is the region, the higher is the average employment income. The results in this paper suggest that if the population density is greater than or equal to 200 persons per square kilometer, the urban effect is significant. The variable PAGE65 that represents the age structure of the population has a significant negative effect on average employment income at the census division level. However, the location variables included in the model, such as the provincial dummy variables and the border area dummy variable, and the official language variable appear to have little effect on average employment income.

To further explore the implications of the parameter estimates, changes in the percentage of the population at four education levels (high school, trade certificate, college diploma and university degree) were simulated. The increments in education increase the average employment income and total employment income for all provinces. In addition, the increases in average employment income affect regional disparities in Canada; disparities across provinces and disparities across census divisions both become smaller after the changes in the simulations, except for the case of changes in the trade certificate level in the second set of simulations.

Although the analysis of my paper was conducted as carefully as possible, the paper does have some limitations. First, the quality of schooling is assumed to be equal across all 288 census divisions. This is probably not the case in a country as large and diverse as Canada. This may influence my regression results to some extent. Second, due to data availability, the selection of explanatory variables may affect my results as well. In Domazlicky et al. (1996) and Sloboda (1999), the female labour participation rate is included in the model, while I do not use it in my model. The third potential problem with the model is endogeneity. As mentioned in some previous papers, not only could the education level influence the employment income, but also income could affect investment in education. Income level has the potential to affect many household choices related to the quality of life of individuals and their children, such as learning new skills and attaining a higher educational level. Therefore, more income could lead to more education. But this endogeneity problem can not be solved easily with a cross-section study such as this one. Fourth, there are some results shown in the paper for which I can not find a clear economic explanation in the Canadian situation, such as: the coefficient of PCOLL, the percentage of the population which has a college diploma, is smaller than the coefficients of lower levels of education; the average employment income controlling for the educational attainment of the population and other factors, is lower in most provinces than in the three territories; and official language skills do not have a significant impact on average employment income.

This paper's analysis of regional economic performance in Canada obtains some results similar to those of previous empirical studies, but also has some limitations. However, to my knowledge it is the only paper that examines the relationship between education and earnings at the census division level in Canada. As time goes by, I believe that newer and more efficient theories will be developed in this field. In future work, I will check out more data sources and more relevant papers and try to perfect my paper.

References

Alonso-Rodriguez, Agustin (2001) "Logistic Regression and World Income Distribution." *International Advances in Economic Research*, Vol. 7, No. 2, pp. 231-42.

Arellano, Armando; and Thomas M. Fullerton, Jr. (2005) "Educational Attainment and Regional Economic Performance in Mexico." *International Advances in Economic Research*, Vol. 11, No. 2, pp. 231-242.

Becker, Gary Stanley (1964) *Human Capital; a Theoretical and Empirical Analysis, with Special Reference to Education*. New York, National Bureau of Economic Research; distributed by Columbia University Press.

Card, David (1999) "The Causal Effect of Education on Earnings." *Handbook of Labor Economics*, Vol. 3A, pp. 1801-1863.

Canton, Erik (2007) "Social Returns to Education: Macro-evidence." *De Economist*, Vol. 155, No. 4, pp. 449-468.

Coulombe, Serge; and Jean-Francois Tremblay (2001) "Human Capital and Regional Convergence in Canada." *Journal of Economic Studies*, Vol. 28, No. 3, pp. 154-80.

Coulombe, Serge; and Jean-Francois Tremblay (2007) "Skills, Education, and Canadian Provincial Disparity." *Journal of Regional Science*, Vol. 47, No. 5, pp. 965-91.

DeBenedictis, L. F.; and D. E. A. Giles (1998) "Diagnostic Testing in Econometrics: Variable Addition, FRESET, and Fourier Approximations." In A. Ullah and D. E. A. Giles, *Handbook of Applied Economic Statistics*. New York: Marcel Dekker, 383-417.

Domazlicky, Bruce R.; Arron Benne, Molly McMahon, Clifford Myers, and Bruce Skinner (1996) "Measuring the Cost of High School Noncompletion in Southeast Missouri." *Journal of Economics*, Vol.22, No.1, pp. 81– 86.

Enchautegui, Maria E. (1993) "Education, Location, and Labor Market Outcomes of Puerto Rican Men during the 1980s." *Eastern Economic Journal*, Vol. 19, No. 3, pp. 295-308.

E-STAT wbesite, http://estat.statcan.ca.proxy.bib.uottawa.ca/cgi-win/cnsmcgi.pgm

Fullerton, Thomas M., Jr. (2001) "Educational Attainment and Border Income Performance." *Federal Reserve Bank of Dallas Economic and Financial Review*, 3rd Quarter 2001, pp. 2-10.

Galindo Martin, Miguel-Angel; and Agustin Alvarez Herran (2004) "Human Capital and Economic Growth in Spanish Regions." *International Advances in Economic Research*, Vol. 10, No. 4, pp. 257-64.

Greene, William H. (2008) *Econometric Analysis*, 6th edition. Upper Saddle River, NJ: Prentice Hall.

Heckman, J. and S. Polachek (1974) "Empirical evidence on the functional form of the earnings-schooling relationship." *Journal of the American Statistical Association*, Vol. 346, No. 69, pp. 350-354.

Iyigun, Murat (2005) "Geography, Demography, and Early Development" *Journal of Population Economics*, Vol. 18, No. 2, pp. 301-21.

Jorgenson, Dale W.; and Barbara M. Fraumeni (1993) "Education and Productivity Growth in a Market Economy." *Atlantic Economic Journal*, Vol. 21, No. 2, pp.1-25.

Koenker, R. (1981) "A Note on Studentizing a Test for Heteroscedasticity." *Journal of Econometrics,* Vol. 17, No. 1, pp. 107-112

Lemieux, Thomas (2001) "The Causal Effect of Education on Earnings in Canada." *Towards evidence-based policy for Canadian education /Vers des politiques canadiennes d'education fondees sur la recherché* . Kingston, Ont.: John Deutsch Institute for the Study of Economic Policy, Queen's University, pp. 105-116.

Mincer, J. (1974) *Schooling, Experience and Earnings.* Columbia University Press for the National Bureau of Economic Research, New York.

Sanmartin, Marta (2001) "Linearity of the Return to Education and Self Selection." *Applied Economics*, Vol. 33, No. 1, pp. 133-42.

Sloboda, Brian W. (1999) "Measuring the Costs of High School Dropouts on the Region of Southern Illinois." *Journal of Economics (MVEA)*, Vol. 25, No. 2, pp. 89-101.

Statistics Canada (2004) *2001 Census Dictionary*, Catalogue No. 92-378-XIE, Published by authority of the Minister responsible for Statistics Canada, Revised in December 2004.
Resource can be seen from:
http://www12.statcan.ca /english/census01/Products/Reference/dict/index.htm

Tilak, Jandhyala B. G. (1989) "Rates of Return to Education and Income Distribution." *De Economist*, Vol.137, No. 4, pp.454-465.

World Development Indicators, Washington, DC: World Bank, 2007.

2001 census reference map, resource can be seen from:
http://geodepot.statcan.ca /Diss/Maps/ReferenceMaps/n_cd_e.cfm

Appendix A. Variables drawn from the 2001 Census Profiles[1]

In the Census Profiles Database, the variables names are as follows:

- Census division name
- Province/Territory code
- Population, 2001 - 100% Data
- Land area in square kilometres, 2001
- Total number of persons 65 years and over - 20% Sample Data
- Total population by knowledge of official languages - 20% Sample Data
    - English only, population by knowledge of official languages
    - French only, population by knowledge of official languages
    - English and French, population by knowledge of official languages
    - Neither English nor French, population by knowledge of official languages
- Total population 20 years and over by highest level of schooling - 20% Sample Data
    - Less than grade 9, population 20 years and over by highest level of schooling
    - Grades 9 to 13, population 20 years and over by highest level of schooling
        - Without high school graduation certificate
        - With high school graduation certificate
    - Trades certificate or diploma, population 20 years and over by highest level of schooling
    - College, population 20 years and over by highest level of schooling
        - Without certificate or diploma
        - With certificate or diploma
    - University, population 20 years and over by highest level of schooling
        - Without degree, population 20 years and over by highest level of schooling
            - Without certificate or diploma
            - With certificate or diploma
        - With bachelor's degree or higher, population 20 years and over by highest level of schooling
- Total population 15 years and over with employment income, - 20% Sample Data
    - Average employment income $, population 15 year and over with employment income

1. In E-STAT, the database selected from the 2001 Census is "2001 Census of Population (Provinces, Census Divisions, Municipalities)". All the data are obtained from the profile named "2001 Census of Population: All Tables" in that database.

Table 1. **Descriptive statistics for the sample[1,2,3]**

Variable Name	Mean	Standard deviation	Minimum	Maximum
INCOME	31757	5106.7	15974	45835
PNOHIGH	17.42	4.61	7.70	33.82
PHIGH	13.91	2.87	3.75	22.26
PTRAD	11.78	2.93	7.07	25.35
PNOCOLL	6.36	1.07	2.91	13.83
PCOLL	16.18	2.37	6.75	21.85
PNODEG	7.04	2.20	1.67	12.51
PDEG	16.86	6.95	4.14	31.88
URBAN100	0.04	0.19	0.00	1.00
URBAN200	0.04	0.20	0.00	1.00
URBAN400	0.05	0.22	0.00	1.00
NF	0.03	0.18	0.00	1.00
PE	0.01	0.10	0.00	1.00
NS	0.06	0.24	0.00	1.00
NB	0.05	0.22	0.00	1.00
QC	0.34	0.48	0.00	1.00
ON	0.17	0.38	0.00	1.00
MB	0.08	0.27	0.00	1.00
SK	0.06	0.24	0.00	1.00
AB	0.07	0.25	0.00	1.00
BC	0.10	0.30	0.00	1.00
ADJUS	0.22	0.42	0.00	1.00
PAGE65	12.01	2.81	1.80	24.74
PNOL	1.49	1.75	0.00	16.66

1. Average employment income is weighted by the population 15 years and over with employment income.
2. Education variables are weighted by the census variable population 20 years and over by highest level of schooling-20% sample.
3. PAGE65 and PNOL are weighted by the population – 100% sample

Table 2. Definitions of variables used in the regression model

Variable name	Definition
INCOME	Average employment income ($) for the population 15 years and over with employment income
Education Attainment	
PNOHIGH	the percentage of the population that completed grade nine but did not complete high school
PHIGH	the percentage of the population whose highest educational qualification is a high school graduation diploma
PTRAD	the percentage of the population whose highest qualification is a trades certificate or diploma
PNOCOLL	the percentage of the population who attended college but did not obtain a certificate or diploma
PCOLL	the percentage of the population whose highest level of educational attainment is a college certificate or diploma
PNODEG	the percentage of the population that attended university but did not receive a degree
PDEG	the percentage of the population whose highest level of attainment is a university degree or a higher degree
Population Density	
URBAN400	=1, if population density greater than or equal to 400 people per square kilometer and population larger than 1000
URBAN200	=1, if population density greater than or equal to 200 but less than 400 and population larger than 1000
URBAN100	=1, if population density greater than or equal to 100 but less than200 and population larger than 1000
Province	
NF	=1, if census division is located in Newfoundland
PE	=1, if census division is located in Prince Edward Island
NS	=1, if census division is located in Nova Scotia
NB	=1, if census division is located in New Brunswick
QC	=1, if census division is located in Quebec
ON	=1, if census division is located in Ontario
MB	=1, if census division is located in Manitoba
SK	=1, if census division is located in Saskatchewan
AB	=1, if census division is located in Alberta
BC	=1, if census division is located in British Columbia
Other Variables	
ADJUS	=1, if census division is adjacent to U.S.
PAGE65	percentage of population age 65 or older
PNOL	percentage of individuals who can not speak official languages

Table 3. Diagnostic tests for equation estimated using OLS

Test Statistic	Test statistic	p-value
Jarque-Bear test	113.095	0.000
Goodness of Fit test	58.840	0.000
Auxiliary R^2s		
PNOHIGH	0.8983	
PHIGH	0.8282	
PTRAD	0.6365	
PNOCOLL	0.8007	
PCOLL	0.7235	
PNODEG	0.8601	
PDEG	0.8758	
URBAN100	0.1736	
URBAN200	0.2364	
URBAN400	0.3415	
NF	0.8414	
PE	0.6119	
NS	0.9060	
NB	0.8907	
QC	0.9735	
ON	0.9500	
MB	0.9190	
SK	0.9010	
AB	0.8748	
BC	0.9082	
ADJUS	0.1465	
PAGE65	0.4294	
PNOL	0.5024	
Breusch-Pagan-Godfrey (BPG) test		
KOENKER(based on R^2)	48.964	0.001
Ramsry's RESRT test		
RESET(2)	20.218	0.000
RESET(3)	10.511	0.000
RESET(4)	7.463	0.000
FRESETL test		
FRESET(1)	10.820	0.000
FRESET(2)	6.796	0.000
FRESET(3)	5.157	0.000
FRESETS test		
FRESET(1)	1.636	0.197
FRESET(2)	2.603	0.036
FRESET(3)	3.389	0.003

Table 4. OLS estimates

Dependent variable: INCOME (average employment income ($) for the population 15 years and over)

Variable Name	Estimated Coefficient	t-ratio	p-value
PNOHIGH	213.91	2.772	0.006
PHIGH	335.44	4.056	0.000
PTRAD	432.69	3.943	0.000
PNOCOLL	236.72	1.123	0.263
PCOLL	170.10	1.985	0.048
PNODEG	137.10	0.694	0.489
PDEG	577.21	6.528	0.000
URBAN100	662.73	1.026	0.306
URBAN200	2431.0	2.482	0.014
URBAN400	2026.9	3.474	0.001
PAGE65	-398.58	-7.348	0.000
PNOL	97.05	0.969	0.334
ADJUS	-357.34	-1.145	0.253
NF	-4285.4	-1.784	0.076
PE	-6412.8	-2.596	0.010
NS	-3268.1	-1.285	0.200
NB	-2859.0	-1.189	0.236
QC	-708.00	-0.326	0.745
ON	1627.2	0.741	0.459
MB	-1787.5	-0.737	0.462
SK	-1989.6	-0.812	0.417
AB	-742.91	-0.342	0.733
BC	-586.75	-0.274	0.785
CONSTANT	6647.0	1.872	0.062
R^2	0.821		
Adjusted R^2	0.805		
F statistic	71.920		0.000

Number of observations: 288 census divisions.

The F-statistic computed manually is for the test of overall significance.

Table 5. Joint Test of groups of coefficients

Test	F-statistic	p-value
Joint test for educational coefficients	27.899	0.000
Joint test for population density coefficients	5.341	0.001
Joint test for provincial coefficients	9.528	0.000

Table 6-1. Descriptive statistics for PHIGH, the percentage of the population whose highest level of educational attainment is a high school diploma

Province	No. of census divisions	Maximum	Minimum	Mean[1]	No.of areas with PHIGH below national average	No.of areas with PHIGH above national average
NF	10	10.12	7.32	9.08	10	0
PE	3	13.03	10.13	11.36	3	0
NS	18	12.53	7.88	9.75	18	0
NB	15	18.93	12.22	14.94	5	10
QC	99	22.26	11.94	16.99	5	94
ON	49	21.03	11.73	14.18	11	38
MB	23	13.12	6.41	11.36	23	0
SK	18	12.93	8.29	10.83	18	0
AB	19	14.33	10.69	11.41	16	3
BC	28	15.26	11.28	12.25	24	4
Canada	288	22.26	3.75	13.91	139	149

1. Means are weighted by the Census variable "population 20 years and over by highest level of schooling-20% sample".

Table 6-2. Effect of increasing PHIGH by five percentage points[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	421.89	23875.36	1677.21	25552.58
PE	1715.24	130.40	22060.94	1677.21	23738.15
NS	12526.99	786.32	26719.97	1677.21	28397.18
NB	9701.78	652.19	24949.61	1677.21	26626.83
QC	113533.80	6398.99	29757.90	1677.21	31435.11
ON	216928.30	10599.18	34326.71	1677.21	36003.92
MB	17073.83	1022.38	28009.63	1677.21	29686.84
SK	13820.85	896.23	25864.54	1677.21	27541.75
AB	56007.57	2966.04	31670.79	1677.21	33348.00
BC	67025.09	3570.04	31488.54	1677.21	33165.75
Canada	514339.178	27443.66	31332.07	1677.21	33009.28

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income".

Table 6-3. Effect of raising PHIGH to the national average level[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	406.26	23875.36	1615.05	25490.41
PE	1715.24	66.80	22060.94	859.21	22920.15
NS	12526.99	657.93	26719.97	1403.36	28123.33
NB	9701.78	40.72	24949.61	104.73	25054.34
QC	113533.80	54.17	29757.90	14.20	29772.10
ON	216928.30	1528.00	34326.71	241.79	34568.50
MB	17073.83	512.64	28009.63	840.99	28850.62
SK	13820.85	548.63	25864.54	1026.72	26891.26
AB	56007.57	1486.11	31670.79	840.36	32511.15
BC	67025.09	1194.97	31488.54	561.40	32049.94
Canada	514339.18	6496.25	31332.07	395.73	31727.81

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "population 15 years and over with employment income."

Table 7-1. Descriptive statistics for PTRAD, the percentage of the population whose highest level of educational attainment is a trade certificate

Province	No. of census divisions	Maximum	Minimum	Mean[1]	No.of areas with PTRAD below national average	No.of areas with PTRAD above national average
NF	10	25.35	14.53	18.46	0	10
PE	3	14.51	13.61	13.80	0	3
NS	18	20.31	13.02	15.29	0	18
NB	15	14.82	10.30	12.29	6	9
QC	99	17.90	8.11	11.50	27	72
ON	49	16.06	7.07	10.18	14	35
MB	23	16.63	8.97	11.72	13	10
SK	18	15.79	12.65	13.86	0	18
AB	19	23.58	12.24	14.04	0	19
BC	28	18.92	10.49	12.81	1	27
Canada	288	25.35	7.07	11.78	62	226

1. Means are weighted by the Census variable "population 20 years and over by highest level of schooling-20% sample".

Table 7-2 Effect of increasing PTRAD by five percentage points[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	544.21	23875.36	2163.46	26038.82
PE	1715.24	168.21	22060.94	2163.46	24224.39
NS	12526.99	1014.28	26719.97	2163.46	28883.42
NB	9701.78	841.27	24949.61	2163.46	27113.07
QC	113533.80	8254.12	29757.90	2163.46	31921.35
ON	216928.30	13671.99	34326.71	2163.46	36490.16
MB	17073.83	1318.77	28009.63	2163.46	30173.08
SK	13820.85	1156.05	25864.54	2163.46	28028.00
AB	56007.57	3825.92	31670.79	2163.46	33834.24
BC	67025.09	4605.03	31488.54	2163.46	33651.99
Canada	514339.18	35399.85	31332.07	2163.46	33495.53

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income."

Table 7-3. Effect of raising PTRAD to the national average level[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	0	23875.36	0	23875.36
PE	1715.24	0	22060.94	0	22060.94
NS	12526.99	0	26719.97	0	26719.97
NB	9701.78	51.48	24949.61	132.39	25082.00
QC	113533.80	1925.24	29757.90	504.62	30262.52
ON	216928.30	5488.03	34326.71	868.43	35195.13
MB	17073.83	150.29	28009.63	246.56	28256.19
SK	13820.85	0	25864.54	0	25864.54
AB	56007.57	0	31670.79	0	31670.79
BC	67025.09	617.58	31488.54	290.14	31778.68
Canada	514339.18	8232.63	31332.07	501.51	31833.58

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income."

Table 8-1. Descriptive statistics about PCOLL, the percentage of the population whose highest level of educational attainment is a college diploma

Province	No. of census divisions	Maximum	Minimum	Mean[1]	No.of areas with PCOLL below national average	No.of areas with PCOLL above national average
NF	10	14.35	6.75	11.90	10	0
PE	3	18.20	12.85	16.35	2	1
NS	18	17.49	10.66	15.91	13	5
NB	15	20.86	10.75	15.01	10	5
QC	99	21.85	8.43	15.28	79	20
ON	49	21.28	13.23	17.09	7	42
MB	23	16.70	6.77	14.30	21	2
SK	18	15.58	10.29	13.50	18	0
AB	19	19.15	12.12	16.95	14	5
BC	28	21.24	12.71	16.68	14	14
Canada	288	21.85	6.75	16.18	191	97

1. Means are weighted by the Census variable "population 20 years and over by highest level of schooling-20% sample".

Table 8-2. Effect of increasing PCOLL by five percentage points[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	213.93	23875.36	850.48	24725.85
PE	1715.24	66.12	22060.94	850.48	22911.42
NS	12526.99	398.73	26719.97	850.48	27570.45
NB	9701.78	330.71	24949.61	850.48	25800.10
QC	113533.80	3244.80	29757.90	850.48	30608.38
ON	216928.30	5374.64	34326.71	850.48	35177.19
MB	17073.83	518.43	28009.63	850.48	28860.11
SK	13820.85	454.46	25864.54	850.48	26715.02
AB	56007.57	1504.02	31670.79	850.48	32521.27
BC	67025.09	1810.30	31488.54	850.48	32339.02
Canada	514339.18	13916.14	31332.07	850.48	32182.55

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income."

Table 8-3. Effect of raising PCOLL to the national average level[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Preditced average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	176.78	23875.36	702.77	24578.13
PE	1715.24	11.87	22060.94	152.69	22213.63
NS	12526.99	59.83	26719.97	127.62	26847.59
NB	9701.78	101.62	24949.61	261.33	25210.94
QC	113533.80	867.48	29757.90	227.37	29985.27
ON	216928.30	610.78	34326.71	96.65	34423.36
MB	17073.83	193.98	28009.63	318.22	28327.85
SK	13820.85	238.90	25864.54	447.09	26311.63
AB	56007.57	110.66	31670.79	62.58	31733.36
BC	67025.09	81.32	31488.54	38.21	31526.74
Canada	514339.18	2453.23	31332.07	149.44	31481.52

1. All other variables are held constant.
2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income."

Table 9-1. Descriptive statistics about PDEG, the percentage of the population whose highest level of educational attainment is a university degree

Province	No. of census divisions	Maximum	Minimum	Mean[1]	No.of areas with PDEG below national average	No.of areas with PDEG above national average
NF	10	14.72	4.60	10.52	10	0
PE	3	17.36	6.44	12.62	2	1
NS	18	22.89	4.76	15.29	16	2
NB	15	23.57	6.83	12.36	14	1
QC	99	23.14	4.14	15.15	88	11
ON	49	31.88	6.06	19.15	41	8
MB	23	18.28	4.37	14.33	22	1
SK	18	18.81	5.98	12.32	17	1
AB	19	22.44	5.82	16.65	16	3
BC	28	22.51	6.74	17.61	26	2
Canada	288	31.88	4.14	16.86	257	31

1. Means are weighted by the Census variable "population 20 years and over by highest level of schooling-20% sample."

Table 9-2. Effect of increasing PDEG by five percentage points[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	725.97	23875.36	2886.06	26761.43
PE	1715.24	224.39	22060.94	2886.06	24947.00
NS	12526.99	1353.06	26719.97	2886.06	29606.03
NB	9701.78	1122.26	24949.61	2886.06	27835.68
QC	113533.80	11011.05	29757.90	2886.06	32643.96
ON	216928.30	18238.53	34326.71	2886.06	37212.77
MB	17073.83	1759.26	28009.63	2886.06	30895.69
SK	13820.85	1542.18	25864.54	2886.06	28750.60
AB	56007.57	5103.80	31670.79	2886.06	34556.85
BC	67025.09	6143.14	31488.54	2886.06	34374.60
Canada	514339.18	47223.64	31332.07	2886.06	34218.13

1. All other variables are held constant.

2. Total employment income and average employment income are weighted by the population 15 years and over with employment income.

Table 9-3. Effect of raising PDEG to the national average level[1,2]

Province	Predicted total employment income (millions of $)	Increase in total employment income (millions of $)	Predicted average employment income ($)	Increase in average employment income ($)	Average employment income after change ($)
NF	6005.73	893.20	23875.36	3550.86	27426.22
PE	1715.24	199.66	22060.94	2568.01	24628.95
NS	12526.99	1014.01	26719.97	2162.86	28882.83
NB	9701.78	1155.32	24949.61	2971.07	27920.68
QC	113533.80	8112.36	29757.90	2126.30	31884.20
ON	216928.30	8196.71	34326.71	1297.05	35623.75
MB	17073.83	1150.38	28009.63	1887.20	29896.83
SK	13820.85	1506.66	25864.54	2819.59	28684.13
AB	56007.57	2378.28	31670.79	1344.85	33015.64
BC	67025.09	3156.16	31488.54	1482.77	32971.31
Canada	514339.18	27762.73	31332.07	1691.23	33023.30

1. All other variables are held constant.

2. Average employment income for each provinces is weighted by "the population 15 years and over with employment income."

Table 10. Measures of regional dispersion across provinces within Canada[1,2]

			mean	Standard deviation	Coefficient of variation
Before change (10 provinces)		Predict average employment income	31542	4527.6	0.144
Simulations (10 provinces)	The first case	Changes in PHIGH	33219	4527.6	0.136
		Changes in PTRAD	33705	4527.6	0.134
		Changes in PCOLL	32392	4527.6	0.140
		Changes in PDEG	34428	4527.6	0.132
	The second case	Changes in PHIGH	32009	4186.2	0.131
		Changes in PTRAD	32145	4906.1	0.153
		Changes in PCOLL	31706	4420.9	0.139
		Changes in PDEG	33239	3885.4	0.117

1. Mean and standard deviation are weighted by population 15 year and over with employment income.
2. Average employment income used to compute dispersion in this table is weighted average employment income in each province.

Table 11. Measures of regional disparities across census divisions within Canada[1]

			mean	Standard deviation	Coefficient of variation
Before change (288 regions)		Predict average employment income	31437	4392.3	0.140
Simulations (288 regions)	The first method	Changes in PHIGH	33114	4392.3	0.133
		Changes in PTRAD	33600	4392.3	0.131
		Changes in PCOLL	32287	4392.3	0.136
		Changes in PDEG	34323	4392.3	0.128
	The second method	Changes in PHIGH	31839	4391.1	0.138
		Changes in PTRAD	31939	4854.2	0.152
		Changes in PCOLL	31587	4274.4	0.135
		Changes in PDEG	33133	3044.1	0.092

1. Mean and standard deviation are weighted by population 15 year and over with employment income.