Cole Sprague
Kai Addae
Economics 312
Canadian Census Project

Introduction

This project is based off of the 2001 Canadian Census data, and examines the relationship between wages and education, while controlling for gender, age, race, location, and immigrant status.

Data

The Canadian Census contains observations on 775,531 people living in Canada throughout all of the provinces. People under the age of 15 were excluded from the sample, as were any people that made $0 for the entire year as they were assumed to be not part of the work force. This left us with a sample size of 412,304 people. The variables are described as follows:

IMM	Is the person a permanent immigrant
NONP	A non permanent resident
Male	Does the person identify as male
CHIN	Person appearing Chinese
SASIAN	Person appearing to be from Southern Asia
BLACK	Person appearing Black
OTHER	Person appears as another visible minority
TOTSCHP	Total years of schooling achieved
YNN	Is the person from the Yukon Peninsula, Northwest Territories, or Nunavut
AGEP	Age of person
AGE2	Age of person squared
agemale	Interaction of AGE*Male
HS	Highest degree High School
TRADES	Highest degree Trade School
CCERT	Highest degree College Certificate below BA
UCERT	Highest degree University Certificate below BA
MED	Highest degree Medical certification
BACH	Highest degree Bachelors Degree

UBACH	Highest degree above University BA
MAST	Highest degree Masters Degree
DOC	Highest degree Doctorate

Exploring the Relationship Between Years of Schooling and Wage

Although the data is truncated so the highest observed wage is $200,000 , only 600 observations are at the limit so we decided that a tobit function was unnecessary for estimation. OLS was used with robust standard errors to account for heteroskedasticity.
Originally we controlled for region with a set of dummy variables for all of the provinces, but after regressing including all but one of the regions we were surprised to find that they were exceedingly collinear with other variables as the variance inflation factors for each was around 30 each regional dummy. Using the akaike information criterion, we settled on this model including controls for immigration status, sex, age, minority status and an interaction variable for age and sex, as well as one regional dummy (YNN) for persons living in the Yukon Peninsula, Northwest Territories, or Nunavut. These areas are well-known for their natural resource abundance (gold and oil) and so controlling for the higher wages there seems reasonable. With an R^2 of .278 we can account for approximately 28% of the wage earned in a year. Each additional year of schooling is predicted to increase your expected wages by $3352.70. Being an immigrant has little to no effect assuming they are a permanent resident. Non-permanent residents earn roughly $1652.22 less than their citizen counterparts. This is likely due to temporary work being paid less, such as seasonal farm work, or construction. If you are any visible race other that white then you are paid anywhere from $4465.60 to $6297.58 less than someone who passes as white. Being male has no economic or statistical effect on wages

```
Linear regression                               Number of obs =  412304
                                                F( 12,412291) =14242.22
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2781
                                                Root MSE      =   22861
```

WAGES	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
IMM	-23.75863	129.0823	-0.18	0.854	-276.756	229.2388
NONP	-1652.217	666.4074	-2.48	0.013	-2958.356	-346.0789
Male	-77.07214	174.4817	-0.44	0.659	-419.0511	264.9068
CHIN	-4464.594	231.156	-19.31	0.000	-4917.653	-4011.535
SASIAN	-4753.904	225.4075	-21.09	0.000	-5195.696	-4312.112
BLACK	-5744.388	226.415	-25.37	0.000	-6188.154	-5300.621
OTHER	-6297.576	187.7168	-33.55	0.000	-6665.495	-5929.656
TOTSCHP	3352.702	20.48224	163.69	0.000	3312.557	3392.846
YNN	4733.454	597.8681	7.92	0.000	3561.651	5905.257
AGEP	3035.217	14.98993	202.48	0.000	3005.837	3064.596
AGE2	-32.62677	.1957172	-166.70	0.000	-33.01037	-32.24317
agemale	349.8895	4.977213	70.30	0.000	340.1343	359.6447
_cons	-61065.74	277.316	-220.20	0.000	-61609.28	-60522.21

although the interaction of age and male is both economically and statistically significant. Men

earn more for experience than women do by about $350 per year. Age has a positive, but decreasing effect, most likely attributed to experience gains and then a decrease in productivity as you age and become less youthful.

The Relationship between Highest Earned Degree and Wage

We also estimated the relationship (using the same controls) between highest earned degree and wage. The theory is that earning the next highest degree is all that matters so someone who attempts some schooling and drops out has roughly the same qualifications as

Linear regression

Number of obs	= 412304
F(20,412283)	= 9289.49
Prob > F	= 0.0000
R-squared	= 0.2980
Root MSE	= 22544

WAGES	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
IMM	-627.2465	127.1552	-4.93	0.000	-876.4668	-378.0263
NONP	-4831.555	666.4825	-7.25	0.000	-6137.841	-3525.27
Male	748.8023	172.3825	4.34	0.000	410.9378	1086.667
CHIN	-5639.341	228.1788	-24.71	0.000	-6086.565	-5192.118
SASIAN	-5424.143	225.4497	-24.06	0.000	-5866.018	-4982.269
BLACK	-4732.16	223.9719	-21.13	0.000	-5171.138	-4293.182
OTHER	-6394.899	186.2646	-34.33	0.000	-6759.972	-6029.826
HS	4141.148	84.26966	49.14	0.000	3975.982	4306.314
TRADES	4726.269	115.6088	40.88	0.000	4499.679	4952.859
CCERT	9014.322	101.9761	88.40	0.000	8814.452	9214.192
UCERT	11181.56	243.2203	45.97	0.000	10704.86	11658.26
BACH	18589.96	146.5953	126.81	0.000	18302.64	18877.28
UBACH	20171.69	352.3057	57.26	0.000	19481.18	20862.2
MED	37002.12	1184.128	31.25	0.000	34681.26	39322.97
MAST	24978.63	311.9787	80.07	0.000	24367.16	25590.1
DOC	30138.33	706.5948	42.65	0.000	28753.42	31523.23
YNN	3170.801	603.5208	5.25	0.000	1987.919	4353.683
AGEP	2940.626	15.09341	194.83	0.000	2911.044	2970.209
AGE2	-32.32844	.196027	-164.92	0.000	-32.71265	-31.94423
agemale	327.2893	4.921515	66.50	0.000	317.6433	336.9353
_cons	-42679.42	256.9936	-166.07	0.000	-43183.12	-42175.72

someone that has never attended the further schooling. As done previously, we decided that 600 truncated observations were not enough to require use of the tobit distribution. Minimizing the akaike information criterion led us to choose the OLS specification, using robust standard errors to account for heteroskedasticity that can be found below.

This estimator has a slightly higher R^2 value than the earlier estimate, and can account for about 30% of the variation of yearly wages across individuals. All of the variables in this regression model are statistically significant, and except for possibly the AGE2 variable early in life, all are economically significant. Unlike the previous model, this regression estimates a statistically significant negative coefficient of on the immigrant variable (IMM) as well as on the

non-permanent resident variable (NONP). A person of immigrant status is estimated to make about $627.25 less than a natural-born Canadian in 2001, and a non-permanent resident is estimated to make $4831.56 less than a natural-born Canadian citizen. Men are estimated to make $748.80 more in yearly wages than women, and men are estimated to make $327.29 for each year older they are than a woman would. Age still has a positive but diminishing effect on wages, with an AGEP coefficient of $2940.63 per year, and a AGE2 coefficient of negative $32.33 dollars a year. This regression using dummies for highest degree attained is useful as it shows non-linear differences between education levels and wage, and relates specific certificate and degree levels to wage gains. All of the coefficient estimates are positive, but they increase as one reaches higher levels of degree attainment (relative to not having a high school degree).

Variable	Definition	Estimated yearly wage gains (rounded to nearest dollar)
HS	Highest degree High School	4141
TRADES	Highest degree Trade School	4726
CCERT	Highest degree College Certificate below BA	9014
UCERT	Highest degree University Certificate below BA	11,182
MED	Highest degree Medical certification	37,002
BACH	Highest degree Bachelors Degree	18,590
UBACH	Highest degree University above BA	20,172
MAST	Highest degree Masters Degree	24,979
DOC	Highest degree Doctorate	30,138

This shows that there is there are some large wage differences depending on degree/certification level attained. For instance, although a Doctorate may take more years of education than a medical certification, someone with a medical certification is predicted to make almost 7,000 dollars a year more than someone with a Doctorate. We also see that the difference in yearly wages between similar (timewise) education levels can be quite extreme-- a $2000+ difference in wages between those with a University and College certificate. This supports the conclusion that it doesn't only matter how much schooling one has, but what type of schooling one has, and the completion of that schooling.

Conclusions and Validity Assessment

Our sample of 412,304 Canadian Residents from 2001 shows a positive relationship between education level and yearly wages. Considering both total years of schooling and highest degree level attained, we find that there is a statistically and economically significant impact of education

level on wages, controlling for age, sex, minority status, immigration status, and to some extent region. Furthermore we can state that while more education in general is better, attaining certain degrees is best with a medical certification earning higher returns than a PhD. We used standard OLS with robust standard errors rather than a tobit regression because even though our sample was truncated, 600 observations out of over 400,000 seemed like too few to significantly affect our regression. Use of OLS is optimal under assumptions SR1 through SR5 on page 47 of HGL. Some issues of concern with the validity of these assumptions are:

Autocorrelation- Autocorrelation between individuals doesn't seem to be too much of a concern. Unfortunately our regional dummies were significantly collinear with other variables, and therefore were not useful enough to be included in the final regression. Locational controls may be useful, as we might expect education levels and wage levels to rise in urban areas, and this may complicate the relationship between education and wage.

Reverse Causality- Reverse causality doesn't seem to me too much of a concern here. It may be plausible in a minority of cases that someone is heavily encouraged or subsidized to receive more schooling because of their high wages, but for the most part causality seems to be one-way, especially since the educational attainment would have been in the past, and it seems unlikely that current high wages could affect past educational attainment.

Homoskedasticity and normality- We used robust standard errors in both OLS regression models to account for heteroskedasticity that we found using the standard STATA hettest. Even so, this heteroskedasticity does not bias our results, it just affects our standard error estimates.

Omitted Variable bias- Unfortunately, we found significant omitted variable bias in both our regression estimators. We found that age was significantly correlated with the error terms, yet even dropping the age variable from our sample didn't quite fix the problem. There seems to be some important variables missing from our regression models that are leading our estimates to be biased. Perhaps information on experience, locational, industry/sector of workplace would help to eliminate bias from our results. Our results may be significantly biased without the missing variables.

Functional Form- We found that there was a quadratic relationship between age and wage, and this suggests that there is the need to explore alternative functional forms when looking relationships with wage. This makes sense as you earn more money from experience over time, but after a while you get slower and worse at your job. Other time-based variables (workplace experience, time in current workplace) may also have a quadratic relationship with yearly wage.

As far as external validity, it seems that Census data would be highly applicable over time and to the whole of Canada, and perhaps to other similar countries. However, this sample is now over a decade old, and we may expect that as increasingly innovative technology advances remake the job market, education's effect on wages may have changed drastically. Furthermore, our estimates our known to be biased, and so we would not advise extrapolation of the estimated trends in this report without other information to found one's claims on. Using time series data to determine if there returns to education are changing would be interesting to examine in the future.