Multiple Linear Regression on the Volunteer Activity Survey Data

Professor, Chris Ferrall
Econometrics 452 Section B
Data Set #377, Screening Sample

Day, Kathleen & Rose Anne Devlin. "The payoff to work without pay: volunteer work as an investment in human capital." *Canadian Journal of Economics*, vol. 31(5) Nov. 1998 pp 1179-91.

Introduction

It is common practice to encourage any entrants to the work force with the old adage "If you can't find a job then volunteer." Besides the benefit of having concrete experience on your resume, authors Day and Devlin explore the financial benefit of volunteer work in, "The payoff to work without pay." The objective of their paper is to empirically verify that volunteer work increases one's earnings with a human capital earnings equation. One can contend that volunteer work leads to the acquisition of useful skills and experience thus, the "human capital model provides a natural framework for studying the returns to volunteering." (p. 1181)

The Model Paper: Data

By using the data from Statistics Canada, compiled in the 1987 Survey of Volunteer Activity (VAS) in conjunction with the November 1987 Labour Force Survey (LFS); Day and Devlin estimate a human capital earnings equation and measure the financial returns of volunteering. More specifically, the data set is the VAS screening file, and contains 37,426 observations. Though the Labour Force Survey is overwhelming to work with, the data set contains respondents' answers to both surveys. For a given individual, observations indicate volunteer and labour market activities. For the authors, the importance of this is paramount because they are estimating a human capital earnings equation and the VAS contains information on income.

Though the data set contains 37,426 observations, Day and Devlin work with a sub-sample of 5,147 observations as they are the only ones that prove useful for empirically verifying the hypothesis that volunteer work increases one's earnings. From the sub-sample, only 3003 individuals had volunteered in the period November 1986 to October 1987.

From the initial set of observations, the data set requires a bit of work before even a regression can be considered. To begin with, the income data is not in ideal form because the income data is available on a household basis and not on an individual basis and no distinction is made between employment income and other income sources. Understandably, the observations in the income variable are presented in ranges to protect confidentiality of the respondents.

As a result of the nature of the income variable, the data set had to be refined. Day and Devlin restricted their sample to employed individuals who were the sole wage earners in their household. The assumption is sole wage earners who volunteered were more likely to do so on a part time basis. Moreover, individuals over the age of sixty-five were eliminated from the sample because "employment income is less likely to be an important share of household income." (p.1181) Lastly, individuals who had not responded to the other questions of interest to their study were excluded.

The VAS screening file continues a multitude of information on the volunteer activities pursued. However, for the purposes of the Day and Devlin equation estimation, a volunteer was strictly defined as an individual "whose volunteer activities are carried out through a formal organization". Therfore, even though the VAS documents individuals who help neighbours or senior citizens informally, those individuals are ignored from the study.

The Model Paper: Method

Day and Devlin use the human capital model to estimate the returns to volunteering. The principal determinant of level of earnings is the individual's stock of human capital which is a function of education and labour market experience. The basic equation is as follows:

$$\ln W_i = b_0 + b_1 S_i + b_2 EXP_i + b_3 EXP^2_i + e_I$$

Where W_i is the earnings of individual *i*, S_i is years of schooling, EXP_1 denotes experience and e_I is a stochastic error term. However, the variable EXP, or experience is difficult to measure. Though Day and Devlin mention Mincer's definition of experience, age minus years of schooling minus six, they offer no other substitute. However, this measure makes it difficult to distinguish volunteer experience from other activities as it relates to the contribution of an individual's stock of human capital. Notably, Day and Devlin use Mincer's definition of experience when creating their variable, EXP.

Furthermore, Day and Devlin devote much of their discussion to measuring the impact of volunteer experience on an individual's human capital. The VAS is inadequate to provide the detailed information on the total lifetime hours volunteered by each individual, or an accurate measure of time spent earning income. Thus, after determing the specifications of a simultaneous model, a simple dummy variable is added to the above question. By adding a dummy variable indicating whether the individual is or has ever been a volunteer, we have the following equation:

$$\ln W_i = b_0 + b_1 S_i + b_2 EXP_i + b_3 EXP^2_i + b_4 VOL_i + e_I$$

Thus, if there is a positive return to volunteering, then the coefficient of the dummy variable, will be positive and statistically significant. Day and Devlin further explain that its magnitude will provide a sample estimate to the payoff to volunteering. Moreover, they expand their analysis by dividing the volunteer organizations into type and analyzing their results.

By estimating the human capital equation, they find their estimated coefficients for the earnings equations "are consistent with other studies." (1186) The findings of Day and Devlin can be summarized as follows:

- Earnings of males are higher than those of females
- Earnings of married individuals from sole-earner households have higher incomes than single individuals
- The higher the education level, the higher the income level of an individual

Day and Devlin conclude there is a positive payoff to volunteerism, of approximately 7 per cent higher incomes. Yet they also acknowledge that they are unable to explain why this is so, and put forward numerous unverified hypothesis; clearly highlighting the need for further research in the area.

Emulating the Data

Determining the payoff to work without pay in the same way as Day and Devlin do in their paper was not difficult because we were able to access the Volunteer Activity Survey Screening file. However, we did encounter difficulties attempting to emulate their sub sample and by extension, the number of individuals who volunteered during the period 1986 to October 1987.

We were able to narrow our 37426 observations to 5159, instead of 5147. Our dummy income variable allowed us to follow their directions perfectly and eliminate individuals whose answer was "not stated/don't know" as well as those in the highest income range, of greater than $60,000. Secondly, we kept all employed individuals, and then further narrowed the observations down by discerning which households had only one employed person. We chose this variable as a measure of the sole wage earners in the household. Lastly, we eliminated all observations for individuals over the age of 65. We encountered difficulty when we had to exclude individuals who had not responded to other questions of interest, as that was ambigious. Having browsed our observations and used the "codebook" command in Stata, we were still unable to reduce the observations in our subsample to less than 5159. In the table in Appendix A, we have listed the Variable names and their definitions.

The only variable we felt we had to construct and make our own assumptions on was EXP, or experience. The authors used its Mincer's definition as age minus years of schooling minus six. Following their method, we chose the mid point of the age range, as the variable was quoted in ranges. However, years of schooling was encoded to general assumptions associated with each level; thus ELEMENT, or no school, was coded as 4; HIGHSCH as 6; POSTSEC as 6.5; DIPLOMA as 7, and UNIVERS as 8. Again, these are mid points of these ranges.

In the article, much discussion is devoted to the type of the organization volunteered with and the occupation of individuals and the separate effects these two variables have on the payoffs associated with volunteerism. For the purposes of this assignment, we ignored the organizational and occupational analysis. Notably, in the paper, Day and Devlin do not indicate occupation or volunteer organization on their comprehensive set of variables, yet both variables are present in their regressions. The summary statistics on the variables we included are indicated in the table below.

The summary statistics are as follows:

Table 1. Summary Statistics		
Variable	**Mean**	**Standard Deviation**
EARNINGS (lnw)	1.51	.398
MALE	.614	.487
MARRIED	.476	.499
ELEMENT	.100	.300
HIGHSCH	.449	.497
POSTSEC	.100	.301
DIPLOMA	.180	.384
UNIVERS	.171	.376
FAMSIZE	2.30	1.22
OWNKIDS1	.153	.411
OWNKIDS2	.153	.408
OWNKIDS3	.458	.838
KIDSCH	.076	.303
FRENCH	.168	.374
ENGLISH	.168	.374
OTHLANG	.032	.176
EXP	25.1	11.9
EXP^2	772	698
RURAL	.247	.431
TOWN	.243	.429
CITY	.510	.500
HOURS (f05q13)	39.7	10.8
VOL	.580	.494

Based on our attempted emulation of Day and Devlin's method, we are confident we were able to work effectively with the Volunteer Activity Survey Screening sample and create a data set comparable to the one in the paper.

Emulating the Results

After working with the data set sub sample, we were able to run a regression on 10 variables. It was not surprising that we were able to achieve similar results to Day and Devlin. Thus, our human capital model was formulated as follows:

$$\ln W_i = b_0 + b_1 MALE_i + b_2 MARRIED_i + b_3 HIGHSCH_i + b_4 DIPLOMA_i + b_5 UNIVERS_i + b_6 EXP_i +$$

$$b_7 EXP^2_1 + b_8 RURAL_i + b_9 CITY_i + b_{10} VOL_i + e_I$$

The results of our regression are in the table below.

Table 2. Regression Table		
Variable	**Coefficient (Standard Error)**	**t-statistic**
MALE	.124 (.011)	11.0
MARRIED	.131 (.011)	11.4
HIGHSCH	.044 (.014)	3.20
DIPLOMA	.160 (.017)	9.45
UNIVERS	.229 (.017)	13.1
EXP	.027 (.017)	13.1
EXP^2	$-4.04e^{04}$ ($3.51e^{05}$)	-11.5
RURAL	-.020 (.015)	.-1.37
CITY	.080 (.013)	.6.34
VOL	..044 (.011)	4.16
_CONS	.863 (.029)	29.4

In specific comparison to the results obtained by Day and Devlin, we were in fact able to simulate their model and show two coefficients exactly the same to their second decimal point; b_2 is 0.136 for Day and Devlin and we estimate it at 0.131 and b_7 is –0.0004 for Day and Devlin which is what we estimate! In terms of general trends, our human capital model supports their conclusions; the earnings of males, married individuals, and higher education levels indicate higher levels of income.

To conclude our results we performed a hypothesis test on b_{10}, or the dummy variable on volunteering. Day and Devlin estimate volunteering to have a 7% increase on earnings, where as we estimate only 4%. Regardless, we formulated our test as follows:

$$H_o\text{: } b_{10} \text{ equal to } 0 \qquad H_a\text{: } b_{10} \text{ not equal to } 0$$

With a t-statistic of 4.157, and a p-value of 0.00 we know we can reject the null hypothesis, H_o, and retain the alternative. Thus, volunteering has a positive effect on earnings.

Summary

The process of choosing an article, and recreating the data set was a surprisingly intriguing experience. Even just initially, the plethora of research and work available was overwhelming yet exciting. We were able to find an article that we personally found interesting. In fact, we were hoping to reach the same conclusion because obviously, as students with relatively little work experience, it is encouraging to know that volunteering does increase one's earnings.

Taking the VAS Screening sample and ending up with a sub sample was a very enlightening experience as it took much of the theory from class and forced us to put it into practice - thereby allowing us to see the relevance and importance of understanding theory. More importantly, the theory was in fact applicable! From our experience regressions do not make sense, unless the data is in a form that you can make sense of it.

Though "The payoff to work without pay" was easy to understand, when it came to emulating the results we found that at times it was difficult to understand what Day and Devlin did with their data. The footnotes were extremely helpful in this endeavour, but as we are only introductory econometric students, it took time to comprehend their actions. As well, though our regression was restricted to 10 variables, we found the 10 we chose were relevant because we ignored any of the volunteer organization results, partly because the authors were not clear in the article how they collapsed some of the observations in to certain groups. Lastly, we were unclear how to approach variables that were indicated as reference groups, for instance, ENGLISH, so we chose to create it as dummy variable.

Overall, at the beginning we were overwhelmed with the task in front of us, but clearly it was not as daunting as it seemed once we understood the nature of our data and how we were going to discern if there is a payoff to work without pay.

APPENDIX A: Table of Variable Names

Variable Names & Definitions	
EARNINGS	Income range of individual
MALE	Dummy variable, 1 if male, 0 otherwise
MARRIED	Dummy variable, 1 if married, 0 otherwise
ELEMENT	Dummy variable, 1 if no school or elementary, 0 otherwise
HIGHSCH	Dummy variable, 1 if high school, 0 otherwise
POSTSEC	Dummy variable, 1 if some post-secondary education, 0 otherwise
DIPLOMA	Dummy variable, 1 if post-secondary diploma, 0 otherwise
UNIVERS	Dummy variable, 1 if university degree, 0 otherwise
FAMSIZE	Number of individuals residing in the household
OWNKIDS1	Number of own children ages 0-2 years old
OWNKIDS2	Number of own children ages 3-5 years old
OWNKIDS3	Number of own children ages 6-15 years old
KIDSCH	Number of children ages 16-24 attending school
FRENCH	Dummy variable, 1 if language at home is French., 0 otherwise
ENGLISH	Dummy variable, 1 if language at home is English, 0 otherwise
OTHLANG	Dummy variable, 1 if language at home is other, 0 otherwise
EXP	Experience
EXP^2	Experience squared
RURAL	Dummy variable, 1 if lives in small urban areas and rural areas, 0 otherwise
TOWN	Dummy variable, 1 if lives in other cities, 0 otherwise
CITY	Dummy variable, 1 if lives in metro areas, 0 otherwise
HOURS (f05q13)	Total Usual weekly hours worked
VOL	Dummy variable, 1 if volunteered in current year, 0 otherwise

APPENDIX B: Log File

(In the interests of saving space, only Stata Commands are here.)

```
-> . Qextract
getting information about file 377 ...
loading variables from 377 (vas87scr) only (no data yet)...  done
-> . do "a:project3.do"
/* Assignment 1: The payoff to work without pay
by Kathleen M. Day & Rose Anne Devlin*/

#delimit;

/* To begin emulating the data set, we need to drop individuals who don’t know
their income or make over $60,000.*/
tab f06_q35c, missing;
tab f06_q35c, gen(dinc);
drop if dinc1==.;
count if dinc8>0;
drop if dinc8>0;

/* We need to only have employed individuals.*/
tab lfstatus, missing;
tab lfstatus, gen(demp);
count if demp1>0;
keep if demp1>0;
```

```
/*We need to have individuals under the age of 65.*/
tab f03q33, missing;
tab f03q33, gen(dage);
count if dage8<1 & dage9<1;
keep if dage8<1 & dage9<1;

/* We need sole wage earners, thus households with only one employed person.*/
tab empfam, missing;
tab empfam, gen(dsole);
count if dsole1>0;
keep if dsole1>0;

/* We need to generate our dependent variable.*/
gen lnw=ln(f06_q35c);
tab lnw;
tab f06_q35c;

/* We now generate our dummy variables*/
tab f03q34;
count if f03q34>1;
count if f03q34<2;
gen male = f03q34<2;

tab f03q35;
count if f03q35<2;
gen married = f03q35<2;

tab f03q38;
count if f03q38<2;
gen element = f03q38<2;
count if f03q38>=2 & f03q38<3;
gen highsch = f03q38>=2 & f03q38<3;
count if f03q38>=3 & f03q38<4;
gen postsec = f03q38>=3 & f03q38<4;
count if f03q38>=4 & f03q38<5;
gen diploma = f03q38>=4 & f03q38<5;
count if f03q38>=5;
generate univers = f03q38>=5;

tab f06_q30b;
count if f06_q30b<2;
gen french = f06_q30b<2;
tab f06_q30a;
count if f06_q30a<2;
gen english = f06_q30b<2;
tab f06_q30c;
count if f06_q30c<2;
gen othlang = f06_q30c<2;
tab f06_q15, missing;
count if f06_q15<8;
gen vol = f06_q15<8;

tab province, gen(dprov);


tab f03q33;
gen mdage = f03q33;
recode mdage 1=15.5 2=18 3=22 4=30 5=40 6=50 7=60;
tab f03q38;
gen educ = f03q38;
recode educ 1=4 2=6 3=6.5 4=7 5=8;

gen exp = mdage - educ -6;
gen exp2 = exp*exp;

tab areaflg;
gen location = areaflg;
recode location 1=1 2=1 3=1 4=2 5=3;
```

```
tab location;
count if location<2;
gen city = location<2;
count if location>=2 & location<3 ;
gen town = location>=2 & location<3;
count if location>2;
gen rural = location>2;

/* Finally we can summarize our variables*/
.  summarize;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   QEDid |    5159    19335.16   10856.56         17      37426
province |    5159    35.02733   16.13405         10         59
  f03q34 |    5159    1.385928   .4868608          1          2
  f03q35 |    5159    1.709827   .7604273          1          3
  f03q33 |    5159    4.704012   1.239381          1          7
  f03q38 |    5159    2.872844   1.301905          1          5
  f05q13 |    5159    39.73638   10.78829          1         65
lfstatus |    5159           1          0          1          1
f05q7374 |    5159    7.723008   3.496222          1         13
  f05q75 |    5159    23.81644   14.35259          1         49
 famsize |    5159    2.300833   1.222535          1          4
ownkids1 |    5159    .1525489   .4109173          0          3
ownkids2 |    5159    .1533243   .4082611          0          3
ownkids3 |    5159    .4582283    .837835          0          5
kidsatsh |    5159    .0763714   .3031148          0          3
  empfam |    5159           1          0          1          1
 areaflg |    5159    3.135104   1.513852          1          5
 f06_q15 |    2991    1.817452    1.09763          1          7
f06_q30a |    5026    1.193593   .3951531          1          2
f06_q30b |    5026    1.827099   .3781992          1          2
f06_q30c |    5026    1.967171   .1782073          1          2
f06_q35c |    5159    4.856755   1.544678          1          7
   dinc1 |    5159    .0155069   .1235693          0          1
   dinc2 |    5159    .0752084   .2637528          0          1
   dinc3 |    5159     .122892   .3283451          0          1
   dinc4 |    5159    .1484784    .355608          0          1
   dinc5 |    5159    .2504361   .4333062          0          1
   dinc6 |    5159    .2362861   .4248411          0          1
   dinc7 |    5159    .1511921   .3582707          0          1
   dinc8 |    5159           0          0          0          0
   demp1 |    5159           1          0          1          1
   demp2 |    5159           0          0          0          0
   demp3 |    5159           0          0          0          0
   dage1 |    5159    .0038767   .0621485          0          1
   dage2 |    5159    .0178329   .1323567          0          1
   dage3 |    5159    .1108742   .3140067          0          1
   dage4 |    5159     .364024   .4812021          0          1
   dage5 |    5159    .2618725   .4396962          0          1
   dage6 |    5159    .1242489   .3298972          0          1
   dage7 |    5159    .1172708   .3217739          0          1
   dage8 |    5159           0          0          0          0
   dage9 |    5159           0          0          0          0
  dsole1 |    5159           1          0          1          1
  dsole2 |    5159           0          0          0          0
  dsole3 |    5159           0          0          0          0
  dsole4 |    5159           0          0          0          0
  dsole5 |    5159           0          0          0          0
  dsole6 |    5159           0          0          0          0
  dsole7 |    5159           0          0          0          0
     lnw |    5159    1.513611   .3982359          0    1.94591
    male |    5159    .6140725   .4868608          0          1
 married |    5159    .4762551   .4994843          0          1
 element |    5159    .0998255   .2997963          0          1
 highsch |    5159    .4489242   .4974326          0          1
 postsec |    5159    .1004071   .3005711          0          1
```

```
 diploma |     5159    .1802675   .3844474          0          1
 univers |     5159    .1705757   .3761742          0          1
  french |     5159    .1684435   .3742959          0          1
 english |     5159    .1684435   .3742959          0          1
 othlang |     5159    .0319829   .1759717          0          1
     vol |     5159    .5797635   .4936446          0          1
  dprov1 |     5159    .0595077   .2365953          0          1
  dprov2 |     5159    .0226788   .1488918          0          1
  dprov3 |     5159    .0674549   .2508325          0          1
  dprov4 |     5159    .0767591   .2662346          0          1
  dprov5 |     5159    .1482846   .3554163          0          1
  dprov6 |     5159    .1752278   .3801986          0          1
  dprov7 |     5159    .0783097   .2686844          0          1
  dprov8 |     5159    .0986625   .2982373          0          1
  dprov9 |     5159    .1517736   .3588361          0          1
 dprov10 |     5159    .1213413   .3265552          0          1
   mdage |     5159    37.46462   11.67113       15.5         60
    educ |     5159    6.371971   1.071837          4          8
     exp |     5159    25.09265   11.93134        3.5         50
    exp2 |     5159    771.9706    697.889      12.25       2500
location |     5159    1.736965   .8293825          1          3
    city |     5159    .5099826   .4999488          0          1
    town |     5159    .2430704   .4289788          0          1
   rural |     5159    .2469471   .4312775          0          1


/* Lastly, we can run a regression.*/
. regress lnw male married highsch diploma univers exp exp2 rural city vol;

  Source |       SS       df       MS                  Number of obs =    5159
---------+------------------------------               F( 10,  5148) =  104.75
   Model |  138.300197    10  13.8300197               Prob > F      =  0.0000
Residual |  679.716544  5148  .132035071               R-squared     =  0.1691
---------+------------------------------               Adj R-squared =  0.1675
   Total |  818.016741  5158  .158591846               Root MSE      =  .36337

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |   .1242691   .0112996     10.998   0.000       .1021172    .1464211
 married |   .1309856   .0114507     11.439   0.000       .1085373    .1534339
 highsch |    .044223   .0138084      3.203   0.001       .0171527    .0712933
 diploma |   .1599195    .016916      9.454   0.000       .1267568    .1930821
 univers |     .22895   .0174548     13.117   0.000       .1947311    .2631689
     exp |   .0268639   .0020546     13.075   0.000        .022836    .0308918
    exp2 |  -.0004037   .0000351    -11.504   0.000      -.0004725   -.0003349
   rural |  -.0199525   .0146007     -1.367   0.172       -.048576     .008671
    city |   .0797786   .0125896      6.337   0.000       .0550976    .1044595
     vol |   .0439993   .0105852      4.157   0.000       .0232478    .0647507
   _cons |   .8634901   .0293548     29.416   0.000       .8059423    .9210379
------------------------------------------------------------------------------

.
end of do-file

-> . BREAK
sending Break to calling program...
Session ended at 22 Mar 2001; 10:35:14
******************************************************************************
```

The Influence of Male and Female Incomes on Patterns of Household Expenditure:

A Study of the *Family Expenditures Survey* with References to Phipps and Burton

The data set for this paper was obtained from the Queen’s Economics Data Archive (file number 21). The reference paper is:

Phipps, Shelley and Peter Burton (1998) ‘What’s Mine is Yours? The Influence of Male and Female Incomes of Patterns of Household Expenditure,’ *Economica* 65, 599-613

1. Introduction

Differences in gender have assumed important roles in determining household spending and consumption patterns. As more women enter the labour force, and assume higher income occupations, the difference between male and female consumption patterns become increasingly relevant to economic analysis. Early research in household spending patterns focused on models that assumed that husband and wife have the same, or household, utility functions. Attempting to correct for this unrealistic assumption, recent theory has focused on relationship models where spouses may disagree over certain areas of spending, yet assume a gendered assignment of responsibility to reach a compromise. To test whether the household utility function theory bears relevance to empirical results, Phipps and Burton study the Family Expenditures Survey conducted by Statistics Canada in 1992 (FAMEX). The FAMEX offers a detailed respondent profile, providing comprehensive spending and income figures, thus permitting more thorough analysis regarding their interrelation. The primary advantage of the FAMEX is that both respondent and spousal income is reported, as well as number of weeks worked, allowing for a full-time/part-time labour status distinction of both income earners. As well, the survey also reports the number of other full-time and part-time earners in the household. This inclusion allows for a more accurate depiction of the typical two-income household. Finally, the fourteen spending categories reported by the survey cover almost all aspects of typical household expenditures. Phipps and Burton use these aspects of the FAMEX to test the relevance and magnitude of husband and wife spending differentials when applied to specific spending categories. Nominal spending and income statistics, descriptive statistics, multiple regressions and probit estimations of observed household purchases of a particular good are the tools used by Phipps and Burton to make inferences regarding gender wage levels and spending differences. The authors restrict their sample, and our attempts to replicate the same restricted sample did not strictly match Phipps and Burton. The method used, as well as the reasons for our deviation will be discussed in Section 2. After fourteen regressions using each of the spending categories as the dependent variable, the author's results show that income-pooling behaviour of husband and wife is statistically valid in only six categories. This conclusion refutes the hypothesis that there is a household utility function over all goods. For the purposes of this paper we will restrict our testing to the donations category[1] of household spending, one of the eight categories where income pooling was rejected. While the regression results are closely related to those derived by the authors, irregularities with the inclusion of component figures in the donations figure result in rejection of the income-pooling hypothesis. The reasons for these deviations will be discussed in Section 3. This paper is organized in four sections. Section 2 provides a description of the FAMEX survey, restricted sample derivation and summary of key variables. Section 3 presents regression results that illustrate how the influence of male and female income, as well as other factors discussed. Section 4 provides a summary and conclusions.

2. Data

The FAMEX was conducted nationwide in 1992 by Statistics Canada and represents data collected during at least one, yet usually more, interviews with a

---

[1] Donations was chosen due to it's ease of aggregation, which reduces the probability of deviation from Phipps and Burton's results.

respondent from a private household. During the interview, respondents were asked to recall expenditures made during 1991, thus repeat visits were needed so that respondents could consult financial records. The final public-use observation count includes 9492 observations. Total spending amounts were reported by the respondent in numerous areas of consumption, falling under the following categories: restaurant food, household food, housing, wife's clothing, husband's clothing, child care, household operations, recreation flows, recreation stocks, donations, transportation flows, transportation stocks, and tobacco and alcohol. In addition, household economic statistics beyond income were reported, such as social assistance received, sources of income or size of residence. In total, 251 figures were reported by each respondent.

We had difficulty restricting our sample exactly as Phipps and Burton restricted their sample. Our sample consisted only of married couples[2], with or without children. Households where there lived any other persons earning income, such as employed teenagers, were excluded. Inclusion of these households would complicate the results, since income earning teenagers will have input on how those earning are spent. To keep labour supply constant across households, any households where either spouse or respondent did not work at least 40 weeks were dropped, thus including only full-time workers. This reduces the possibility that differences in patterns of expenditure are explainable as a result of labour related cost. Based on this criteria, our restricted sample consisted of 950[3] observations, which is greater Phipps and Burton sample of 921. The difference is explained as the authors control for hours of full-time paid employment, since there is less variation compared to simple full-time workers. How this is

[2] The definition of married couple includes both married and common-law couples.
[3] Regression observations number 893, due to lack of sufficient donation of some households in the restricted set.

implemented, given the absence of explicit or implicit indication of hours of full-time paid employment, the authors do not make clear. However, the sample means of both male and female income is very close to Phipps and Burton's figures despite this discrepancy. For these figures, as well as others used in our regression, refer to Table 1.

The authors do not present their own calculations for all the variables found in Table 1, so comparison is not possible[4]. However, the authors do present mean male and female income, and their figures do differ from our own mean calculations, thus indicating that our other calculations in Table 1 differ from the authors. This is due to our inability to narrow the sample size to Phipps and Burtons 921 observations, however the difference between our calculations of mean income and the authors differ by less than 2%. The small difference may be indicative of the similarity, and our relative success, in paralleling Phipps and Burton's observations.

Looking at the data, we can note some interesting trends. First, the mean income levels are 44,441.30 for males and 31,592.80 for females[5]. This is consistent with empirical evidence that finds that male income is higher than female income. Secondly, Phipps and Burton neglect to mention criteria for estimating total donations, such as whether donations are net of receipts. Given that donations comprised of 8 possible sub-categories[6], our estimation of 3,759 total donations was not the same as Phipps and Burton. Continuing, it is interesting to note that 59% of respondents lived in urban areas[7], yet respondents living in a suburban environment would have more in common with city dwellers than their rural counterparts they are grouped with. Both males and

[4] The authors do state that figures are available upon request. At the current time the authors have not responded.
[5] Income is calculated on a pre-tax basis.
[6] 8 sub-categories do not include possible receipts categories.
[7] The survey defines living in an urban area as residing in one of the 15 major Canadian cities.

females report roughly the same level of education, where 3.3 represents a point between some post-secondary education (3) and a college certificate (4). The males average occupation score is 4.9, while females average is 3.4. Since traditionally male occupations were assigned higher numbers, this figure is also sensical.

The remainder of the figures presented are calculated and used in the regression, as presented in Section 3. Age-squared and wage-squared are tabulated to test the marginal effect of age and weight on donations. Multiplying husband income by wife income to create HWINCOME independent variable is used to test the effect of interaction of husband and wife income levels on donations. It is interesting to note that Phipps and Burton state only male-age squared is included, but male and female ages are included. This problem in interpretation is representative of the author's brief explanation of demographic control variables. This, combined with an absence of their complete calculated coefficient estimates makes our own accurate replication more frustrating.

Empirical Results

Table 2 contains the estimation results for the expression relating total donations given by the household to the explanatory variables found in Table 1. In addition, dummy variables were calculating for provincial estimates and included to test regional differences. Since the focus of their paper are income effects on expenditures, Phipps and Burton report only coefficient results for the income based variables. Although our paper deals with only donation expenditures, Phipps and Burton include figures for all 14 expenditures[8]. Further, they conduct likelihood ratio tests for the pooling restriction that $\beta_{HINC} = \beta_{WINC}$ and $\beta_{HINCSQ} = \beta_{WINCSQ} = \beta_{HWINCOME}*0.5$[9]. Continuing this discussion, the authors derive iso-expenditure curves to illustrate the income-expenditure relation in cases where the pooling hypothesis was rejected. Since our analysis focuses exclusively on donation expenditure, we will discuss all coefficient estimates.

HINC has a positive effect positive effect on donations, yet WINC has a negative effect. This result contradicts Phipps and Burton, who find both factors have a negative effect on donations. Further, HINCSQ has a negative effect and WINCSQ a positive effect on donations, where Phipps and Burton find both have positive effects. Given the deviation that exists in the sample set, and in the lackluster definition of donation calculations, a discrepancy is not surprising. Also, HINCSQ and WINCSQ have such a minimal effect, inconsistencies would not have a great impact on these variables.

Education levels for both men and women have significant positive effects on donation expenditure, although the effect is greater for men than for women. Looking at HOCC and WOCC, we see a significant negative effect on donations. This can be interpreted as a move towards from 'white-collar' towards 'blue-collar' type employment negatively affects donation expenditure. It is interesting to note that blue-collar positions tend to be lower paying. As well, education levels have also been linked to higher income levels. These two results contradict the previous observation that income is negatively related to donations.

---

[8] For complete description of this analysis, see Phipps and Burton p. 604.

[9] At the 10% significance level, Phipps and Larkin find pooling is only evident for 'big-ticket' expenditures, or 6 of the 14 categories.

Also interesting to note is the fact that living in a city is negatively related to donation expenditure. This is congruent with the idea that rural Canadians donate more than do their urban counterparts. When HAGE and WAGE is considered, we see that age is of both male and female age is negatively related to donation expenditure. Yet the relation is small for females, and significantly larger for males, indicating that donation expenditure is more sensitive to male aging than female.

The hypothesis put forth by Phipps and Burton found that the null-hypothesis $\beta_{HINC} = \beta_{WINC}$ and $\beta_{HINCSQ} = \beta_{WINCSQ} = \beta_{HWINCOME}*0.5$ could be rejected at the 10% significance level. Our own F-test of this hypothesis derives a p-value of .085, showing that the income pooling theory can also be rejected at the 10% significance level. It is interesting to note that Phipps and Burton p-value of .019 would have been rejected at the 5% significance level, while our income pooling hypothesis would have been retained.

The adjusted R^2 was lower than the amount Phipps and Burton reported. This is significant as the authors state that the adjusted R^2 was already lower than microdata figures usually indicate[10]. Therefore, decreased emphasis should be placed on our results due to low proportion of sample variation explained by the regression function.

4. Summary and Conclusions

The method and results presented above present some mixed conclusions. The socio-economic variables considered and tested are ones that would normally impact household expenditure. Education levels and occupation control variables have positive effect on donation expenditure, as would be expected. Also, our results show that age is negatively related to donation. Since Phipps and Burton do not provide comparable

[10] Phipps and Burton note this significance of a low adjusted R^2 on p. 603

figures in their discussion, we do not know how congruent our conclusions are with the authors.

The inconsistency with the coefficients that are presented by the authors, those associated with income, are indicative of the problems we experienced in replicating Phipps and Burton's data set. Although the positive relation between HINC and DONTOT contrast Phipps and Burton, the coefficient estimate for WINC was closer to the value determined by the authors. Our own conclusion that income pooling is not evident in donation expenditure at the 10% significance level is consistent with the author's own conclusions, despite our p-value discrepancy.

This problem is derived from the author's lack of clarity in outlining component sub-categories of the 14 main categories they analyze. Our attempts to minimize this error, by choosing donations which has the smallest number of potential sub-categories, still resulted in our creating a donations expenditure variable which was not the same as the authors. Added clarity regarding these figures would be beneficial, however we understand that available space constrained the author's inclusion of these figures.

One further problem encountered is the absence of some other defining socio-economic variables, such as if the household received social assistance beyond U.I. This would certainly impact expenditure patterns of the typical Canadian household. Also, some of the variables that are included, such as male age and it's square, would lead us to presume that female age squared would be included. We are unsure whether this exclusion was purposeful. If it was, then the reasons for this should be explained. If it was not, we wonder what other variables are we not including simply because they were not mentioned.

In closing, some questions raised by this analysis are worth commenting on. There is a distinct difference between men and women in their propensity to donate. This seems to due to a number of different gender factors, such a more distinct effect of male aging than female aging. Although the Phipps and Burton wished to test for the simple existence of a household utility through an income pooling hypothesis, it would be interesting to break down each expenditure grouping and discern the sources of different gender behaviour. For instance, would age be more significant in determining donation expenditure gender differences than in transportation expenditures. Analysis in these areas would certainly be beneficial in uncovering the sources of the gender-household utility gap.

REFERENCES

Phipps, Shelley and Peter Burton (1998) 'What's Mine is Yours? The Influence of Male and Female Incomes of Patterns of Household Expenditure,' *Economica* 65, 599-613

Stata Corporation Website: http://www.stata.com

Appendix A

TABLE 1: VARIABLE DESCRIPTIONS, SAMPLE MEANS AND STANDARD DEVIATIONS

Variable Name	Description	Mean
DONTOT	Total Donations	3759.4 (6321.6)
HAGE	Husband's age	39.9 (7.1)
WAGE	Wife's age	37.8 (6.9)
HINC	Husband's income	44441.3 (23144.6)
WINC	Wife's income	31592.8 (15977)
HEDUC*	Husband's education level	3.30 (1.3)
WEDUC*	Wife's education level	3.31 (1.2)
CITY	Probability of respondent living in a city	.59 (.5)
HAGESQ	Husband's age squared	950 (1640.4)
HINCSQ	Husband's income squared	$2.51*10^9$ ($3.60*10^9$)
WINCSQ	Wife's income squared	$1.25*10^9$ ($1.94*10^9$)
HWINCOME	Husband*wife income	$1.52*10^9$ ($1.41*10^9$)
HWINCOME	Husband*wife income	$1.52*10^9$ ($1.41*10^9$)
HOCC	Husband's occupation level	4.9 (3.6)
WOCC	Wife's occupation level	3.4 (2.1)

*education is based on an increasing 5 point scale, 1=elementary level : 5=university degree

Appendix B

TABLE 2: THE IMPACT OF INCOME POOLING ON HOUSEHOLD DONATIONS

Independent Variable	OLS Parameter Esitmates
HAGE	-549.51
	(-1.77)
WAGE	-2.96
	(-0.051)
HINC	0.014
	(0.286)
WINC	-0.027
	(-0.495)
HEDUC	227.97
	(1.05)
WEDUC	187.89
	(0.82)
CITY	-815.76
	(-1.78)
HAGESQ	7.596167
	(1.972)
HINCSQ	$-6.15*10^{-8}$
	(-0.434)
WINCSQ	$4.15*10^{-7}$
	(1.922)
HWINCOME	$1.02*10^{-6}$
	(2.37)
HOCC	-15.8312
	(-0.24)
WOCC	-51.8126
	(-0.44)
NUIREC	1387.813
	(1.339)
Constant	11889.25
	(1.93)
Adjusted R^2	0.0684

Note: *t*-ratios are presented in parentheses

Appendix C
LOG File

The following is the log file representing how the figures presented were calculated in Stata:

gen hsex = cond(sex==1,sex,ssex)
husbands sex
gen wsex=cond(sex==1,ssex,sex)
wifes sex
gen hage= cond(sex==1,age,sage)
husbands age
gen wage=cond(sex==1,sage,age)
wifes age
gen heduc=cond(sex==1,educ,seduc)
husbands education
gen weduc=cond(sex==1,seduc,educ)
wifes education
gen hinc=cond(sex==1,ibt,sibt)
husbands income before taxes
gen winc=cond(sex==1,sibt,ibt)
wifes income before taxes
gen hocc=cond(sex==1,occu,socc)
husbands occupation
(36 missing values generated)
gen wocc=cond(sex==1,socc,occu)
wifes occupation
(35 missing values generated)
Creating dummy variables
tabulate prov, gen(dvp)

geographic code (province)	Freq.	Percent	Cum.
masked records	10	1.01	1.01
newfoundland	59	5.97	6.98
prince edward island	31	3.13	10.11
nova cotia	52	5.26	15.37
new brunswick	69	6.98	22.35
quebec	181	18.30	40.65
ontario	270	27.30	67.95
manitoba	49	4.95	72.90
saskachewan	88	8.90	81.80
alberta	98	9.91	91.71
british columbia	82	8.29	100.00

Total | 989 100.00

Generating the aggregate donation total:

- ➔ . gen dontot = giftc+mgc+otgif+chaor+relor+oco
- ➔

GIFTC - GIFTS AND CONTRIBUTIONS
MGC - MONETARY GIFTS AND CONTRIBUTIONS
OTGIF - OTHER GIFTS
CHAOR - CHARITABLE ORGANIZATIONS
RELOR - RELIGIOUS ORGANIZATIONS
OCO - OTHER CHARITABLE ORGANIZATIONS

- ➔ . gen hincsq = hinc*hinc
- ➔ Husbands income before taxes squared
- ➔ . gen wincsq = winc*winc
- ➔ Wifes income before taxes squared

-> . gen hwincome = hinc*winc
Husbands multiplied by wifes income before taxes

Tabulations:
summ hage

Variable	Obs	Mean	Std. Dev.	Min	Max
hage	950	39.86947	7.138537	25	54

-> . summ wage

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	950	37.89895	6.997086	25	54

-> . summ hagesq

Variable	Obs	Mean	Std. Dev.	Min	Max
hagesq	950	1640.48	570.8558	625	2916

-> . summ hsex

Variable	Obs	Mean	Std. Dev.	Min	Max
hsex	950	1	0	1	1

-> . summ wsex

Variable	Obs	Mean	Std. Dev.	Min	Max
wsex	950	2	0	2	2

```
-> . summ heduc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   heduc |     949    3.302424   1.300034          1          5
-> . summ weduc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   weduc |     950    3.305263   1.226453          1          5
-> . summ hinc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    hinc |     950    44441.31    23144.6     -22688     234336
-> . summ winc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    winc |     950    31592.86   15977.02      -7700     220000
-> . summ hocc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    hocc |     916    4.973799   3.594071          1         11
-> . summ wocc

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    wocc |     915    3.419672   2.022533          1         11
-> . summ hincsq

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
  hincsq |     950    2.51e+09   3.60e+09          0   5.49e+10
-> . summ wincsq

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
  wincsq |     950    1.25e+09   1.94e+09          0   4.84e+10
-> . summ hwincome

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
hwincome |     950    1.52e+09   1.41e+09  -2.98e+08   1.65e+10
-> .
```

summm city

Variable	Obs	Mean	Std. Dev.	Min	Max
city	950	.5884211	.4923788	0	1

-> . summ dontot

Variable	Obs	Mean	Std. Dev.	Min	Max
dontot	950	3759.42	6321.634	30	122950

-> .

regress dontot city nuirec hage wage heduc weduc hinc winc hocc wocc dvp1 dvp2 dvp3 dvp4 dvp5 dvp6 dvp7 dvp8 dvp9 dvp10 hagesq hincsq wincsq hwincome

Source	SS	df	MS
Model	3.4668e+09	25	138673295
Residual	3.3215e+10	867	38309872.5
Total	3.6681e+10	892	41122748.7

Number of obs = 893
F(25, 867) = 3.62
Prob > F = 0.0000
R-squared = 0.0945
Adj R-squared = 0.0684
Root MSE = 6189.5

dontot	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
city	-815.7664	456.4707	-1.787	0.074	-1711.683	80.15046
nuirec	1387.813	1036.65	1.339	0.181	-646.8238	3422.449
hage	-549.5136	309.0352	-1.778	0.076	-1156.058	57.03099
wage	-2.95748	58.19512	-0.051	0.959	-117.1773	111.2623
heduc	227.9706	216.1409	1.055	0.292	-196.2499	652.1912
weduc	187.8909	226.5848	0.829	0.407	-256.8279	632.6097
hinc	.0138123	.0483688	0.286	0.775	-.0811213	.1087459
winc	-.0279641	.0564535	-0.495	0.620	-.1387655	.0828374
hocc	-15.83118	65.2511	-0.243	0.808	-143.8998	112.2374
wocc	-51.81263	115.8241	-0.447	0.655	-279.1411	175.5158
dvp1	2260.092	2116.013	1.068	0.286	-1893.015	6413.199
dvp2	347.1528	1177.402	0.295	0.768	-1963.739	2658.045
dvp3	-716.0433	1373.261	-0.521	0.602	-3411.348	1979.261
dvp4	252.0214	1159.193	0.217	0.828	-2023.13	2527.173
dvp5	330.5415	1085.067	0.305	0.761	-1799.124	2460.207
dvp6	-1523.192	897.8458	-1.696	0.090	-3285.397	239.0139
dvp7	-326.2074	838.0809	-0.389	0.697	-1971.112	1318.697
dvp8	-103.7304	1218.645	-0.085	0.932	-2495.569	2288.108
dvp9	1916.459	1011.838	1.894	0.059	-69.48015	3902.398
dvp10	1734.798	990.4305	1.752	0.080	-209.1242	3678.72

```
  hagesq |   7.596167   3.851476     1.972   0.049     .0368611    15.15547
  hincsq |  -6.15e-08   1.42e-07    -0.434   0.664    -3.40e-07    2.17e-07
  wincsq |   4.15e-07   2.16e-07     1.922   0.055    -8.78e-09    8.38e-07
hwincome |   1.02e-06   4.27e-07     2.376   0.018     1.77e-07    1.85e-06
   _cons |   11889.25   6162.088     1.929   0.054    -205.1008    23983.61
------------------------------------------------------------------------------


-> . summ hage

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    hage |     950    39.86947   7.138537         25         54
-> . summ wage

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    wage |     950    37.89895   6.997086         25         54
-> . summ hagesq

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  hagesq |     950     1640.48   570.8558        625       2916
-> . summ hsex

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    hsex |     950           1          0          1          1
-> . summ wsex

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    wsex |     950           2          0          2          2
-> . summ heduc

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   heduc |     949    3.302424   1.300034          1          5
-> . summ weduc

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   weduc |     950    3.305263   1.226453          1          5
-> . summ hinc

Variable |     Obs        Mean   Std. Dev.       Min        Max
```

```
---------+-----------------------------------------------------
    hinc |     950    44441.31    23144.6     -22688     234336
-> . summ winc

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    winc |     950    31592.86   15977.02      -7700     220000
-> . summ hocc

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    hocc |     916    4.973799   3.594071          1         11
-> . summ wocc

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    wocc |     915    3.419672   2.022533          1         11
-> . summ hincsq

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  hincsq |     950    2.51e+09   3.60e+09          0   5.49e+10
-> . summ wincsq

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  wincsq |     950    1.25e+09   1.94e+09          0   4.84e+10
-> . summ hwincome

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
hwincome |     950    1.52e+09   1.41e+09  -2.98e+08   1.65e+10
-> . summ city

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    city |     950    .5884211   .4923788          0          1
-> . summ dontot

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  dontot |     950     3759.42   6321.634         30     122950

-> . regress dontot  city nuirec hage wage heduc  weduc hinc winc hocc wocc dvp1 dvp2
dvp3 dvp4 dvp5 dvp6 dvp7 dvp8 dvp9 dvp10 hagesq hincsq wincsq hwincome
```

```
  Source |       SS       df       MS              Number of obs =     893
---------+------------------------------           F( 25,   867) =    3.62
   Model |  3.4668e+09    25   138673295           Prob > F      =  0.0000
Residual |  3.3215e+10   867  38309872.5           R-squared     =  0.0945
---------+------------------------------           Adj R-squared =  0.0684
   Total |  3.6681e+10   892  41122748.7           Root MSE      =  6189.5

------------------------------------------------------------------------------
  dontot |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    city |  -815.7664   456.4707     -1.787   0.074      -1711.683    80.15046
  nuirec |   1387.813    1036.65      1.339   0.181      -646.8238    3422.449
    hage |  -549.5136   309.0352     -1.778   0.076      -1156.058    57.03099
    wage |   -2.95748   58.19512     -0.051   0.959      -117.1773    111.2623
   heduc |   227.9706   216.1409      1.055   0.292      -196.2499    652.1912
   weduc |   187.8909   226.5848      0.829   0.407      -256.8279    632.6097
    hinc |   .0138123   .0483688      0.286   0.775      -.0811213    .1087459
    winc |  -.0279641   .0564535     -0.495   0.620      -.1387655    .0828374
    hocc |  -15.83118    65.2511     -0.243   0.808      -143.8998    112.2374
    wocc |  -51.81263   115.8241     -0.447   0.655      -279.1411    175.5158
    dvp1 |   2260.092   2116.013      1.068   0.286      -1893.015    6413.199
    dvp2 |   347.1528   1177.402      0.295   0.768      -1963.739    2658.045
    dvp3 |  -716.0433   1373.261     -0.521   0.602      -3411.348    1979.261
    dvp4 |   252.0214   1159.193      0.217   0.828      -2023.13     2527.173
    dvp5 |   330.5415   1085.067      0.305   0.761      -1799.124    2460.207
    dvp6 |  -1523.192   897.8458     -1.696   0.090      -3285.397    239.0139
    dvp7 |  -326.2074   838.0809     -0.389   0.697      -1971.112    1318.697
    dvp8 |  -103.7304   1218.645     -0.085   0.932      -2495.569    2288.108
    dvp9 |   1916.459   1011.838      1.894   0.059      -69.48015    3902.398
   dvp10 |   1734.798   990.4305      1.752   0.080      -209.1242     3678.72
  hagesq |   7.596167   3.851476      1.972   0.049       .0368611    15.15547
  hincsq |  -6.15e-08   1.42e-07     -0.434   0.664      -3.40e-07    2.17e-07
  wincsq |   4.15e-07   2.16e-07      1.922   0.055      -8.78e-09    8.38e-07
hwincome |   1.02e-06   4.27e-07      2.376   0.018       1.77e-07    1.85e-06
   _cons |   11889.25   6162.088      1.929   0.054      -205.1008    23983.61
------------------------------------------------------------------------------
```

-> . regress dontot city nuirec hage wage heduc weduc hinc winc hocc wocc dvp1 dvp2
dvp3 dvp4 dvp5 dvp6 dvp7 dvp8 dvp9 dvp10 hagesq hincsq wincsq hwincome

```
  Source |       SS       df       MS              Number of obs =     893
---------+------------------------------           F( 25,   867) =    3.62
   Model |  3.4668e+09    25   138673295           Prob > F      =  0.0000
Residual |  3.3215e+10   867  38309872.5           R-squared     =  0.0945
---------+------------------------------           Adj R-squared =  0.0684
   Total |  3.6681e+10   892  41122748.7           Root MSE      =  6189.5
```

dontot	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
city	-815.7664	456.4707	-1.787	0.074	-1711.683	80.15046
nuirec	1387.813	1036.65	1.339	0.181	-646.8238	3422.449
hage	-549.5136	309.0352	-1.778	0.076	-1156.058	57.03099
wage	-2.95748	58.19512	-0.051	0.959	-117.1773	111.2623
heduc	227.9706	216.1409	1.055	0.292	-196.2499	652.1912
weduc	187.8909	226.5848	0.829	0.407	-256.8279	632.6097
hinc	.0138123	.0483688	0.286	0.775	-.0811213	.1087459
winc	-.0279641	.0564535	-0.495	0.620	-.1387655	.0828374
hocc	-15.83118	65.2511	-0.243	0.808	-143.8998	112.2374
wocc	-51.81263	115.8241	-0.447	0.655	-279.1411	175.5158
dvp1	2260.092	2116.013	1.068	0.286	-1893.015	6413.199
dvp2	347.1528	1177.402	0.295	0.768	-1963.739	2658.045
dvp3	-716.0433	1373.261	-0.521	0.602	-3411.348	1979.261
dvp4	252.0214	1159.193	0.217	0.828	-2023.13	2527.173
dvp5	330.5415	1085.067	0.305	0.761	-1799.124	2460.207
dvp6	-1523.192	897.8458	-1.696	0.090	-3285.397	239.0139
dvp7	-326.2074	838.0809	-0.389	0.697	-1971.112	1318.697
dvp8	-103.7304	1218.645	-0.085	0.932	-2495.569	2288.108
dvp9	1916.459	1011.838	1.894	0.059	-69.48015	3902.398
dvp10	1734.798	990.4305	1.752	0.080	-209.1242	3678.72
hagesq	7.596167	3.851476	1.972	0.049	.0368611	15.15547
hincsq	-6.15e-08	1.42e-07	-0.434	0.664	-3.40e-07	2.17e-07
wincsq	4.15e-07	2.16e-07	1.922	0.055	-8.78e-09	8.38e-07
hwincome	1.02e-06	4.27e-07	2.376	0.018	1.77e-07	1.85e-06
_cons	11889.25	6162.088	1.929	0.054	-205.1008	23983.61

-> . test hinc=winc hincsq=wincsq=hwincome*.5

(1) hinc - winc = 0.0 hincsq – wincsq – hwincome*5 = 0.0

F(1, 867) = 4.91
Prob > F = 0.0847

Do previous smoking behaviors affect adolescent smoking status?

Modeled from the paper, “Does cigarette price influence adolescent experimentation?”
from the Journal of Health Economics, 20(2): 261-270
By Sherry Emery, Martha M. White, John P. Pierce
Dataset #386: Youth Smoking Survey

March 25, 2001

INTRODUCTION

The paper we selected is entitled "Does cigarette price influence adolescent experimentation?" by Emery, White, and Pierce. The objective of their paper was to adequately measure the adolescents' price sensitivity to smoking as well as test whether this price responsiveness is affected by smoking experience. They chose their data set from the 1993 US national survey of youth smoking: TAPS (longitudinal Teenage Attitudes and Practices Survey). They estimated a two-part model of demand, the first one being a model of smoking participation, and the second part was a model of conditional demand for cigarettes among current and established smokers. These models are represented by:

$$SMOKE_i = \alpha PRICE_i + \beta_1 TOBACON_i + \beta_2 SOCDEM_i + \beta_3 PSYCHSOC_i + \varepsilon_I$$

Where PRICE is real price/pack of cigarettes, determined by the US consumer price index (CPI), and TOBACON is an index that indicates state-level tobacco control activity. SOCDEM represents a set of socio-demographic variables, such as gender, parental marital status, job possession, and household income. PSYCHSOC is representative of a group of psycho-social variables like school performance, parental bond, and family smoking, and belief that there are health risks associated with the occasional cigarette. The survey targeted adolescents in the range of 10-22 years of age, but the study restricted the sample to those ≥ 14 years old. The authors of the study utilized the answers to various questions in order to categorize the respondents into experimental, current, and established smokers. Some of their measures of smoking behaviour include questions such as: “Have you smoked a cigarette?”, “Have you smoked more than 100 cigarettes in your lifetime?", and "Have you smoked in the past 30 days?". Those who had smoked in the past 30 days were identified as current smokers. Established smokers were defined as those who had smoked in the past 30 days, as well as those that attested to smoking at least 100 cigarettes in their lives. Experimenters answered positively to trying a cigarette, but had not smoked within 30 days prior to the survey. The model was regressed for each category, and a table of parameter estimations for adolescent smoking behaviour was provided.

The results from these tests coincide with previous studies in that price and other state-controlled policies are not significantly associated with smoking behaviours of the adolescents in the experimental group. This can be attributed to the sporadic consumption of cigarettes by experimenters. However, price is an important indicator of smoking behaviours within more advanced smokers, i.e. established smokers. Their estimates suggest that price plays an influential role in moderating the amount smoked within adolescents.

DATA AND MODEL SELECTION

The data set we used was the Youth Smoking Survey, which contains 9,491 observations from adolescents throughout Canada. The ages of the respondents varied from 15-19 years, whereas the TAPS surveyed people who were in between the ages of 10 and 22. The Youth Smoking Survey included subjects from all provinces, whereas TAPS used respondents from only 48 states.

We eliminated the adolescents who responded negatively to the question, "Have you ever tried cigarette smoking?". We divided the remaining subjects into two categories: those who had smoked within the last year, and those who hadn't. We labeled those who hadn't smoked within the last year, along with those who were labeled "valid skip" as "non-smokers". Those respondents who didn't answer the question (i.e. "not stated") were dropped from the sample. From this, we calculated the sample populations for both "Smokers" (n = 2530) and "Non-smokers" (n = 3758).

ESTIMATION AND RESULTS

We replicated a similar model as described in the paper, using 9 unique variables. We selected a list of variables to describe our independent variable, average number of cigarettes smoked per day (NCIGS):

$$NCIGS_i = \beta_1 + \beta_2 SEX_i + \beta_3 JOB_i + \beta_4 AGE_i + \beta_5 HOME_i + \beta_6 FAPP_i + \beta_7 MAPP_i + \beta_8 FSMOKE_i + \beta_9 MSMOKE_i + \beta_{10} RISK_i + \varepsilon_i$$

Where SEX_i is the gender of the respondent, JOB_i represents whether or not the subject holds a job, AGE_i is the age (in years) of the adolescent, $HOME_i$ indicates whether or not the subject smokes in their own home, $FAPP_i$ defines whether the father approves of the adolescent smoking while $MAPP_i$ is whether the mother approves of the adolescent smoking, $FSMOKE_i$ signifies whether the father of the subject smokes and $MSMOKE_i$ is whether the mother of the subject smokes, and finally, $RISK_i$ indicates whether the subject believes there to be a health risk associated with an occasional cigarette.

After selecting the descriptive variables, we tabulated the summary statistics for both non-smokers and smokers, which are provided in Table 1.

Variable	Smokers (Smoke = 1), n = 2530	Non-Smokers (Smoke = 0), n = 3758
Male (%)	51.5 (50.0)	49.6 (50.0)
Working at a job (%)	54.2 (49.8)	53.4 (49.9)
Age (yrs.)	17.1 (1.4)	16.8 (1.4)
Smokes in own home (%)	59.6 (49.1)	1.1 (10.5)
Father approves (%)	3.8 (19.1)	0.05 (2.3)
Mother approves (%)	4.8 (21.4)	0.05 (2.3)
Father smokes (%)	42.1 (49.4)	33.8 (47.3)
Mother smokes (%)	42.2 (49.4)	29.8 (45.8)
Health risk (%)	61.3 (48.7)	65.6 (47.5)

TABLE 1: DESCRIPTIVE STATISTICS → MEAN (STANDARD DEVIATION)

We generated variables that interacted the dependent variables with the dummy variable, which specifies whether they had smoked in the past year. We then regressed the average number of cigarettes smoked per day variable (ncigs) on these newly generated interaction terms, along with the original dependent variables, to produce parameter estimates. Table 2 presents the parameter estimates from the regression, with the respective p-values in parenthesis.

Dependent variables	
Male	$1.74e^{-11}$ (1.000)
Job possession	$-2.13e^{-11}$ (1.000)
Age	$4.47e^{-11}$ (1.000)
Smokes in own home	$3.04e^{-10}$ (1.000)
Father approves	$2.77e^{-10}$ (1.000)
Mother approves	$-7.19e-^{10}$ (1.000)
Father smokes	$-2.22e^{-11}$ (1.000)
Mother smokes	$3.14e^{-11}$ (1.000)
Health risk	$-9.18e^{-13}$ (1.000)
Dummy variable	
Smoked in last year	16.815 (0.346)
Interaction variables	
Sex * smoked in last year	-14.184 (0.000)
Job possession * smoked in last year	-4.154 (0.054)
Age * smoked in last year	-8.846 (0.000)
Smokes in own home * smoked in last year	-26.255 (0.000)
Father approves * smoked in last year	-2.189 (0.513)
Mother approves * smoked in last year	-15.480 (0.000)
Father smokes * smoked in last year	8.574 (0.000)
Mother smokes * smoked in last year	10.823 (0.000)
Health risk * smoked in last year	5.372 (0.000)
Intercept	$7.40e^{-11}$ (1.000)

TABLE 2: PARAMETER ESTIMATES AND RELATIVE P-VALUES

We set up a hypothesis test to see whether the variables had a significant impact on the average number of cigarettes smoked per day. After taking the expected values of each category (non-smokers and smokers) and taking the difference of these values, we arrive at our null hypothesis:

H_o: $\beta_2 = \beta_4 = \beta_6 = \beta_8 = \beta_{10} = \beta_{12} = \beta_{14} = \beta_{16} = \beta_{18} = \beta_{20} = 0$
H_A: $\beta_2 \neq 0$ &/or $\beta_4 \neq 0$ &/or $\beta_6 \neq 0$ &/or $\beta_8 \neq 0$ &/or $\beta_{10} \neq 0$ &/or $\beta_{12} \neq 0$ &/or $\beta_{14} \neq 0$ &/or $\beta_{16} \neq 0$ &/or $\beta_{18} \neq 0$ &/or $\beta_{20} \neq 0$

We performed this test on STATA at the 5% significance level, and generated an F-statistic of 125.23, with a p-value of 0.000. Given these values, we can reject the null hypothesis in favour of the alternative, at any significance level, that the variables have no impact on the number of cigarettes smoked per day. The interaction terms of the dummy variable with job possession and father’s approval have shown to be the most significant of the interacted variables (with p-values of 0.054 and 0.513 respectively).

SUMMARY

Upon further analysis, another interesting correlation can be drawn from the data. For instance, out of 2,411 respondents labeled as "Smokers", only 3.8% of their fathers and 4.8% of their mothers approve of their children's smoking habits. However, of these smokers, 42% have at least one parent that smokes as well. This supports the argument that children tend to mimic their parents' behavior, regardless of their parents' approval. Also, the high number of parental disapproval for their children's smoking habits is indicative of their concern for the health of their kids. Therefore, the risks associated with cigarette smoking are realized by parents and more established smokers than within adolescents and experimental smokers.

The model paper concludes that price does affect the level of cigarette consumption within certain types of smokers, yet further studies need to be done to find what factors lead to the deterence of youth smoking in order to decrease the amount of smokers within the adolescent population. There are many factors in determining how much adolescent experimentation occurs, and how many cases of experimentation lead to established smoking patterns. Surveys and consequent studies should focus upon alternative public policy approaches that specifically address experimentation with smoking.

```
********************************************************************************

This is a Stata log file for a QED session

Course: Econ 452
Students:
Date and time: Fri, 23 Mar 2001, 15:22:10


At the end of the QED session, this file will be copied to:
82_265_Fri_ng.log
82_265_Fri_cao.log
These files will also be uploaded to:
http://edith.econ.queensu.ca/statausr/logfiles/Econ452
Type help QEDstata for a list of QED commands

Student work begins below this line
********************************************************************************


pause:  "Type BREAK to end session started at 23 Mar 2001 15:22:10"
-> . do a:proj3

. * Project3 Do-File Commands*/
. #delimit;
delimiter now ;
. Qextract QEDid age sex q72_62 q11a_19a q3_9a q26_40 q27_36 q28_38 q41_35a q43_37a q49b_44b
dvamtsmk, ds(386);
getting information about file 386 ...
loading variables from 386 (yss94lfs) only (no data yet)...  done

. /* Extracts the variables from the data set */
>
> keep if q3_9a == 1;
(3203 observations deleted)

. /* Criteria --> Person must answer: 'yes' to having tried cigs */
>
> tab q11a_19a, gen(dhab);

   smoked in |
   last week |      Freq.     Percent        Cum.
-------------+-----------------------------------
   did smoke |       2530       40.24       40.24
did not smoke |       285        4.53       44.77
  valid skip |       3424       54.45       99.22
  not stated |         49        0.78      100.00
-------------+-----------------------------------
       Total |       6288      100.00

. /* creates dummy variable whether or not smoked in the last year */
>
> gen smoke = q11a_19a;
```

```
. /* Generates new variable = # smokers */
>
> drop if smoke == 9;
(49 observations deleted)

. replace smoke = 0 if  smoke > 1;
(3709 real changes made)

. /* Changes all other responses to 'whether smoked' =  non-smokers --> explicit assumption1*/
>
> gen ncigs = dvamtsmk;

. replace ncigs = 0 if dvamtsmk == 996;
(3709 real changes made)

. replace ncigs = . if dvamtsmk == 999;
(119 real changes made, 119 to missing)

. drop if ncigs == .;
(119 observations deleted)

. /* Drops missing values */
>
> /*To generate summary statistics:*/
>
> tab sex, gen(dsex);

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
      male? |       3083       50.38       50.38
    female? |       3037       49.62      100.00
------------+-----------------------------------
      Total |       6120      100.00

. summ dsex1 if smoke==0;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   dsex1 |    3709    .4963602   .5000542          0          1

. summ dsex1 if smoke==1;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   dsex1 |    2411    .5151389   .4998744          0          1

. tab q72_62, gen(djob);

        job |
 possession |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       3284       53.66       53.66
         no |       2833       46.29       99.95
 not stated |          3        0.05      100.00
```

```
------------+-----------------------------------
      Total |       6120      100.00

. summm djob1 if smoke==0;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   djob1 |    3709    .5332974    .4989573          0          1

. summm djob1 if smoke==1;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   djob1 |    2411    .5416839    .4983628          0          1

. summm age if smoke==0;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
     age |    3709    16.84255     1.39289         15         19

. summm age if smoke==1;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
     age |    2411    17.05848    1.404314         15         19

. tab q26_40, gen(home);

  smokes in |
   own home |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       1477       24.13       24.13
         no |       1204       19.67       43.81
 valid skip |       3424       55.95       99.75
 not stated |         15        0.25      100.00
------------+-----------------------------------
      Total |       6120      100.00

. summm home1 if smoke==0;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   home1 |    3709    .0110542    .1045703          0          1

. summm home1 if smoke==1;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
   home1 |    2411    .5956035    .4908767          0          1

. tab q27_36, gen(fapp);

                      father |      Freq.     Percent        Cum.
```

```
-----------------------------+-----------------------------------
              he approves |         93        1.52        1.52
           he doesn't care |        598        9.77       11.29
         he doesn't like it |       1166       19.05       30.34
he doesn't know that i smoke |        574        9.38       39.72
    i don't have a father * |        235        3.84       43.56
               valid skip |       3424       55.95       99.51
               not stated |         30        0.49      100.00
-----------------------------+-----------------------------------
                    Total |       6120      100.00

. summ fapp1 if smoke==0;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   fapp1 |    3709    .0005392   .0232182          0          1

. summ fapp1 if smoke==1;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   fapp1 |    2411    .0377437   .1906152          0          1

. tab q28_38, gen(mapp);

                    mother |      Freq.     Percent        Cum.
------------------------------+-----------------------------------
              she approves |        118        1.93        1.93
           she doesn't care |        570        9.31       11.24
         she doesn't like it |       1464       23.92       35.16
she doesn't know that i smoke |        475        7.76       42.92
     i don't have a mother * |         48        0.78       43.71
                valid skip |       3424       55.95       99.66
                not stated |         21        0.34      100.00
------------------------------+-----------------------------------
                     Total |       6120      100.00

. summ mapp1 if smoke==0;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   mapp1 |    3709    .0005392   .0232182          0          1

. summ mapp1 if smoke==1;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   mapp1 |    2411    .0481128    .214049          0          1

. tab q41_35a, gen(fsmoke);

              father |      Freq.     Percent        Cum.
------------------------+-----------------------------------
                 yes |       2268       37.06       37.06
```

```
                     no |       3506       57.29       94.35
i don't have a father * |        300        4.90       99.25
             don't know |         32        0.52       99.77
             not stated |         14        0.23      100.00
------------------------+-----------------------------------
                  Total |       6120      100.00

. summ fsmoke1 if smoke==0;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
 fsmoke1 |    3709    .3375573    .4729405          0          1

. summ fsmoke1 if smoke==1;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
 fsmoke1 |    2411    .4214019    .4938861          0          1

. tab q43_37a, gen(msmoke);

                 mother |      Freq.     Percent        Cum.
------------------------+-----------------------------------
                    yes |       2124       34.71       34.71
                     no |       3924       64.12       98.82
i don't have a mother * |         45        0.74       99.56
             don't know |         13        0.21       99.77
             not stated |         14        0.23      100.00
------------------------+-----------------------------------
                  Total |       6120      100.00

. summ msmoke1 if smoke==0;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
 msmoke1 |    3709    .2981936    .4575266          0          1

. summ msmoke1 if smoke==1;

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
 msmoke1 |    2411    .4222314    .4940175          0          1

. tab q49b_44b, gen(risk);

health risk |
       from |
 occasional |
  cigarette |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       3911       63.91       63.91
         no |       2005       32.76       96.67
 don't know |        190        3.10       99.77
 not stated |         14        0.23      100.00
```

```
------------+-----------------------------------
      Total |       6120      100.00

. summm risk1 if smoke==0;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   risk1 |    3709     .655972   .4751143          0          1

. summ risk1 if smoke==1;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   risk1 |    2411    .6130236   .4871592          0          1

. /*Regression of variables by creating interaction terms.*/
>
>  gen sexdhab1 = sex*dhab1;

. gen agedhab1 = age*dhab1;

. gen homdhab1 = q26_40*dhab1;

. gen fapdhab1 = q27_36*dhab1;

. gen mapdhab1 = q28_38*dhab1;

. gen fsmdhab1 =  q41_35a*dhab1;

. gen msmdhab1 = q43_37a*dhab1;

. gen rskdhab1 = q49b_44b*dhab1;

. gen jobdhab1 = q72_62*dhab1;

. regress ncigs dhab1 sex sexdhab1 q72_62 jobdhab1 age agedhab1 q26_40 homdhab1 q27_36 fapdhab1
q28_38 mapdhab1 q41_35a fsmdhab1 q43_37a msmdhab1 q49b_44b rskdhab1;

  Source |       SS       df       MS                  Number of obs =    6120
---------+------------------------------               F( 19,  6100) =  401.36
   Model |  14442067.2    19   760108.80               Prob > F      =  0.0000
Residual |  11552407.5  6100  1893.83729               R-squared     =  0.5556
---------+------------------------------               Adj R-squared =  0.5542
   Total |  25994474.7  6119  4248.15732               Root MSE      =  43.518

------------------------------------------------------------------------------
   ncigs |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   dhab1 |   16.81541   17.82806      0.943   0.346      -18.13387    51.76469
     sex |   1.74e-11   1.433509      0.000   1.000      -2.810184    2.810184
sexdhab1 |  -14.18429   2.293374     -6.185   0.000      -18.68011   -9.688469
  q72_62 |  -2.13e-11   1.393649      0.000   1.000      -2.732044    2.732044
jobdhab1 |   -4.15395   2.151983     -1.930   0.054      -8.372597    .0646962
     age |   4.47e-11   .5145252      0.000   1.000      -1.008651    1.008651
```

```
agedhab1 |   8.846009   .8269603     10.697   0.000       7.224875    10.46714
  q26_40 |   3.04e-10   2.033621      0.000   1.000      -3.986615    3.986615
homdhab1 |  -26.25456   2.628243     -9.989   0.000      -31.40685   -21.10228
  q27_36 |   2.77e-10   3.171268      0.000   1.000      -6.216804    6.216804
fapdhab1 |  -2.189185    3.34344     -0.655   0.513      -8.743506    4.365136
  q28_38 |  -7.19e-10    3.56988      0.000   1.000      -6.998224    6.998224
mapdhab1 |  -15.48041   3.799516     -4.074   0.000       -22.9288   -8.032013
 q41_35a |  -2.22e-11   1.103071      0.000   1.000      -2.162408    2.162408
fsmdhab1 |   8.573838   1.740675      4.926   0.000         5.1615    11.98618
 q43_37a |   3.14e-11   1.334784      0.000   1.000      -2.616648    2.616648
msmdhab1 |    10.8228   2.081453      5.200   0.000       6.742417    14.90318
q49b_44b |  -9.18e-13   .6096182      0.000   1.000      -1.195067    1.195067
rskdhab1 |   5.372864   1.075629      4.995   0.000       3.264251    7.481477
   _cons |   7.40e-11   12.69186      0.000   1.000      -24.88053    24.88053
------------------------------------------------------------------------------

. test dhab1 sexdhab1 agedhab1 fapdhab1 mapdhab1 fsmdhab1 msmdhab1 rskdhab1 jobdhab1 homdhab1;

 ( 1)  dhab1 = 0.0
 ( 2)  sexdhab1 = 0.0
 ( 3)  agedhab1 = 0.0
 ( 4)  fapdhab1 = 0.0
 ( 5)  mapdhab1 = 0.0
 ( 6)  fsmdhab1 = 0.0
 ( 7)  msmdhab1 = 0.0
 ( 8)  rskdhab1 = 0.0
 ( 9)  jobdhab1 = 0.0
 (10)  homdhab1 = 0.0

       F( 10,  6100) =  125.23
            Prob > F =    0.0000
```

Econ 452B

Using OLS Regression techniques to estimate Chaloupka and Wechsler's "Price, Tobacco Control Policies and Smoking among young adults"[*]

By:

Data Set: Survey of Smoking in Canada (# 373)

[*] ***"Price, Tobacco Control Policies and Smoking among young adults", Journal of Health Economics, Vol:16, 1997, pp 359-373***

Introduction

Much is known about the health consequences of cigarette smoking and other forms of tobacco use. Despite this cigarette smoking remains high, especially among the youths and young adults. After declining rapidly throughout 1970's, the decrease in smoking participation has reached a plateau in recent years.

Chaloupka and Wechsler's paper (1997) examines the effectiveness of two major tobacco control policies in discouraging smoking among young adults. It studies the effects of cigarette prices (which are directly proportional to changes in excise taxes) and restrictions on smoking in public places on smoking participation among students in U.S. colleges and universities. The reason this age group was chosen is that at this age smoking practices become firmly established. Almost all smokers first use cigarettes by high school graduation, while nearly no first use occurs after age 20. Almost 45% of people who ever smoked daily began daily smoking between the ages of 18 and 29 years. Therefore, a sample of college and university students was used by Chaloupka and Wechsler to capture this age range and to thus explore how price and policies may be used to discourage smoking.

The data utilized in the model paper was taken from the 1993 Harvard College Alcohol Study, which focused on binge drinking in colleges. In the study, all respondents were asked about their current/past smoking participation as well as their daily cigarette consumption. These data were used to construct various variables that reflect cigarette demand (like smoking participation, frequency of cigarette consumption, average daily cigarette consumption and so on). Additionally, a variety of independent variables were constructed to control for other factors affecting cigarette demand, which include age, age squared, indicators of gender, race/ethnicity, marital status, parental education, on-campus living, fraternity or sorority membership and employment. Indicators of price were included, by Chaloupka and Wechsler, using site-specific data, while that on restrictions included a set of five dichotomous indicators including restrictions in the workplace, schools, restaurants, retail stores and other public places. Chaloupka and Wechsler test the effect of age on cigarette consumption as well as how price and various restrictions affect cigarette consumption.

In their paper Chaloupka and Wechsler do not use ordinary least squares techniques and deem them “inappropriate" for this data set, due to the "limiting" nature of the data set. Instead, two alternative approaches are used. In the fist step, probit measures are used to estimate a smoking participation equation. In the second step, least square methods are used to estimate average daily cigarette consumption by smokers.

Based on these testing procedures, they find that age is significant in affecting cigarette consumption. They report that the price of cigarettes has a negative and statistically significant impact on smoking amongst college students. Finally they also conclude that policies restricting smoking in public places and private work-sites appear also to have a negative impact but of a smaller scale as compared to price effects.

Our analysis attempts the more modest goal of establishing whether or not average cigarette consumption is affected by various restrictions on smoking in the workplace, as well as whether or not a variety of other variables affect cigarette consumption.

Information for our analysis is adapted from the Survey of Smoking in Canada (SOIC). This survey was conducted in 1995 across all provinces and involved telephone interviews with 15, 804 individuals aged 15 to 78 years old. The survey deals with the smoking patterns, practices and beliefs of interviewees. SOIC is useful in a variety of ways. Firstly, it contains information on a random sample of Canadians within the appropriate age group, who are either subject to or not subject to a variety of smoking restrictions. Secondly, it details individuals' actual average cigarette usage as a 'continuous' variable, in contrast to Chalupka and Weschsler's (1995) constructed continuous variable, which is essentially the mid-point of a variety of consumption ranges. Additionally, most of the variables in the model paper, including data on age, marital status, sex, and workplace restrictions are available within the SOIC. The data on workplace restrictions is particularly useful as it lays out a variety of levels of restrictions for comparison.

Despite these favourable aspects of the SOIC, there are many disadvantages which persist. Firstly, we have data only on restrictions within the workplace, thus forcing our analysis to focus on a sample of working individuals. Since is unlikely that one would not work because of smoking restrictions, we can eliminate non-working individuals to focus on how daily cigarette consumption differs among working individuals who face a variety of levels of smoking restrictions in their place of work. This serves as the main distinguishing feature between our analysis and the model paper which focused on university students. However, as the emphasis of both analyses is on cigarette consumption within a particular age group, rather than occupation status, this assumption seems fairly robust. Finally, we do not have data on ethnicity/race or parental education, however the variable 'language spoken at home' will be used as an arguable substitute for ethnicity.

Data

The data used in our analysis was manipulated as follows. Firstly, our study focuses on young adults between the ages of 18 and 29, thus individuals outside of this range are excluded (11327). Secondly, individuals who were not working are excluded as we are interested in looking at how restrictions on smoking in the workplace affect average cigarette consumption in this younger portion of the labour force. This eliminates a further 3085 individuals. Also, individuals who smoke no cigarettes are excluded to focus on how smoking consumption varies among individuals who already smoke (3). Finally, individuals with incomplete data in any of the appropriate areas (average cigarette use, age, sex, marital status, language spoken and restrictions in the workplace) were excluded (895). After all this our final sample is reduced to 494 participants.

The variable containing information on restrictions ranges from restricted completely (24.5% of our sample), allowed to smoke only in designated areas (36.0%), restricted only in certain places (10.9%), not restricted at all (27.9%). To control for the effects of men and women facing differing work environments, we also explore the interactions of sex with restrictions on cigarette consumption.

Table 1.shows the summary statistics for the appropriate variables.

Results

Table 2. has been constructed to reflect the regression results of average cigarette consumption on the respective variables as well as dummy variables so generated to account for categorical data.

Table 2 shows that, based on our sample, differences in age, age-squared, sex, marital status, as well as language spoken have no significant effect on a person's average cigarette consumption. Additionally, Table 2 indicates that the difference between both people who are completely restricted from smoking at work and those who face no restrictions at work is significant in having an impact on an individuals average cigarette consumption (p-value = 0.000). It is also evident from the table that the difference between those who are allowed to smoke only in certain areas of the workplace and those who face no restrictions are significant (p-value = 0.000). The effect of restrictions on average cigarette consumption are strongly negative in both cases, and imply that an increase in 'restrictions' reduce average daily cigarette consumption. It is also evident that the coefficients of both these types of restrictions are, together, significantly different from zero- thus supporting our previous conclusion (F= 6.86, prob>F = 0.000).

The effect of being male and having various restrictions in the workplace is insignificant, thus indicating that multicollinearity is not a factor here. More specifically, the difference between one being male and facing a particular restriction is not significantly different from one being female and having the same restriction in affecting average cigarette consumption. Table 2 again stands testimony to these results.

Discussion and Conclusions

While numerous econometric studies of cigarette demand have been published over the past several decades, most of these studies have used diverse data and methods to estimate the effects of cigarette prices and taxes on smoking participation and cigarette consumption in the overall population. One general conclusion emerges from such studies: higher cigarette prices significantly reduce cigarette smoking. However, relatively few studies use individual-level data to focus on the price responsiveness of cigarette smoking among youths and young adults. Furthermore, a more recent phenomenon in the anti-smoking campaign is restrictions on smoking in public places, where studies have gradually taken precedence only recently. Thus, the research undertaken by Chaloupka and Wechsler addresses these issues by studying the impact of cigarette prices and restrictions on smoking in public places and private worksites on smoking participation and the frequency of cigarette consumption in a large, nationally representative sample of college/university students. This paper attempts to reflect upon the results of the article by considering the effect of restrictions on smoking in the work place on the average daily consumption of cigarettes by individuals in the appropriate age group.

The results gathered from the OLS estimation indicate that there is a statistically significant difference between no restrictions employed compared with (1)full restrictions employed and (2)restrictions for smoking only in certain areas on average cigarette consumption. This ties in well with Chaloupka and Wechsler's results, which indicate that "relatively stringent limits on smoking in public places and workplaces can influence the decision to smoke by young adults". However, due to skipped, invalid or unknown data, present in our survey (which were subsequently eliminated), our sample size was greatly reduced. Furthermore, the fact that our survey data was only indirectly related to the journal article required the elimination of further records to more closely model our data to that of the article, as well as the use of proxy variables in the place of more appropriate ones. The sum of all this is that these factors may have affected the robustness of our OLS estimation. This reasoning may be applied to our finding that age did not significantly affect average cigarette consumption. This differs from the findings in other empirical studies such as that done in our model paper as well as that conducted by Grossman et al (1983). Both these studies found that age is estimated to be statistically significant in affecting cigarette consumption. This discrepancy may be due to the limited sampling size as mentioned above.

The following discussion will briefly outline some of the issues and concerns that we came across in undertaking the survey data study. It is interesting that the authors chose to study the effects of cigarette prices and restrictions on smoking in public places and private workplaces among students in colleges/universities. Clearly, this is an interesting age group to study since it contains individuals whose smoking practices are likely to be in the process of becoming firmly set. Thus, it is important to use such an age range during which policies to discourage smoking can have a significant impact. As well, given that college/university students will be the leaders of tomorrow and will shape future public policy, understanding the determinants of cigarette smoking for this population is particularly important. The fact that restrictions in the workplace have a significant impact suggests a possible role for them as a policy instrument.

A few problem areas with the original journal article as well as results based on our survey data stand out quite clearly. Firstly, it is highly probable that cigarette smoking is underreported in survey data. However, no information is available on the extent of underreporting or on how underreporting varies with consumption. Assuming that all smokers underreport by the same degree then this may not affect the estimates for policy variables. The loss of information for many individuals reduces the certainty of our results. For example, for the variable "average number of cigarettes smoked daily", it is calculated that approximately 72.3% of the survey data set is coded missing, and these are not valid skips. Equally the model article does not seem able to underline sufficient information to estimate cigarette demand equations or other economic models of addictive behavior. As well, details on the selection of colleges and universities, the sampling procedure, questionnaire, and response rates for this survey were lacking in the model article.

It is unclear as to why the authors decided to use ordered probit methods and least squares methods instead of ordinary least squares. It is only observed that the probit methods provide a general sense of the relationships between prices, policies, and cigarette smoking. However, is it due to the lack of or limited nature of dependent variables that the ordinary least squares method cannot be executed? Or is it a problem regarding the specific independent variable; that is, smoking participation or average daily consumption that renders the OLS technique inappropriate? Such questions need to be clarified more thoughtfully by the article. Our OLS estimates concur in terms of restrictions on smoking in the workplace, but not in terms of the effect of age on smoking. Nonetheless, Chaloupka and Wechsler's contribution proves to be significant in the study of tobacco control policies and smoking among young adults.

Appendix

TABLE 1
Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Average # of cigarettes smoked daily	494	1259.567	870.3722	3	5357
Age Group	494	22.82389	2.95115	18	29
Sex	494	1.481781	.5001745	1	2
Marital Status	494	3.293522	1.253344	1	4
Language spoken at home	494	1.251012	.6123716	1	4
Smoking restrictions at date of work	494	2.453441	1.191952	1	7

Table 2
Regression Results from OLS Estimation

Variable	Co-efficient (Standard Error)
Age	332.98 (173.34)
Age^2	- 6.49 (3.71)
Sex	203.22 (78.69)
Marital Status – Married	-225.85 (248.60)
Marital Status – Single	-417.20 (244.62)
Language Spoken at home – English	356.99 (236.08)
Language Spoken at home – French	277.88 (254.25)
Language Spoken at home – Both English and French	669.83 (391.85)
Smoking Restrictions at place of work - Restricted Completely	-512.95 (107.78)
Smoking Restrictions at place of work - Allowed only in designated places	-350.34 (96.07)
Smoking Restrictions at place of work - Restricted only in certain places	-59.06 (133.93)
Smoking Restrictions at place of work - Not Restricted at all	-57.95 (491.09)

Break-up by Age group:

Age	Freq.	Percent	Cum.
18	36	7.29	7.29
19	37	7.49	14.78
20	39	7.89	22.67
21	58	11.74	34.41
22	68	13.77	48.18
23	59	11.94	60.12
24	83	16.80	76.92
25	21	4.25	81.17
26	27	5.47	86.64
27	20	4.05	90.69
28	20	4.05	94.74
29	26	5.26	100.00
Total	**494**	**100.00**	

Break-up by Marital Status:

Marital Status	Freq.	Percent	Cum.
Married	112	22.67	22.67
Separated-Divorced	13	2.63	25.30
Single	369	74.70	100.00
Total	**494**	**100.00**	

Break-up by Sex:

Sex	Freq.	Percent	Cum.
Male	256	51.82	51.82
Female	238	48.18	100.00
Total	**494**	**100.00**	

Break-up by Language most often spoken at home:

Language most often spoken at home	Freq.	Percent	Cum.
English	403	81.58	81.58
French	71	14.37	95.95
Both equally	7	1.42	97.37
Other	13	2.63	100.00
Total	**494**	**100.00**	

Break-up by Smoking restrictions at place of work:

Smoking restrictions at place of work	Freq.	Percent	Cum.
Restricted completely	121	24.49	24.49
Allowed only in designated areas	178	36.03	60.53
Restricted only in certain places	54	10.93	71.46
Not restricted at all	138	27.94	99.39
Don't know	3	0.61	100.00
Total	**494**	**100.00**	

STATA LOG

```
********************************************************************************

This is a Stata log file for a QED session

Course: Econ 452
Students: wen
Date and time: Sun, 25 Mar 2001, 12:32:08


At the end of the QED session, this file will be copied to:
84_222_Sun_wen.log
These files will also be uploaded to:
http://edith.econ.queensu.ca/statausr/logfiles/Econ452
Type help QEDstata for a list of QED commands

Student work begins below this line
********************************************************************************


pause:  "Type BREAK to end session started at 25 Mar 2001 12:32:08"
-> . Qextract
getting information about file 373 ...
loading variables from 373 (sosic95) only (no data yet)...  done
-> . browse
-> . drop if agegp1 <18 | agegp1>29
(11327 observations deleted)
-> . gen agegp1sq = agegp1*agegp1
```

```
-> . browse
-> . drop if sex == .
(0 observations deleted)
-> . QEDmerge c2q21 , ds(373)
unrecognized command:  QEDmerge
r(199);
-> . Qmerge c2q21 , ds(373)
working some more ...

QEDid     QEDmerge  c2q21

 Was observation in memory, data set |
                       373, or both? |      Freq.     Percent
Cum.
-------------------------------------+---------------------------------
--
                obs. from using data |      11327       71.67
71.67
obs. from both master and using data |       4477       28.33
100.00
-------------------------------------+---------------------------------
--
                               Total |      15804      100.00
-> . browse
-> . drop if c2q21==.
(11099 observations deleted)
-> . browse
-> . drop if mst==.
(3025 observations deleted)
-> . drop if c4q28==.
(201 observations deleted)
-> . drop if c2q22==.
(0 observations deleted)
-> . browse
-> . Qmerge c1avgcig , ds(373)
QEDmerge already defined
r(110);
-> . browse
-> . drop QEDmerge
-> . browse
-> . Qmerge c1avgcig , ds(373)
working some more ...

QEDid     QEDmerge  c1avgcig

 Was observation in memory, data set |
                       373, or both? |      Freq.     Percent
Cum.
-------------------------------------+---------------------------------
--
                obs. from using data |      14325       90.64
90.64
obs. from both master and using data |       1479        9.36
100.00
-------------------------------------+---------------------------------
--
                               Total |      15804      100.00
```

```
-> . browse
-> . tab sex, no label
no invalid
r(198);
-> . tab sex, nolabel

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        743       50.24       50.24
          2 |        736       49.76      100.00
------------+-----------------------------------
      Total |       1479      100.00
-> . tab sex, nolabel, gen (fem)
invalid 'gen'
r(198);
-> . tab sex, nolabel, gen(fem)
invalid 'gen'
r(198);
-> . tab sex, nolabel gen(fem)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        743       50.24       50.24
          2 |        736       49.76      100.00
------------+-----------------------------------
      Total |       1479      100.00
-> . tab mst, nolabel gen(fem)

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        409       27.65       27.65
          3 |         30        2.03       29.68
          4 |       1040       70.32      100.00
------------+-----------------------------------
      Total |       1479      100.00
fem1 already defined
r(110);
-> . tab mst, nolabel gen(mst)

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        409       27.65       27.65
          3 |         30        2.03       29.68
          4 |       1040       70.32      100.00
------------+-----------------------------------
      Total |       1479      100.00
-> . tab mst, nolabel gen(mg)

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        409       27.65       27.65
          3 |         30        2.03       29.68
          4 |       1040       70.32      100.00
------------+-----------------------------------
```

```
      Total |       1479      100.00
-> . tab c4q28, nolabel gen(lang)

   language |
 most often |
   speak at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       1179       79.72       79.72
          2 |        238       16.09       95.81
          3 |         10        0.68       96.48
          4 |         52        3.52      100.00
------------+-----------------------------------
      Total |       1479      100.00
-> . drop mst1
-> . drop mst2
-> . drop mst3
-> . tab c2q22, nolabel gen(res)

    smoking |
restriction |
 s at place |
    of work |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        498       33.67       33.67
          2 |        476       32.18       65.86
          3 |        165       11.16       77.01
          4 |        335       22.65       99.66
          7 |          5        0.34      100.00
------------+-----------------------------------
      Total |       1479      100.00
-> . browse
-> . drop if sex==.
(14325 observations deleted)
-> . browse
-> . browse
-> . browse
-> . browse
-> . browse
-> . drop if  c1avgcig==.
(982 observations deleted)
-> . browse
-> . clear
-> . use "A:\final data.dta", clear
(373 : sosic95 : survey of smoking in canada)
-> . browse
-> . browse
-> . browse
-> . drop QEDmerge
-> . Qmerge c1ciguse , ds(373)
working some more ...

QEDid     QEDmerge  c1ciguse

 Was observation in memory, data set |
                       373, or both? |      Freq.     Percent
Cum.
```

```
------------------------------------+----------------------------------
--
                obs. from using data |      14325       90.64
90.64
obs. from both master and using data |       1479        9.36
100.00
------------------------------------+----------------------------------
--
                               Total |      15804      100.00
-> . browse
-> . browse
-> . drop if mst==.
(14325 observations deleted)
-> . browse
-> . save "A:\finaluse.dta"
file A:\finaluse.dta saved
-> . browse
-> . browse
-> . drop if  c1avgcig==.
(982 observations deleted)
-> . gen lnavgcig = ln(c1avgcig)
(3 missing values generated)
-> . browse
-> . browse
-> . drop if  c1avgcig==0
(3 observations deleted)
-> . summ

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   QEDid |     494    13565.24   1354.997      11361      15801
QEDmerge |     494           3          0          3          3
     age |     494          99          0         99         99
     sex |     494    1.481781   .5001745          1          2
     mst |     494    3.293522   1.253344          1          4
  agegp1 |     494    22.82389    2.95115         18         29
   c2q22 |     494    2.453441   1.191952          1          7
   c4q28 |     494    1.251012   .6123716          1          4
c4smuggl |     368    3.404891   1.113053          1          4
agegp1sq |     494    529.6215   137.9701        324        841
   c2q21 |     494    1.214575   .4792977          1          7
c1avgcig |     494    1259.567   870.3722          3       5357
    fem1 |     494    .5182186   .5001745          0          1
    fem2 |     494    .4817814   .5001745          0          1
     mg1 |     494    .2267206   .4191348          0          1
     mg2 |     494    .0263158    .160235          0          1
     mg3 |     494    .7469636   .4351924          0          1
   lang1 |     494    .8157895   .3880487          0          1
   lang2 |     494    .1437247    .351166          0          1
   lang3 |     494      .01417   .1183114          0          1
   lang4 |     494    .0263158    .160235          0          1
    res1 |     494    .2449393   .4304871          0          1
    res2 |     494    .3603239    .480581          0          1
    res3 |     494    .1093117   .3123462          0          1
    res4 |     494    .2793522   .4491357          0          1
    res5 |     494    .0060729   .0777704          0          1
c1ciguse |     494    3.117409   1.260405          2          6
```

```
lnavgcig |     494     6.66784   1.298693   1.098612    8.58616
-> . browse
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
3.86
   Model |  73.1000615    12  6.09167179               Prob > F      =
0.0000
Residual |   758.39583   481  1.57670651               R-squared     =
0.0879
---------+------------------------------               Adj R-squared =
0.0652
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2557

----------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+------------------------------------------------------------
-------
  agegp1 |   .4666611    .262477      1.778   0.076      -.0490822
.9824043
agegp1sq |  -.0092326    .005614     -1.645   0.101      -.0202635
.0017984
    fem1 |   .2320113   .1191534      1.947   0.052      -.0021143
.4661368
    fem2 |  (dropped)
     mg1 |  -.2828236    .376439     -0.751   0.453      -1.022492
.4568444
     mg2 |  (dropped)
     mg3 |  -.5308499   .3704168     -1.433   0.152      -1.258685
.1969851
   lang1 |   .2294103   .3574846      0.642   0.521       -.473014
.9318346
   lang2 |   .1814799    .384997      0.471   0.638      -.5750038
.9379637
   lang3 |   .7792978   .5933538      1.313   0.190      -.3865879
1.945183
   lang4 |  (dropped)
    res1 |  -.7111408   .2116747     -3.360   0.001      -1.127062   -
.2952194
    res2 |  -.4581373   .1979715     -2.314   0.021      -.8471332   -
.0691415
    res3 |  (dropped)
    res4 |  -.0041952   .2028035     -0.021   0.984      -.4026856
.3942951
    res5 |   .1567687   .7569808      0.207   0.836      -1.330629
1.644166
   _cons |   1.362201   3.074986      0.443   0.658      -4.679864
7.404266
----------------------------------------------------------------------
-------
```

```
-> . regress   c1avgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
   Total |   373471061   493  757547.792               Root MSE      =
829.25

-----------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
    fem2 |  (dropped)
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |   356.9957   236.0838      1.512   0.131      -106.8872
820.8787
   lang2 |   277.8875   254.2531      1.093   0.275      -221.6965
777.4715
   lang3 |   669.8344   391.8525      1.709   0.088      -100.1197
1439.789
   lang4 |  (dropped)
    res1 |  -453.8946   139.7906     -3.247   0.001      -728.5702    -
179.219
    res2 |  -291.2785   130.7409     -2.228   0.026      -548.1725   -
34.38456
    res3 |  (dropped)
    res4 |   59.06489    133.932      0.441   0.659      -204.0992
322.229
    res5 |   1.105058   499.9122      0.002   0.998      -981.1765
983.3866
   _cons |  -2786.574   2030.729     -1.372   0.171       -6776.77
1203.623
-----------------------------------------------------------------------
-------
-> . regress  lnavgcig  res1 res2 res3 res4 res5
```

```
  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  4,   489) =
7.00
   Model |  45.0099524     4  11.2524881               Prob > F      =
0.0000
Residual |  786.485939   489   1.6083557               R-squared     =
0.0541
---------+------------------------------               Adj R-squared =
0.0464
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2682

-----------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
-------
    res1 |  -.6752258   .2075489     -3.253   0.001      -1.083024   -
.2674281
    res2 |  -.4382136   .1970281     -2.224   0.027      -.8253397   -
.0510874
    res3 |  (dropped)
    res4 |   .0629174    .203566      0.309   0.757      -.3370545
.4628894
    res5 |   .4170866   .7522652      0.554   0.580      -1.060984
1.895158
   _cons |   6.971019   .1725815     40.393   0.000       6.631927
7.310112
-----------------------------------------------------------------------
-------
-> . regress  lnavgcig

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  0,   493) =
.
   Model |        0.00     0           .               Prob > F      =
.
Residual |  831.495892   493  1.68660424               R-squared     =
0.0000
---------+------------------------------               Adj R-squared =
0.0000
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2987

-----------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
-------
   _cons |    6.66784    .058431    114.115   0.000       6.553036
6.782645
-----------------------------------------------------------------------
-------
```

```
-> . browse
-> . regress c1avgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
   Total |   373471061   493  757547.792               Root MSE      =
829.25

-----------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
    fem2 |  (dropped)
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |   356.9957   236.0838      1.512   0.131      -106.8872
820.8787
   lang2 |   277.8875   254.2531      1.093   0.275      -221.6965
777.4715
   lang3 |   669.8344   391.8525      1.709   0.088      -100.1197
1439.789
   lang4 |  (dropped)
    res1 |  -453.8946   139.7906     -3.247   0.001      -728.5702    -
179.219
    res2 |  -291.2785   130.7409     -2.228   0.026      -548.1725   -
34.38456
    res3 |  (dropped)
    res4 |   59.06489    133.932      0.441   0.659      -204.0992
322.229
    res5 |   1.105058   499.9122      0.002   0.998      -981.1765
983.3866
   _cons |  -2786.574   2030.729     -1.372   0.171       -6776.77
1203.623
-----------------------------------------------------------------------
-------
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res4 res5
```

```
  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
3.86
   Model |  73.1000615    12  6.09167179               Prob > F      =
0.0000
Residual |   758.39583   481  1.57670651               R-squared     =
0.0879
---------+------------------------------               Adj R-squared =
0.0652
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2557

------------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
  agegp1 |   .4666611    .262477      1.778   0.076      -.0490822
.9824043
agegp1sq |  -.0092326    .005614     -1.645   0.101      -.0202635
.0017984
    fem1 |   .2320113   .1191534      1.947   0.052      -.0021143
.4661368
    fem2 |  (dropped)
     mg1 |  -.2828236    .376439     -0.751   0.453      -1.022492
.4568444
     mg2 |  (dropped)
     mg3 |  -.5308499   .3704168     -1.433   0.152      -1.258685
.1969851
   lang1 |   .2294103   .3574846      0.642   0.521       -.473014
.9318346
   lang2 |   .1814799    .384997      0.471   0.638      -.5750038
.9379637
   lang3 |   .7792978   .5933538      1.313   0.190      -.3865879
1.945183
   lang4 |  (dropped)
    res1 |  -.7111408   .2116747     -3.360   0.001      -1.127062   -
.2952194
    res2 |  -.4581373   .1979715     -2.314   0.021      -.8471332   -
.0691415
    res3 |  (dropped)
    res4 |  -.0041952   .2028035     -0.021   0.984      -.4026856
.3942951
    res5 |   .1567687   .7569808      0.207   0.836      -1.330629
1.644166
   _cons |   1.362201   3.074986      0.443   0.658      -4.679864
7.404266
------------------------------------------------------------------------
-------
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
```

```
---------+------------------------------              F( 12,   481) =
3.86
   Model |  73.1000615    12  6.09167179              Prob > F      =
0.0000
Residual |   758.39583   481  1.57670651              R-squared     =
0.0879
---------+------------------------------              Adj R-squared =
0.0652
   Total |  831.495892   493  1.68660424              Root MSE      =
1.2557

------------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
  agegp1 |   .4666611    .262477      1.778   0.076      -.0490822
.9824043
agegp1sq |  -.0092326    .005614     -1.645   0.101      -.0202635
.0017984
    fem1 |   .2320113   .1191534      1.947   0.052      -.0021143
.4661368
    fem2 |  (dropped)
     mg1 |  -.2828236    .376439     -0.751   0.453      -1.022492
.4568444
     mg2 |  (dropped)
     mg3 |  -.5308499   .3704168     -1.433   0.152      -1.258685
.1969851
   lang1 |   .2294103   .3574846      0.642   0.521       -.473014
.9318346
   lang2 |   .1814799    .384997      0.471   0.638      -.5750038
.9379637
   lang3 |   .7792978   .5933538      1.313   0.190      -.3865879
1.945183
   lang4 |  (dropped)
    res1 |  -.7069455   .1632037     -4.332   0.000      -1.027626   -
.3862653
    res2 |  -.4539421   .1454726     -3.120   0.002      -.7397824   -
.1681018
    res3 |   .0041952   .2028035      0.021   0.984      -.3942951
.4026856
    res5 |   .1609639   .7436285      0.216   0.829      -1.300198
1.622126
   _cons |   1.358006   3.070716      0.442   0.659      -4.675669
7.391681
------------------------------------------------------------------------
-------
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang3 lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------              F( 12,   481) =
3.86
   Model |  73.1000615    12  6.09167179              Prob > F      =
0.0000
```

```
Residual |   758.39583   481  1.57670651               R-squared     =
0.0879
---------+------------------------------               Adj R-squared =
0.0652
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2557

-----------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
-------
  agegp1 |   .4666611    .262477      1.778   0.076      -.0490822
.9824043
agegp1sq |  -.0092326    .005614     -1.645   0.101      -.0202635
.0017984
    fem1 |   .2320113   .1191534      1.947   0.052      -.0021143
.4661368
    fem2 |  (dropped)
     mg1 |  -.2828236    .376439     -0.751   0.453      -1.022492
.4568444
     mg2 |  (dropped)
     mg3 |  -.5308499   .3704168     -1.433   0.152      -1.258685
.1969851
   lang1 |   .0479303   .1638462      0.293   0.770      -.2740123
.369873
   lang3 |   .5978178   .5000657      1.195   0.232      -.3847653
1.580401
   lang4 |  -.1814799    .384997     -0.471   0.638      -.9379637
.5750038
    res1 |  -.7069455   .1632037     -4.332   0.000      -1.027626   -
.3862653
    res2 |  -.4539421   .1454726     -3.120   0.002      -.7397824   -
.1681018
    res3 |   .0041952   .2028035      0.021   0.984      -.3942951
.4026856
    res5 |   .1609639   .7436285      0.216   0.829      -1.300198
1.622126
   _cons |   1.539486   3.063196      0.503   0.615      -4.479412
7.558384
-----------------------------------------------------------------------
-------
-> . browse
-> . regress c1avgcig lang1  lang2 lang3 lang4 res1 res2 res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  7,   486) =
5.24
   Model |  26225465.9     7  3746495.12               Prob > F      =
0.0000
Residual |   347245595   486   714497.11               R-squared     =
0.0702
---------+------------------------------               Adj R-squared =
0.0568
```

```
   Total |   373471061   493  757547.792               Root MSE      =
845.28

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
   lang1 |    394.476   239.5065      1.647   0.100      -76.12003
865.072
   lang2 |   300.8077   257.9815      1.166   0.244      -206.0892
807.7046
   lang3 |   617.3326   398.2758      1.550   0.122      -165.2225
1399.888
   lang4 |  (dropped)
    res1 |   -429.541   139.2387     -3.085   0.002      -703.1251   -
155.9569
    res2 |  -287.5235   131.6011     -2.185   0.029      -546.1009   -
28.94615
    res3 |  (dropped)
    res4 |   109.7906   135.7797      0.809   0.419      -156.9972
376.5784
    res5 |   204.8489    501.838      0.408   0.683      -781.1911
1190.889
   _cons |   1062.675   261.4907      4.064   0.000       548.8832
1576.467
------------------------------------------------------------------------
-------
-> . regress c1avgcig res1 res2 res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  4,   489) =
8.19
   Model |  23456792.9     4  5864198.23               Prob > F      =
0.0000
Residual |   350014268   489   715775.60               R-squared     =
0.0628
---------+------------------------------               Adj R-squared =
0.0551
   Total |   373471061   493  757547.792               Root MSE      =
846.04

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
    res1 |   -432.062    138.458     -3.121   0.002      -704.1079   -
160.0161
    res2 |  -277.6685   131.4394     -2.113   0.035      -535.9242   -
19.41286
    res3 |  (dropped)
    res4 |   112.9493   135.8009      0.832   0.406       -153.876
379.7745
```

```
    res5 |      229.5   501.8436      0.457   0.648      -756.5358
1215.536
   _cons |     1432.5   115.1308     12.442   0.000       1206.288
1658.712
------------------------------------------------------------------------
-------
-> . browse
-> . regress   c1avgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang3 lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
   Total |   373471061   493  757547.792               Root MSE      =
829.25

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
    fem2 |  (dropped)
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |    79.1082   108.2045      0.731   0.465      -133.5036
291.72
   lang3 |   391.9469   330.2448      1.187   0.236      -256.9537
1040.847
   lang4 |  -277.8875   254.2531     -1.093   0.275      -777.4715
221.6965
    res1 |  -512.9595   107.7802     -4.759   0.000      -724.7376   -
301.1813
    res2 |  -350.3434    96.0705     -3.647   0.000      -539.1131   -
161.5737
    res3 |  -59.06489    133.932     -0.441   0.659       -322.229
204.0992
    res5 |  -57.95983   491.0943     -0.118   0.906      -1022.915
906.9954
   _cons |  -2449.621   2022.943     -1.211   0.227      -6424.518
1525.276
```

```
------------------------------------------------------------------------------
-------
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 lang1 lang2 res1 res2
res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  9,   484) =
4.46
   Model |  63.7122039     9  7.07913377               Prob > F      =
0.0000
Residual |  767.783688   484  1.58632993               R-squared     =
0.0766
---------+------------------------------               Adj R-squared =
0.0595
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2595

------------------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------------
-------
  agegp1 |   .4624112   .2631787      1.757   0.080      -.0547027
.9795252
agegp1sq |  -.0087817   .0056269     -1.561   0.119      -.0198379
.0022745
    fem1 |   .1865439   .1180242      1.581   0.115      -.0453593
.418447
    fem2 |  (dropped)
   lang1 |  -.0483528    .290405     -0.167   0.868      -.6189631
.5222574
   lang2 |  -.1186269   .3227804     -0.368   0.713      -.7528508
.5155969
    res1 |  -.7073429    .211706     -3.341   0.001      -1.123319   -
.2913666
    res2 |   -.448087   .1984448     -2.258   0.024      -.8380067   -
.0581673
    res3 |  (dropped)
    res4 |   .0127567    .203233      0.063   0.950      -.3865713
.4120847
    res5 |   .2830401   .7496289      0.378   0.706      -1.189889
1.755969
   _cons |   1.054056   3.054983      0.345   0.730       -4.94861
7.056722
------------------------------------------------------------------------------
-------
-> . save "A:\sssaws.dta"
file A:\sssaws.dta saved
-> . regress  c1avgcig  agegp1 agegp1sq fem1 fem2 lang1 lang2 res1 res2
res3 res4 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F(  9,   484) =
5.86
```

```
   Model |  36669586.4     9  4074398.49               Prob > F      =
0.0000
Residual |   336801475   484  695870.816               R-squared     =
0.0982
---------+------------------------------               Adj R-squared =
0.0814
   Total |   373471061   493  757547.792               Root MSE      =
834.19

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
  agegp1 |   329.3406   174.3084      1.889   0.059      -13.15403
671.8353
agegp1sq |  -6.132983   3.726819     -1.646   0.100      -13.45573
1.18976
    fem1 |   166.8859   78.16976      2.135   0.033       13.29193
320.4799
    fem2 |  (dropped)
   lang1 |   118.7268   192.3409      0.617   0.537      -259.1994
496.6531
   lang2 |   21.98779   213.7837      0.103   0.918      -398.0711
442.0466
    res1 |  -451.5866    140.217     -3.221   0.001      -727.0959   -
176.0773
    res2 |   -282.746   131.4339     -2.151   0.032      -540.9974   -
24.49459
    res3 |  (dropped)
    res4 |   73.01398   134.6052      0.542   0.588      -191.4687
337.4967
    res5 |   101.7522   496.4938      0.205   0.838      -873.7972
1077.302
   _cons |  -3004.129   2023.375     -1.485   0.138      -6979.813
971.5536
------------------------------------------------------------------------
-------
-> . regress c1avgcig agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3 lang1 lang3
lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
   Total |   373471061   493  757547.792               Root MSE      =
829.25

------------------------------------------------------------------------
-------
```

```
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
    fem2 |  (dropped)
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |    79.1082   108.2045      0.731   0.465      -133.5036
291.72
   lang3 |   391.9469   330.2448      1.187   0.236      -256.9537
1040.847
   lang4 |  -277.8875   254.2531     -1.093   0.275      -777.4715
221.6965
    res1 |  -512.9595   107.7802     -4.759   0.000      -724.7376   -
301.1813
    res2 |  -350.3434    96.0705     -3.647   0.000      -539.1131   -
161.5737
    res3 |  -59.06489    133.932     -0.441   0.659       -322.229
204.0992
    res5 |  -57.95983   491.0943     -0.118   0.906      -1022.915
906.9954
   _cons |  -2449.621   2022.943     -1.211   0.227      -6424.518
1525.276
----------------------------------------------------------------------
-------
-> . test
last test not found
r(302);
-> . test  res1 res2 res3 res4 res5
res4 not found
r(111);
-> . browse
-> . test  res1 res2 res4 res5
res4 not found
r(111);
-> . test   res4 res5
res4 not found
r(111);
-> . gen fres1 =  fem1* res1
-> . gen fres2 =  fem1* res2
-> . gen fres3 =  fem1* res3
-> . gen fres4 =  fem1* res4
-> . gen lres4 =  fem1* res4
-> . gen lres3 =  fem1* res3
-> . gen lres2 =  fem1* res2
-> . xi: regress c1avgcig  agegp1 agegp1sq fem1*i.c2q22  mst
fem1*i.c2q22
i:  operator invalid
```

```
r(198);
-> . xi: regress c1avgcig  agegp1 agegp1sq fem1*i.c2q22  mst
fem1*i.c2q22
i:  operator invalid
r(198);
-> . regress  lnavgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res4 res5  fres1 fres2 fres3 fres4

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 15,   478) =
3.16
   Model |  74.9540559    15  4.99693706               Prob > F      =
0.0001
Residual |  756.541836   478  1.58272351               R-squared     =
0.0901
---------+------------------------------               Adj R-squared =
0.0616
   Total |  831.495892   493  1.68660424               Root MSE      =
1.2581

------------------------------------------------------------------------------
-------
lnavgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------------
-------
  agegp1 |   .4653588   .2634117      1.767   0.078      -.0522291
.9829468
agegp1sq |  -.0091926   .0056353     -1.631   0.103      -.0202655
.0018804
    fem1 |   .2763878   .2388895      1.157   0.248      -.1930156
.7457911
    fem2 |  (dropped)
     mg1 |  -.2535987   .3786869     -0.670   0.503      -.9976954
.490498
     mg2 |  (dropped)
     mg3 |  -.5009791   .3728798     -1.344   0.180      -1.233665
.2317069
   lang1 |   .2261884   .3590023      0.630   0.529      -.4792292
.9316061
   lang2 |    .184086   .3863492      0.476   0.634      -.5750667
.9432388
   lang3 |   .7764261   .5948289      1.305   0.192      -.3923764
1.945229
   lang4 |  (dropped)
    res1 |  -.6205406   .3253636     -1.907   0.057       -1.25986
.0187791
    res2 |  -.3821635   .3159927     -1.209   0.227       -1.00307
.2387429
    res3 |  (dropped)
    res4 |   .2405764   .3453432      0.697   0.486      -.4380019
.9191548
    res5 |   .2279672   .8295743      0.275   0.784      -1.402096
1.85803
   fres1 |  (dropped)
```

```
   fres2 |   .0214531   .3049182      0.070   0.944      -.5776927
.6205989
   fres3 |   .1184468   .4309045      0.275   0.784      -.7282543
.9651478
   fres4 |  -.2509959   .3291604     -0.763   0.446      -.8977761
.3957842
   _cons |   1.238128   3.085392      0.401   0.688       -4.82448
7.300735
------------------------------------------------------------------------
-------
-> . xi: regress c1avgcig i.c2q22*i.fem1 i.c4q28*i.c2q28
i.c2q22               Ic2q22_1-7   (naturally coded; Ic2q22_1 omitted)
i.fem1                Ifem1_0-1    (naturally coded; Ifem1_0 omitted)
i.c2q22*i.fem1        IcXf_#-#     (coded as above)
i.c4q28               Ic4q28_1-4   (naturally coded; Ic4q28_1 omitted)
c2q28 not found
r(111);
-> . xi: regress c1avgcig i.c2q22*i.fem1 i.c4q28*i.c2q22
i.c2q22               Ic2q22_1-7   (naturally coded; Ic2q22_1 omitted)
i.fem1                Ifem1_0-1    (naturally coded; Ifem1_0 omitted)
i.c2q22*i.fem1        IcXf_#-#     (coded as above)
i.c4q28               Ic4q28_1-4   (naturally coded; Ic4q28_1 omitted)
i.c4q28*i.c2q22       IcXc_#-#     (coded as above)

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 19,   474) =
2.71
   Model |  36627408.0    19  1927758.31               Prob > F      =
0.0001
Residual |   336843653   474  710640.619               R-squared     =
0.0981
---------+------------------------------               Adj R-squared =
0.0619
   Total |   373471061   493  757547.792               Root MSE      =
843.00

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
Ic2q22_2 |   150.8893   136.0257      1.109   0.268      -116.3985
418.1772
Ic2q22_3 |    364.733   226.7216      1.609   0.108      -80.77062
810.2367
Ic2q22_4 |   565.0646   164.7734      3.429   0.001        241.288
888.8412
Ic2q22_7 |   422.8657   504.0264      0.839   0.402      -567.5367
1413.268
 Ifem1_1 |   286.3597   159.7771      1.792   0.074      -27.59935
600.3187
IcXf_2_1 |  -161.5793    204.386     -0.791   0.430      -563.1939
240.0353
IcXf_3_1 |   82.64645   289.7133      0.285   0.776      -486.6348
651.9277
```

```
IcXf_4_1 |   -300.659   221.8701     -1.355   0.176      -736.6295
135.3115
IcXf_7_1 |  (dropped)
Ic4q28_2 |  -347.6595   309.1965     -1.124   0.261      -955.2248
259.9058
Ic4q28_3 |   -23.7746   848.9713     -0.028   0.978      -1691.987
1644.438
Ic4q28_4 |  -616.2658     330.77     -1.863   0.063      -1266.223
33.69108
Ic2q22_2 |  (dropped)
Ic2q22_3 |  (dropped)
Ic2q22_4 |  (dropped)
Ic2q22_7 |  (dropped)
IcXc_2_2 |   338.2834    359.118      0.942   0.347      -367.3767
1043.943
IcXc_2_3 |  -254.2518   428.1891     -0.594   0.553      -1095.635
587.1318
IcXc_2_4 |   467.9627    357.429      1.309   0.191      -234.3787
1170.304
IcXc_2_7 |  (dropped)
IcXc_3_2 |  -11.08443   950.8081     -0.012   0.991      -1879.405
1857.236
IcXc_3_3 |  (dropped)
IcXc_3_4 |   1048.585   1040.858      1.007   0.314      -996.6822
3093.852
IcXc_3_7 |  (dropped)
IcXc_4_2 |   449.3418   592.8248      0.758   0.449      -715.5479
1614.231
IcXc_4_3 |  -70.24793   918.5222     -0.076   0.939      -1875.127
1734.631
IcXc_4_4 |    477.226   688.5976      0.693   0.489      -875.8554
1830.307
IcXc_4_7 |  (dropped)
   _cons |   952.7746   100.5566      9.475   0.000       755.1827
1150.366
----------------------------------------------------------------------
-------
-> . xi: regress c1avgcig  agegp1 agegp1sq fem1 i.mst i.c2q22 i.c4q28
fem1*i.c2q22
i.mst                Imst_1-4     (naturally coded; Imst_1 omitted)
i.c2q22              Ic2q22_1-7   (naturally coded; Ic2q22_1 omitted)
i.c4q28              Ic4q28_1-4   (naturally coded; Ic4q28_1 omitted)
i:  operator invalid
r(198);
-> . regress c1avgcig  agegp1 agegp1sq fem1 mg1 mg2 mg3  lang1  lang2
lang3 lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
```

```
   Total |   373471061   493  757547.792              Root MSE      =
829.25

------------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |   356.9957   236.0838      1.512   0.131      -106.8872
820.8787
   lang2 |   277.8875   254.2531      1.093   0.275      -221.6965
777.4715
   lang3 |   669.8344   391.8525      1.709   0.088      -100.1197
1439.789
   lang4 |  (dropped)
    res1 |  -512.9595   107.7802     -4.759   0.000      -724.7376   -
301.1813
    res2 |  -350.3434    96.0705     -3.647   0.000      -539.1131   -
161.5737
    res3 |  -59.06489    133.932     -0.441   0.659       -322.229
204.0992
    res5 |  -57.95983   491.0943     -0.118   0.906      -1022.915
906.9954
   _cons |  -2727.509   2027.909     -1.345   0.179      -6712.164
1257.147
------------------------------------------------------------------------
-------
-> . gen fres5 =  fem1* res5
-> . regress c1avgcig  agegp1 agegp1sq fem1 mg1 mg2 mg3  lang1  lang2
lang3 lang4 res1 res2 res3 res5  fres1 fres2 fres3 fres5

  Source |       SS       df       MS                   Number of obs =
494
---------+-----------------------------               F( 15,   478) =
4.34
   Model |  44745670.8    15  2983044.72              Prob > F      =
0.0000
Residual |   328725390   478  687710.022              R-squared     =
0.1198
---------+-----------------------------               Adj R-squared =
0.0922
   Total |   373471061   493  757547.792              Root MSE      =
829.28
```

c1avgcig	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegp1	322.0075	173.6342	1.855	0.064	-19.17312	663.1881
agegp1sq	-6.235821	3.714622	-1.679	0.094	-13.53483	1.063186
fem1	48.12873	151.9789	0.317	0.752	-250.5006	346.758
mg1	-214.3176	249.6206	-0.859	0.391	-704.8069	276.1716
mg2	(dropped)					
mg3	-407.0655	245.7927	-1.656	0.098	-890.0332	75.90223
lang1	371.6738	236.645	1.571	0.117	-93.3192	836.6669
lang2	295.2998	254.6714	1.160	0.247	-205.1141	795.7136
lang3	678.5087	392.0958	1.730	0.084	-91.93569	1448.953
lang4	(dropped)					
res1	-666.6986	156.8281	-4.251	0.000	-974.8563	-358.5408
res2	-433.5549	150.0635	-2.889	0.004	-728.4206	-138.6893
res3	-317.0472	227.6413	-1.393	0.164	-764.3485	130.2541
res5	-4.585765	493.4933	-0.009	0.993	-974.27	965.0985
fres1	290.6287	216.974	1.339	0.181	-135.712	716.9694
fres2	107.6269	196.6962	0.547	0.585	-278.8691	494.1229
fres3	393.3329	281.3855	1.398	0.163	-159.5726	946.2383
fres5	(dropped)					
_cons	-2532.388	2031.341	-1.247	0.213	-6523.849	1459.073

```
-> . test res1 res2 res3 res5

 ( 1)  res1 = 0.0
 ( 2)  res2 = 0.0
 ( 3)  res3 = 0.0
 ( 4)  res5 = 0.0

       F(  4,   478) =    4.59
            Prob > F =    0.0012

-> . test fres1 fres2 fres3 fres5

 ( 1)  fres1 = 0.0
 ( 2)  fres2 = 0.0
```

```
 ( 3)  fres3 = 0.0
 ( 4)  fres5 = 0.0
       Constraint 4 dropped

       F(  3,   478) =    0.99
            Prob > F =    0.3989

-> . regress c1avgcig  agegp1 agegp1sq fem1 fem2 mg1 mg2 mg3  lang1
lang2 lang3 lang4 res1 res2 res3 res5

  Source |       SS       df       MS                  Number of obs =
494
---------+------------------------------               F( 12,   481) =
5.18
   Model |  42710512.0    12  3559209.33               Prob > F      =
0.0000
Residual |   330760549   481   687651.87               R-squared     =
0.1144
---------+------------------------------               Adj R-squared =
0.0923
   Total |   373471061   493  757547.792               Root MSE      =
829.25

----------------------------------------------------------------------
-------
c1avgcig |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+------------------------------------------------------------
-------
  agegp1 |   332.9815   173.3406      1.921   0.055      -7.616724
673.5798
agegp1sq |  -6.490118   3.707489     -1.751   0.081      -13.77499
.7947563
    fem1 |   203.2189   78.68926      2.583   0.010       48.60175
357.8361
    fem2 |  (dropped)
     mg1 |  -225.8352   248.6014     -0.908   0.364       -714.314
262.6436
     mg2 |  (dropped)
     mg3 |  -417.2045   244.6243     -1.705   0.089      -897.8688
63.45969
   lang1 |   356.9957   236.0838      1.512   0.131      -106.8872
820.8787
   lang2 |   277.8875   254.2531      1.093   0.275      -221.6965
777.4715
   lang3 |   669.8344   391.8525      1.709   0.088      -100.1197
1439.789
   lang4 |  (dropped)
    res1 |  -512.9595   107.7802     -4.759   0.000      -724.7376   -
301.1813
    res2 |  -350.3434    96.0705     -3.647   0.000      -539.1131   -
161.5737
    res3 |  -59.06489    133.932     -0.441   0.659       -322.229
204.0992
    res5 |  -57.95983   491.0943     -0.118   0.906      -1022.915
906.9954
```

```
   _cons |  -2727.509   2027.909     -1.345   0.179      -6712.164
1257.147
-----------------------------------------------------------------------
-------
-> . test  res1 res2 res3 res5

 ( 1)  res1 = 0.0
 ( 2)  res2 = 0.0
 ( 3)  res3 = 0.0
 ( 4)  res5 = 0.0

       F(  4,   481) =    6.86
            Prob > F =    0.0000

-> . test  res1 res2

 ( 1)  res1 = 0.0
 ( 2)  res2 = 0.0

       F(  2,   481) =   12.22
            Prob > F =    0.0000

-> . BREAK
sending Break to calling program...
Session ended at 25 Mar 2001; 16:51:39
***********************************************************************
********
```

Economics 452 Assignment 3

Why Bother: an Analysis of Volunteer Work in Canada, 1987

2001-03-24

Introduction

Volunteer work is an integral but understudied component of Canadian economic activity. In 1987 alone, whether individuals reported participating in non-standard labour such as fundraising for charity or coaching a youth sports team, 5 337 000 Canadians contributed over 1 017 548 000 hours of unpaid work to at least one volunteer organization. Assuming a thirty-five hour work week and fifty-two weeks per year, these hours amounted to approximately 522 000 full-time workers, an informal labour force that employed more individuals across Canada than agriculture. With this in mind, there can be little doubt as to the economic and social importance of volunteer work within Canadian society. Nonetheless, there does not appear to be a consensus as to what motivates individuals to dedicate their free time to volunteering, an allotment time that could otherwise be used for leisure, and also which people are more likely to do so.

These notions are evaluated and tested in To Volunteer or not: Canada, 1987, an article published in the Canadian Journal of Economics by Francois Vaillancourt. The purpose of his article was to "examine how the benefits and costs of doing volunteer work, as they are represented by individual characteristics such as age, education, or occupation, explain the choice to carry out or to not carry out volunteer work." Within the context of the Time Allocation/ Household Production Model, Vaillancourt hypothesised the effect of fifteen different explanatory variables on doing or nor doing volunteer work, and subsequently tested his predictions using probit analysis. Our paper will consist of three sections. First, we will briefly review the theories, methodologies, and results recorded in the latter article. Second, using similar data and an OLS regression model, we will attempt to replicate Vaillancourt's analysis. Third, we shall report our results and offer some concluding remarks.

I. Literature and Economic Theory

François Vaillancourt's "To Volunteer or not: Canada, 1987" examines two dimensions of economic theory, the Time Allocation/ Household Production Model and the Human Capital Model. His goal is to link these theories with 1987 National Labour Force Survey data in order to explain the benefits, costs, and choices that will make it likely that an individual will engage in volunteer work.

For the purpose of his analysis, Vaillancourt uses the broad definition of volunteer work provided by Statistics Canada. This definition includes all unpaid labour that is likely to replicate organized market activity and be explained by economic variables. Work such as involvement in charitable organizations, clubs, pressure groups, sports teams, unions, and other volunteer associations would fall under this classification. Volunteer work does not, however, include informal market activities such as minding children and aiding in domestic chores. With this definition in mind, Vaillancourt looks to economic theory to explain the reasons why individuals will chose to volunteer.

In order to determine the benefits of volunteer work, Vaillancourt looks to the Time Allocation/ Household Production Model and Jacob Mincer's Human Capital Model. The former deems that time not spent doing wage-earning labour can be allocated to either leisure or performing non-market work- volunteering is one such activity. As such, the returns from volunteer work have to equal those of other preferred activities. The Human Capital Model postulates that many individuals will be motivated only by their enlightened self-interest. More specifically, many will volunteer if they can reap direct benefits from their efforts, including improving themselves through the development of skills, increased networking, or/ and a better public image. In both models, an individual will volunteer until the benefits of his labour equals its opportunity cost. Unfortunately, although these models are informative as to the motivations behind why individuals commit or refuse to dedicate their time to volunteer work, the costs and benefits described are vague, subjective to individual, group, or regional preferences, and are extremely difficult to test empirically, especially within the bounds of any available census data. As a consequence, Vaillancourt decides to look at some prominent characteristics of those individuals who do and do not volunteer, such as their age, income, work status, and employment, that will be reliably contained in the survey data and can be used to explain the choices people have made.

In his analysis, Vaillancourt examines the multivariate impact of sixteen explanatory variables on doing volunteer work. The cross sectional data used in his paper was collected by Statistics Canada through the 1987 Labour Force Survey and by a follow-up questionnaire for the participants who responded positively to the volunteer-specific questions. Volunteer work served as the dependant variable in the analysis, taking on a value of one for those doing volunteer work and zero otherwise. The explanatory variables taken from the survey are as follows:

1) Sex
2) Age
3) Education

4) Marital Status
5) Occupation
6) Work Status
7) Total Income
8) Number of children (0-2) (3-5) (6-15)
9) Occupation
10) Work Status
11) Hours of Work
12) Size of city residence
13) Usual language
14) Religion
15) Region of residence
16) Number of hours worked in a week

Using a probit technique, Vaillancourt then estimated the relationship between these variables and volunteer work. The results of his analysis are mixed, and can be observed in Appendix A. They generally show, however, that an individual's choice to do volunteer work can be meaningfully explained by economic analysis. More precisely, the results indicate that these choices will differ between men and women, reflecting the existing social arrangements that exist between the sexes, and that both consumption (household production) and investment (human capital) play a role in the decisions that individuals make.

II. Data and Analysis

The data used for both the Vaillancourt article and this paper comes from the same Statistics Canada survey from October, 1987. The sample size of the data set is 26,757, with well over 100 variables. In order to emulate the regression carried out in the article, there were a number issues that were encountered in the data.

Firstly, the fact that the survey itself was a two-part process requires recognition. The first question asked of participants was: "There are many ways in which people may give their time and skills to various groups and organizations. It is hard to remember all the things one could have done during a year, so let me ask you specifically…". Upon a positive answer regarding volunteer work, a follow up questionnaire was then administered. The resulting effect on the data is that there are a large number of missing values. The variable that was selected for this paper was the response to the question "Did you volunteer at any time in the past year?" However, this question was administered after the initial screening of the Labour Force survey, and hence creates the large number of missing values.

The second issue encountered is with selection of the availability of choices upon which to create dependent variable for our model. The author's choice of the simple question "Did you volunteer in the last year?" was a poor and arbitrary choice, and does not lend itself to much interpretation by way of the amount of volunteer work that occurs in Canada, and the characteristics that affect the level of contribution. That being said, it was the best available source for the purposes of this paper, and it was retained.

The third issue encountered was in the data itself. In selecting variables of the model, we wanted to accurately follow what Vaillancourt had set out for his regression. However, the "Occupation" variable was a large number of contradictions in it, especially when compared with the "Work Status" variable. When diagnostic tests were run on the individual variables, they proved to be significant. However the data itself claims that although an individual may be unemployed, it is still a full-time job. There are 110 Canadians who are unemployed, yet working full-time, and 16 more unemployed, while engaging in part-time work. Additionally, the “Unemployed” category was dropped from the article’s regression, while also dropping the “Production Workers” category. Due to this confusion, Occupation was not included in one of the regressions we ran.

There were some definite problems that arose when handing the data for provinces. Vaillancourt had amalgamated several provinces into regions (Atlantic, Prairies) for the purposes of commonality, yet retained British Columbia and Quebec. This distorted the data, and perhaps might have been more effective if it had remained in province form. However, we felt that these distortions were minimal, and retained the regions.

The final issue encountered was the “How many children do you have?” variables. The data has three separate questions that were asked in order to assess the number of children that the sample individual has, and the number of children in each age category. This was incredibly distortionary, and even the author has trouble printing it in his table, seen in Appendix 1.

There was a great deal of recoding and generation of variables that was needed in order to arrive at suitable model to regress. There were three steps that we took in creating our “do” file that enabled us to obtain the most desirable and accurate model. Firstly, we recoded a large number of variables for the purposes of both identification and regression. Next, we created dummies variables for most of the existing variables in the data. This is because most of the variables in the data set that we selected were categorical, and as such did not have a meaningful value for Stata without any alteration. For example, the "Age" variable was divided up into categories for various age groups. A value for the category of 15-16 years is assigned a value of say, 1, and 16-20 is 2. They need to be divided up in order to be regressed properly. Finally, the model was constructed using a interaction terms, so as to analyze the effect of certain variables when combined with each other, upon the dependent variable. In the case of marital status, there were only two values, and a missing value, so we made an assumption that a missing variable denoted "not married".

The dummy variables were created on the following variables: age, sex, education, marital status, work status, total income, children, size of city of residence, language, religion, and region.

The regression model is constructed as follows:

$$\mathbf{vol} = \beta_1 + \beta_2\mathbf{sex} + \beta_3\mathbf{agegroup} + \beta_4\mathbf{educ} + \beta_5\mathbf{marital} + \beta_6\mathbf{occ} + \beta_7\mathbf{workf} + \beta_8\mathbf{income} + \beta_9\mathbf{ownkids1} + \beta_{10}\mathbf{ownkids2} + \beta_{11}\mathbf{ownkids3} + \beta_{12}\mathbf{citysize} + \beta_1\mathbf{eng} + \beta_{14}\mathbf{fren} + \beta_{15}\mathbf{other} + \beta_{16}\mathbf{religion} + \beta_{17}\mathbf{region} + \beta_{18}\mathbf{workhrs} + \beta_{19}\mathbf{kids} + \varepsilon$$

Vol is the dependent variable, ε is an error term, and the remaining variables are explanatory variables.

Analysis of this data set is somewhat unconventional given the author's choice of using an indicator variable as the dependent variable. This is a particularly interesting choice given that, at minimum in our similar data set, there is also an average number of hours volunteered variable that could have been chosen. The consequence of this choice is that our regressor coefficients represent the percentage change in likelihood of volunteering, ergo 0.06 would indicate a 6% greater likelihood for some group possessing that attribute to volunteer. However, it would **not** be indicative of any underlying population value equal to that, as any member of population either did (1) or did not (0) volunteer.

As mentioned above, the final regression chosen is quite similar to that modelled by Villancourt (Appendix A) with two notable exceptions:

1. Addition of the variable KIDS, a dummy variable constructed to indicate whether the respondent had any children
2. Omission of the variables expressing occupation sector (e.g. health and education)

These additions were made because of examinations of the data. KIDS was added because the data on children had been disaggregated into particular groups, and an understanding of the overall effect was desired. The omission of variables expressing occupation and/or occupation sector were because a cross-tabulation of workforce participation (not working, part-time, full-time) with occupation, which included a category for not working, was contradictory, as seen in figure 1.1

Figure 1.1: Occupation Statistics

	Type of Job			
Occupation Category	**Not Working**	**Full Time**	**Part Time**	**Total**
1	0	2889	202	3091
2	0	1719	700	2419
3	0	2380	906	3286
4	0	3547	1842	5389
5	0	5832	735	6567
6	5880	110	15	6005
Total	**5880**	**16477**	**4400**	**26757**

Note, occupation Category 6 is not working. Notice then, that there people who said that they were full-time non-working (110) and part-time non-working (15). Given no criteria to judge which was more accurate, we felt that examining the effect of more non-work time (i.e. as expressed through the Type of Job) would be a more interesting examination than occupational categories which are somewhat arbitrary anyway (e.g. Medical doctors with primary school teachers and university deans).

The final results of the regression are shown in Figure 1.2, and demonstrate that at the 5% significance level we retain the impacts of sex, all of our age-group effects, all of our educational effects, being single, type of job (as described above), those at the highest end of the income bracket ($40,000+), those who have many older children (see ownkids5b in description), being in non-metro cities, having a home language other than English or French, and being in an undefined religion.

Table 1.2: Impact Of individual Characteristics On Doing Volunteer Work

Canada, 1987, All Canadians

Variable	All Canadians	P-Values
Volunteer Work		
Male	0.0405 (0.0083)	0.00
Age (15-16 years omitted)		
17-19	-0.0583 (0.0247)	0.0190
20-24	-0.0593 (0.0239)	0.0130
25-34	-0.0679 (0.0237)	0.0040
35-44	-0.1293 (0.0243)	0.0000

45-54	-0.1755 (0.0261)	0.0000
55-64	-0.1743 (0.0265)	0.0000
65-69	-0.2329 (0.0295)	0.0000
70 and over	-0.2809 (0.0280)	0.0000
Education (none or elementary omitted)		
Secondary	-0.0836 (0.0106)	0.0000
Post-Secondary, no degree	-0.1644 (0.0172)	0.0000
Post Secondary, with degree	-0.1727 (0.0153)	0.0000
University	-0.2168 (0.0186)	0.0000
Marital Status (Married omitted)		
Single	0.0278 (0.0125)	0.0270
Separated, divorced, widowed	0.0127 (0.0128)	0.3220
Work Status (full time omitted)		
Part time	-0.0315 (0.0133)	0.0018
Does not work	-0.0522 (0.0147)	0.0000
Total Income $ (less than 5,000 omitted)		
5, 000-9,999	0.0255 (0.0309)	0.4110
10,000-14.999	-0.0104 (0.0304)	0.7340
15,000-19,999	-0.0232 (0.0307)	0.4410
20,000-29,999	-0.0246 (0.0302)	0.4150
30,000-39,000	-0.0406 (0.0303)	0.1810

40,000-59,999	-0.0665 (0.0308)	0.0310
60,000 and over	-0.1011 (0.0332)	0.0020
Children Aged 0-2 years+ (ownkids0 omitted)		
1	0.0197 (0.0198)	0.3190
2	0.0570 (0.0353)	0.1060
3	0.1445 (0.1838)	0.4320
Children Aged 3-5 years+ (ownkidsa0 omitted)		
1	0.0246 (0.0161)	0.4040
2	0.0322 (0.0385)	0.0510
3	-0.7971 (0.4083)	0.1480
Children Aged 6-15 years+ (ownkidsb0 omitted)		
1	0.0302 (0.0209)	0.4560
2	0.001 (0.0236)	0.4510
3	0.0245 (0.0329)	0.0800
4	0.0436 (0.0329)	0.1897
5	0.2279 (0.0303)	0.2540
Size of city of residence (500,000 and over omitted)		
100,000 - 499,999	-0.0075 (0.0135)	0.5790
30,000-99,999	-0.0072 (0.0132)	0.5830
Less than 30,000	-0.0314 (0.0119)	0.0080

Rural	-0.0095 (0.0119)	0.4220
Usual Language		
English	0.0375 (0.0325)	0.2560
French	0.0318 (0.0350)	0.3640
Others	-0.0893 (0.0311)	0.0040
Religion (Protestant omitted)		.
None	-0.0198 (0.0138)	0.1540
Catholic	-0.0439 (0.0132)	0.0010
Others	-0.0308 (0.0168)	0.0680
Region (Ontario omitted)		
Atlantic	0.0539 (0.0182)	0.0030
Quebec	0.0204 (0.0124)	0.1010
Prairies	-0.0144 (0.0115)	0.2120
British Columbia	0.0238 (0.0151)	0.1140
Hours Worked	-0.0001 (0.0003)	0.6700
Kids	2.11405 (1.1584)	0.0000
Constant	-0.0223 (0.0222)	0.0000
R-Sqaured	0.0681	
Sample Size	26757	

Note: Standard Error in ( )

Testing more generally shows what categories we can dismiss as not being significant influencers of likelihood to volunteer. This is performed through F-tests as shown below in Figure 1.3.

F-Tests	F Value	Prob > F	Variables
Age Groups	20.46	0.0000	Iagegr_2 Iagegr_3 Iagegr_4 Iagegr_5 Iagegr_6 Iagegr_7 Iagegr_8 Iagegr_9
Education Levels	53.45	0.0000	Ieduc_2 Ieduc_3 Ieduc_4 Ieduc_5
Marital Status	2.66	0.0700	Imarit_2 Imarit_3
Type of Job	6.34	0.0018	Iworkf_1 Iworkf_2
Income Bracket	7.06	0.0000	Iincom_2 Iincom_3 Iincom_4 Iincom_5 Iincom_6 Iincom_7 Iincom_8
Having Young Children	1.09	0.3537	Iownki_1 Iownki_2 Iownki_3
Having Middle Children	2.18	0.0878	Iownkia1 Iownkia2 Iownkia3
Having Older Children	1.49	0.1897	Iownkib1 Iownkib2 Iownkib3 Iownkib4 Iownkib5
City Size	2.06	0.0831	Icitys_2 Icitys_3 Icitys_4 Icitys_5
Speaking an official language	0.68	0.5048	eng fren
Language	15.88	0.0000	eng fren other
Religious beliefs	4.27	0.0051	Irelig_2 Irelig_3 Irelig_4
Geographic Region	5.29	0.0003	Iregio_2 Iregio_3 Iregio_4 Iregio_5

Figure 1.3: F-Tests

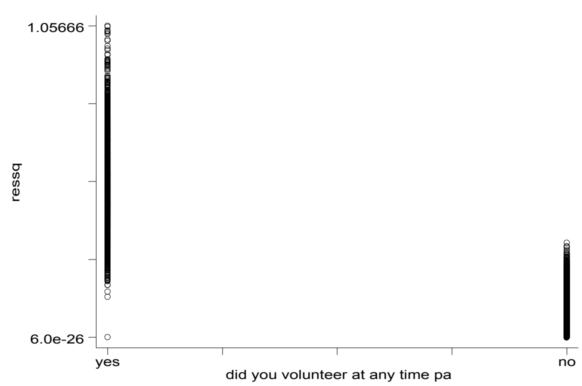
Notice that we are unable to reject the null about the effect of Marital Status, having children (any age), city size, or speaking an official language at the 5% significance level.

We also tested our model for heteroskedasticity and specification error. Note that our tests for specification error were above simply the F-test above, but rather looked at whether we introduced specification error by dropping the occupation or occupation category variables. The specification error of our model shows that removing **all** the occupations introduces a systemic error, whereas removing only the occupational categories does not (performed using a Hausman test).

Perhaps more interestingly, all our models showed significant heteroskedasticity. This is, of course, a significant problem given that we modelled our regression on that of the author. This suggests that there may be significant heteroskedastic errors in that paper as well. This is shown through the Cook-Weisberg test (see Figure 1.4) and through correlations with the residuals of our estimation (see Graph A and Figure 1.5).

Cook-Weisberg test on Estimation	**Chi Squared**	**Prob > Chi Squared**
With Occupation	652.47	0.0000
With Occupation Categories	627.62	0.0000
Without any Occupation	600.67	0.0000

Figure 1.4: note the null hypothesis is constant errors



Graph A

Does Volunteer

$Residual^2$	Freq.	Percent	Cum.
0.0-0.1	1	0.04	0.04
0.1-0.2	9	0.33	0.37
0.2-0.3	134	4.94	5.31
0.3-0.4	424	15.63	20.94
0.4-0.5	683	25.18	46.13
0.5-0.6	649	23.93	70.06
0.6-0.7	473	17.44	87.5
0.7-0.8	236	8.7	96.2
0.8-0.9	71	2.62	98.82
0.9-1.0	22	0.81	99.63
1.0-1.1	10	0.37	100
Total	**2712**	**100**	

Does Not Volunteer

$Residual^2$	Freq.	Percent	Cum.
0.0-0.1	7514	83.17	83.17
0.1-0.2	1383	15.31	98.47
0.2-0.3	134	1.48	99.96
0.3-0.4	4	0.04	100
Total	**9035**	**100**	

Does Volunteer

Variable	Obs	Mean	Std. Dev.	Min	Max
$Residuals^2$	2712	5.75	1.51745	1	11

Does Not Volunteer

Variable	Obs	Mean	Std. Dev.	Min	Max
$Residuals^2$	9035	1.184062	0.427226	1	4

Figure 1.5, Note that for presentation purposes the residuals squared were grouped.

Notice that with these large heteroskedasticity errors, we would expect there to be specification errors. This suggests the potential for significant questioning of the original paper and its findings.

III. Conclusion

Data analysis on similar, if not identical data, shows that the author may very well have chosen a model with significant heteroskedasticity errors. In part this is not surprising given that the dependent variable is an indicator variable. A more useful, although perhaps not as media friendly, analysis would have been on the number of hours volunteered. This would have more fully captured any dependency, and we would expect at least a potential for more constant errors.

Appendix A: Results from Volunteer or not: Canada, 1987

Table 1: Impact of individual characteristics on doing volunteer work

Canada, 1987, All Canadians

(As published in the Vaillancourt paper)

Variable	All Canadians
Volunteer Work	
Male	-0.0960 (-5.620)
Age (15-16 years omitted)	
17-19	-0.1433 (-2.850)
20-24	-0.224 (-4.530)
25-34	-0.0456 (-0.940)
35-44	0.1034 (2.090)
45-54	0.1782 (3.400)
55-64	0.1782 (3.340)
65-69	0.1843 (3.080)
70 and over	-0.0311 (-0.540)
Education (none or elementary omitted)	
Secondary	0.2938 (12.000)

Post-Secondary, no degree	0.5879 (17.340)
Post Secondary, with degree	0.5798 (18.810)
University	0.7983 (23.490)
Marital Status (Married omitted)	
Single	-0.0703 (-2.790)
Separated, divorced, widowed	-0.0559 (-2.150)
Occupation (production workers omitted)	
Managers and Professionals	0.2499 (9.860)
Education and health workers	0.2601 (8.540)
Office workers	0.1058 (4.260)
Sales people	0.1192 (4.020)
Work Status (full time omitted)	
Part time	0.1549 (6.270)
Does not work	-0.0108 (-0.360)
Total Income $ (less than 10,000 omitted)	
10,000-14.999	0.0313 (1.030)
15,000-19,999	0.0867 (2.950)
20,000-29,999	0.1321 (5.310)
30,000-39,000	0.2203 (9.380)

40,000-59,999	0.2309 (9.560)
60,000 and over	0.3774 (12.810)
Children Aged 0-2 years+	-0.0785 (-3.420)
Children Aged 3-5 years+	0.1185 (5.380)
Children Aged 6-15 years+	0.1794 (14.970)
Size of city of residence (500,000 and over omitted)	
100,000 - 499,999	0.1421 (5.650)
30,000-99,999	0.1933 (7.540)
Less than 30,000	0.2825 (12.070)
Rural	0.3725 (17.140)
Usual Language (English omitted)	
French	-0.0720 (-2.140)
Others	-0.5607 (-14.380)
Religion (Protestant omitted)	
None	-0.3414 (-14.010)
Catholic	-0.1136 (-5.540)
Others	0.1692 (6.550)

Region (Ontario omitted)	
Atlantic	0.0917 (3.260)
Quebec	-0.1013 (-3.170)
Prairies	0.3731 (17.650)
British Columbia	0.1801 (7.170)
Hours Worked	-0.0023 (-4.440)

Note: T-Statisitcs in ( )

Appendix B: Log File

```
. use d:\neil.dta;
(378 : vas87vol : volunteer activity survey - volunteer file)

. rename f03q34 sex;

. rename  f03q35 marital;

. rename  f03q33 agegroup;

. rename f03q38 educ;

. rename  f05ftpt workf;

. rename  f05q75 occ;

. rename areaflg citysize;

. rename f05q18 workhrs;

. rename f06_q17 vol;

. rename  f06_q30a eng;

. rename  f06_q30b fren;

. rename  f06_q30c other;

. rename  f06_q31 religion;

. rename f06_q35c income;

. rename province region;

. /*
> Data Information
> */
> tab sex;

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      12270       45.86       45.86
     female |      14487       54.14      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab sex, nol;

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      12270       45.86       45.86
          2 |      14487       54.14      100.00
------------+-----------------------------------
      Total |      26757      100.00
```

```
. recode sex 2=0;
(14487 changes made)

. tab marital;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
    married |      17911       66.94       66.94
     single |       5972       22.32       89.26
      other |       2874       10.74      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab marital, nol;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      17911       66.94       66.94
          2 |       5972       22.32       89.26
          3 |       2874       10.74      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab agegroup;

        age group |      Freq.     Percent        Cum.
------------------+-----------------------------------
      15-16 years |       1062        3.97        3.97
      17-19 years |       1308        4.89        8.86
      20-24 years |       2380        8.89       17.75
      25-34 years |       6170       23.06       40.81
      35-44 years |       5434       20.31       61.12
      45-54 years |       3429       12.82       73.94
      55-64 years |       3259       12.18       86.12
      65-69 years |       1390        5.19       91.31
70 years and over |       2325        8.69      100.00
------------------+-----------------------------------
            Total |      26757      100.00

. tab agegroup, nol;

  age group |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       1062        3.97        3.97
          2 |       1308        4.89        8.86
          3 |       2380        8.89       17.75
          4 |       6170       23.06       40.81
          5 |       5434       20.31       61.12
          6 |       3429       12.82       73.94
          7 |       3259       12.18       86.12
          8 |       1390        5.19       91.31
          9 |       2325        8.69      100.00
------------+-----------------------------------
      Total |      26757      100.00
```

```
. tab educ;

                          education |      Freq.     Percent        Cum.
------------------------------------+-----------------------------------
                none or elementary |       4774       17.84       17.84
                       high school |      13348       49.89       67.73
              some post-secondary |       2205        8.24       75.97
post-secondary certificate,diploma |       3616       13.51       89.48
                        university |       2814       10.52      100.00
------------------------------------+-----------------------------------
                             Total |      26757      100.00

. tab educ, nol;

  education |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       4774       17.84       17.84
          2 |      13348       49.89       67.73
          3 |       2205        8.24       75.97
          4 |       3616       13.51       89.48
          5 |       2814       10.52      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab workf;

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
  full-time |      16477       78.92       78.92
  part-time |       4400       21.08      100.00
------------+-----------------------------------
      Total |      20877      100.00

. tab workf, nol;

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      16477       78.92       78.92
          2 |       4400       21.08      100.00
------------+-----------------------------------
      Total |      20877      100.00

. tab occ;

                    type of occupation |      Freq.     Percent
Cum.
----------------------------------------+------------------------------
----
officials and administrators, governmen |        132        0.49
0.49
      other managers and administrators |       1385        5.18
5.67
  management and administration related |        575        2.15
7.82
                  physical,life science |        124        0.46
8.28
```

math,stats,systems analysis and related	138	0.52	8.80
architects and engineers	151	0.56	9.36
architecture and engineering related	139	0.52	9.88
social science and related	377	1.41	11.29
religion	70	0.26	11.55
university and related	84	0.31	11.87
elementary,secondary and related	738	2.76	14.62
other teaching and related	243	0.91	15.53
health diagnosing and treating	85	0.32	15.85
nursing,therapy and related	765	2.86	18.71
medecine and health related	216	0.81	19.52
artistic and recreation	288	1.08	20.59
stenographic and typing	790	2.95	23.55
bookkeeping,account-recording and relat	980	3.66	27.21
office machine and edp operators	166	0.62	27.83
material recording,scheduling and distr	263	0.98	28.81
reception,info, mail and message distri	414	1.55	30.36
library,files,corres,other clerical and	673	2.52	32.87
sales,commodities	1465	5.48	38.35
sales,services and other sales	296	1.11	39.46
protective services	335	1.25	40.71
food,beverage preparation, rel lodging	1224	4.57	45.28
personal,apparel and furnishing service	932	3.48	48.76
other service occupations	698	2.61	51.37
farmers and farm management	753	2.81	54.19
other farming,horticulture and animal h	718	2.68	56.87
fishing,hunting,trapping and related	179	0.67	57.54
forestry and logging	151	0.56	58.10

```
mining and quarrying including gas and  |        122        0.46
58.56
              food beverage and related |        439        1.64
60.20
           other processing occupations |        318        1.19
61.39
  metal shaping and forming occupations |        186        0.70
62.08
            other machining occupations |        116        0.43
62.52
                    metal products,nec |        117        0.44
62.96
electrical,electronics and related equi |        157        0.59
63.54
        textiles,furs and leather goods |        178        0.67
64.21
wood products,rubber,plastics and other |        211        0.79
65.00
mechanics and repairman,except electric |        582        2.18
67.17
  excavating,grading,paving and related |        269        1.01
68.18
electrical power, lighting and wire com |        205        0.77
68.94
              other construction trades |        863        3.23
72.17
               motor tranport operators |        660        2.47
74.63
         other transportation operators |        114        0.43
75.06
                      material handling |        449        1.68
76.74
  other crafts and equiptment operators |        219        0.82
77.56
                           never worked |       1436        5.37
82.92
last worked more 5 yrs ago,or perm unab |       4569       17.08
100.00
----------------------------------------+-------------------------------
----
                                  Total |      26757      100.00

. tab occ, nol;

    type of |
 occupation |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        132        0.49        0.49
          2 |       1385        5.18        5.67
          3 |        575        2.15        7.82
          4 |        124        0.46        8.28
          5 |        138        0.52        8.80
          6 |        151        0.56        9.36
          7 |        139        0.52        9.88
          8 |        377        1.41       11.29
          9 |         70        0.26       11.55
         10 |         84        0.31       11.87
```

11	738	2.76	14.62
12	243	0.91	15.53
13	85	0.32	15.85
14	765	2.86	18.71
15	216	0.81	19.52
16	288	1.08	20.59
17	790	2.95	23.55
18	980	3.66	27.21
19	166	0.62	27.83
20	263	0.98	28.81
21	414	1.55	30.36
22	673	2.52	32.87
23	1465	5.48	38.35
24	296	1.11	39.46
25	335	1.25	40.71
26	1224	4.57	45.28
27	932	3.48	48.76
28	698	2.61	51.37
29	753	2.81	54.19
30	718	2.68	56.87
31	179	0.67	57.54
32	151	0.56	58.10
33	122	0.46	58.56
34	439	1.64	60.20
35	318	1.19	61.39
36	186	0.70	62.08
37	116	0.43	62.52
38	117	0.44	62.96
39	157	0.59	63.54
40	178	0.67	64.21
41	211	0.79	65.00
42	582	2.18	67.17
43	269	1.01	68.18
44	205	0.77	68.94
45	863	3.23	72.17
46	660	2.47	74.63
47	114	0.43	75.06
48	449	1.68	76.74
49	219	0.82	77.56
50	1436	5.37	82.92
51	4569	17.08	100.00
Total	26757	100.00	

```
. tab citysize;
```

area population size codes	Freq.	Percent	Cum.
major metro area	5603	20.94	20.94
other large metro area	3177	11.87	32.81
minor metro area	3641	13.61	46.42
other cities	6137	22.94	69.36
small urban areas and rural areas	8199	30.64	100.00
Total	26757	100.00	

```
. tab citysize, nol;
```

area population size codes	Freq.	Percent	Cum.
1	5603	20.94	20.94
2	3177	11.87	32.81
3	3641	13.61	46.42
4	6137	22.94	69.36
5	8199	30.64	100.00
Total	26757	100.00	

```
. tab workhrs;
```

total actual hours worked last week	Freq.	Percent	Cum.
0	11872	44.37	44.37
1	15	0.06	44.43
2	66	0.25	44.67
3	67	0.25	44.92
4	87	0.33	45.25
5	99	0.37	45.62
6	119	0.44	46.06
7	57	0.21	46.28
8	237	0.89	47.16
9	40	0.15	47.31
10	221	0.83	48.14
11	35	0.13	48.27
12	201	0.75	49.02
13	35	0.13	49.15
14	89	0.33	49.48
15	236	0.88	50.36
16	285	1.07	51.43
17	54	0.20	51.63
18	92	0.34	51.98
19	41	0.15	52.13
20	352	1.32	53.44
21	82	0.31	53.75
22	99	0.37	54.12
23	62	0.23	54.35
24	407	1.52	55.87
25	164	0.61	56.49
26	77	0.29	56.77
27	102	0.38	57.16
28	602	2.25	59.41
29	151	0.56	59.97
30	1104	4.13	64.10
31	182	0.68	64.78
32	3270	12.22	77.00
33	86	0.32	77.32
34	183	0.68	78.00
35	287	1.07	79.07

36	410	1.53	80.61
37	194	0.73	81.33
38	263	0.98	82.31
39	68	0.25	82.57
40	1736	6.49	89.06
41	71	0.27	89.32
42	260	0.97	90.29
43	53	0.20	90.49
44	117	0.44	90.93
45	210	0.78	91.71
46	73	0.27	91.99
47	67	0.25	92.24
48	267	1.00	93.24
49	16	0.06	93.30
50	413	1.54	94.84
51	19	0.07	94.91
52	98	0.37	95.28
53	21	0.08	95.35
54	31	0.12	95.47
55	77	0.29	95.76
56	78	0.29	96.05
57	18	0.07	96.12
58	25	0.09	96.21
59	5	0.02	96.23
60	337	1.26	97.49
61	12	0.04	97.53
62	28	0.10	97.64
63	8	0.03	97.67
64	11	0.04	97.71
65	613	2.29	100.00
Total	26757	100.00	

. tab workhrs, nol;

total actual hours worked last week	Freq.	Percent	Cum.
0	11872	44.37	44.37
1	15	0.06	44.43
2	66	0.25	44.67
3	67	0.25	44.92
4	87	0.33	45.25
5	99	0.37	45.62
6	119	0.44	46.06
7	57	0.21	46.28
8	237	0.89	47.16
9	40	0.15	47.31
10	221	0.83	48.14
11	35	0.13	48.27
12	201	0.75	49.02
13	35	0.13	49.15
14	89	0.33	49.48
15	236	0.88	50.36

16	285	1.07	51.43
17	54	0.20	51.63
18	92	0.34	51.98
19	41	0.15	52.13
20	352	1.32	53.44
21	82	0.31	53.75
22	99	0.37	54.12
23	62	0.23	54.35
24	407	1.52	55.87
25	164	0.61	56.49
26	77	0.29	56.77
27	102	0.38	57.16
28	602	2.25	59.41
29	151	0.56	59.97
30	1104	4.13	64.10
31	182	0.68	64.78
32	3270	12.22	77.00
33	86	0.32	77.32
34	183	0.68	78.00
35	287	1.07	79.07
36	410	1.53	80.61
37	194	0.73	81.33
38	263	0.98	82.31
39	68	0.25	82.57
40	1736	6.49	89.06
41	71	0.27	89.32
42	260	0.97	90.29
43	53	0.20	90.49
44	117	0.44	90.93
45	210	0.78	91.71
46	73	0.27	91.99
47	67	0.25	92.24
48	267	1.00	93.24
49	16	0.06	93.30
50	413	1.54	94.84
51	19	0.07	94.91
52	98	0.37	95.28
53	21	0.08	95.35
54	31	0.12	95.47
55	77	0.29	95.76
56	78	0.29	96.05
57	18	0.07	96.12
58	25	0.09	96.21
59	5	0.02	96.23
60	337	1.26	97.49
61	12	0.04	97.53
62	28	0.10	97.64
63	8	0.03	97.67
64	11	0.04	97.71
65	613	2.29	100.00
Total	26757	100.00	

. tab vol;

did you volunteer

```
at any time |
  past year |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       3532       22.45       22.45
         no |      12202       77.55      100.00
------------+-----------------------------------
      Total |      15734      100.00

. tab vol, nol;

    did you |
  volunteer |
at any time |
  past year |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       3532       22.45       22.45
          2 |      12202       77.55      100.00
------------+-----------------------------------
      Total |      15734      100.00

. tab eng;

   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
yes english |      20730       79.05       79.05
not english |       5493       20.95      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab eng, nol;

   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      20730       79.05       79.05
          2 |       5493       20.95      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab fren;

   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
 yes french |       4714       17.98       17.98
 not french |      21509       82.02      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab fren, nol;
```

```
   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       4714       17.98       17.98
          2 |      21509       82.02      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab other;

   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
  yes other |       1112        4.24        4.24
  not other |      25111       95.76      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab other, nol;

   language |
 speak most |
   often at |
       home |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       1112        4.24        4.24
          2 |      25111       95.76      100.00
------------+-----------------------------------
      Total |      26223      100.00

. tab region;

 region and province |      Freq.     Percent        Cum.
---------------------+-----------------------------------
        newfoundland |       1757        6.57        6.57
prince edward island |        694        2.59        9.16
         nova scotia |       1941        7.25       16.41
       new brunswick |       2033        7.60       24.01
              quebec |       4113       15.37       39.38
             ontario |       5063       18.92       58.31
            manitoba |       2067        7.73       66.03
        saskatchewan |       2668        9.97       76.00
             alberta |       3660       13.68       89.68
    british columbia |       2761       10.32      100.00
---------------------+-----------------------------------
               Total |      26757      100.00

. tab region, nol;

 region and |
   province |      Freq.     Percent        Cum.
------------+-----------------------------------
```

```
         10 |       1757        6.57        6.57
         11 |        694        2.59        9.16
         12 |       1941        7.25       16.41
         13 |       2033        7.60       24.01
         24 |       4113       15.37       39.38
         35 |       5063       18.92       58.31
         46 |       2067        7.73       66.03
         47 |       2668        9.97       76.00
         48 |       3660       13.68       89.68
         59 |       2761       10.32      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab income;

     total household |
 income for 86-before |
     tax   deductions |      Freq.     Percent        Cum.
---------------------+-----------------------------------
    less than $5,000 |        294        1.38        1.38
       $5,000-$9,999 |       1710        8.02        9.40
     $10,000-$14,999 |       2496       11.70       21.10
     $15,000-$19,999 |       2496       11.70       32.81
     $20,000-$29,999 |       3875       18.17       50.98
     $30,000-$39,999 |       4620       21.66       72.64
     $40,000-$59,999 |       4044       18.96       91.60
    $60,000 and more |       1791        8.40      100.00
---------------------+-----------------------------------
               Total |      21326      100.00

. tab income, nol;

      total |
  household |
 income for |
  86-before |
        tax |
 deductions |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        294        1.38        1.38
          2 |       1710        8.02        9.40
          3 |       2496       11.70       21.10
          4 |       2496       11.70       32.81
          5 |       3875       18.17       50.98
          6 |       4620       21.66       72.64
          7 |       4044       18.96       91.60
          8 |       1791        8.40      100.00
------------+-----------------------------------
      Total |      21326      100.00

. tab region;

 region and province |      Freq.     Percent        Cum.
---------------------+-----------------------------------
        newfoundland |       1757        6.57        6.57
prince edward island |        694        2.59        9.16
         nova scotia |       1941        7.25       16.41
```

```
      new brunswick |       2033        7.60       24.01
             quebec |       4113       15.37       39.38
            ontario |       5063       18.92       58.31
           manitoba |       2067        7.73       66.03
       saskatchewan |       2668        9.97       76.00
            alberta |       3660       13.68       89.68
   british columbia |       2761       10.32      100.00
--------------------+-----------------------------------
              Total |      26757      100.00

. tab region, nol;

 region and |
   province |      Freq.     Percent        Cum.
------------+-----------------------------------
         10 |       1757        6.57        6.57
         11 |        694        2.59        9.16
         12 |       1941        7.25       16.41
         13 |       2033        7.60       24.01
         24 |       4113       15.37       39.38
         35 |       5063       18.92       58.31
         46 |       2067        7.73       66.03
         47 |       2668        9.97       76.00
         48 |       3660       13.68       89.68
         59 |       2761       10.32      100.00
------------+-----------------------------------
      Total |      26757      100.00

. /*
> Therefore Sex is our dummy variable, and has females as the base group
> */
>
> tab marital;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
    married |      17911       66.94       66.94
     single |       5972       22.32       89.26
      other |       2874       10.74      100.00
------------+-----------------------------------
      Total |      26757      100.00

. tab marital, nol;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      17911       66.94       66.94
          2 |       5972       22.32       89.26
          3 |       2874       10.74      100.00
------------+-----------------------------------
      Total |      26757      100.00

. /*
> Therefore, marital needs to be transformed into three dummies
>
```

```
> If we wanted to test if having kids is relevant:
> */
>
> gen kids =ownkids1+ownkids2+ownkids3;

. recode kids 9=1 8=1 7=1 6=1 5=1 4=1 3=1 2=1;
(6521 changes made)

. recode region 10=3 11=3 12=3 13=3 24=2 35=1 46=4 47=4 48=4 59=5;
(26757 changes made)

. tab workf, nol m;

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      16477       61.58       61.58
          2 |       4400       16.44       78.02
          . |       5880       21.98      100.00
------------+-----------------------------------
      Total |      26757      100.00

. recode workf .=0;
(5880 changes made)

. /*
> With occupation
> */
>
> xi: regress vol sex i.agegroup i.educ i.marital i.occ i.workf i.income
> i.ownkids1 i.ownkids2 i.ownkids3 i.citysize eng fren other i.religion
>  i.region workhrs kids;
i.agegroup          Iagegr_1-9   (naturally coded; Iagegr_1 omitted)
i.educ              Ieduc_1-5    (naturally coded; Ieduc_1 omitted)
i.marital           Imarit_1-3   (naturally coded; Imarit_1 omitted)
i.occ               Iocc_1-51    (naturally coded; Iocc_1 omitted)
i.workf             Iworkf_0-2   (naturally coded; Iworkf_0 omitted)
i.income            Iincom_1-8   (naturally coded; Iincom_1 omitted)
i.ownkids1          Iownki_0-3   (naturally coded; Iownki_0 omitted)
i.ownkids2          Iownkia0-3   (naturally coded; Iownkia0 omitted)
i.ownkids3          Iownkib0-5   (naturally coded; Iownkib0 omitted)
i.citysize          Icitys_1-5   (naturally coded; Icitys_1 omitted)
i.religion          Irelig_1-4   (naturally coded; Irelig_1 omitted)
i.region            Iregio_1-5   (naturally coded; Iregio_1 omitted)

  Source |       SS       df       MS                  Number of obs =
11747
---------+------------------------------               F(101, 11645) =
9.42
   Model |  157.563146   101  1.56003115               Prob > F      =
0.0000
Residual |  1928.32431 11645   .16559247               R-squared     =
0.0755
---------+------------------------------               Adj R-squared =
0.0675
   Total |  2085.88746 11746  .177582791               Root MSE      =
.40693
```

vol	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sex	.0321131	.0097823	3.283	0.001	.0129381	.051288
Iagegr_2	-.0555099	.0251196	-2.210	0.027	-.1047485	-.0062713
Iagegr_3	-.0576734	.0244881	-2.355	0.019	-.1056741	-.0096726
Iagegr_4	-.0625111	.0243877	-2.563	0.010	-.110315	-.0147071
Iagegr_5	-.1191445	.0250855	-4.750	0.000	-.1683162	-.0699727
Iagegr_6	-.1612816	.0268619	-6.004	0.000	-.2139353	-.1086278
Iagegr_7	-.1606298	.0273436	-5.874	0.000	-.2142279	-.1070318
Iagegr_8	-.2133149	.0305162	-6.990	0.000	-.2731318	-.1534981
Iagegr_9	-.2600447	.0292718	-8.884	0.000	-.3174224	-.202667
Ieduc_2	-.0786713	.0107354	-7.328	0.000	-.0997144	-.0576282
Ieduc_3	-.1512455	.0174629	-8.661	0.000	-.1854758	-.1170153
Ieduc_4	-.1587024	.0159757	-9.934	0.000	-.1900175	-.1273873
Ieduc_5	-.1836109	.0207474	-8.850	0.000	-.2242793	-.1429425
Imarit_2	.0273076	.0125958	2.168	0.030	.0026177	.0519975
Imarit_3	.0098108	.0128415	0.764	0.445	-.0153606	.0349822
Iocc_2	.1468405	.0792653	1.853	0.064	-.0085328	.3022138
Iocc_3	.1968155	.0824341	2.388	0.017	.0352308	.3584002
Iocc_4	.0775776	.1001416	0.775	0.439	-.1187167	.273872
Iocc_5	.212465	.0930875	2.282	0.022	.0299979	.3949322
Iocc_6	.2576839	.0936274	2.752	0.006	.0741585	.4412092
Iocc_7	.0984358	.0945537	1.041	0.298	-.0869053	.283777
Iocc_8	.1335662	.0872164	1.531	0.126	-.0373925	.304525
Iocc_9	.3058433	.1564534	1.955	0.051	-.0008315	.6125182
Iocc_10	.0116874	.1238243	0.094	0.925	-.231029	.2544038
Iocc_11	.1591471	.0851522	1.869	0.062	-.0077655	.3260597

Iocc_12	.2258927	.0929991	2.429	0.015	.043599	.4081865
Iocc_13	.1741267	.1143506	1.523	0.128	-.0500196	.398273
Iocc_14	.1724134	.0810271	2.128	0.033	.0135866	.3312401
Iocc_15	.2746732	.0921562	2.981	0.003	.0940317	.4553147
Iocc_16	.1530604	.0867157	1.765	0.078	-.0169169	.3230377
Iocc_17	.2144804	.0809358	2.650	0.008	.0558327	.3731281
Iocc_18	.2006034	.0799492	2.509	0.012	.0438895	.3573173
Iocc_19	.2645921	.0900636	2.938	0.003	.0880523	.4411318
Iocc_20	.2454638	.0846036	2.901	0.004	.0796266	.411301
Iocc_21	.2230298	.0831469	2.682	0.007	.060048	.3860116
Iocc_22	.2171132	.0810645	2.678	0.007	.0582131	.3760132
Iocc_23	.1516495	.0789293	1.921	0.055	-.0030652	.3063642
Iocc_24	.1182127	.0857945	1.378	0.168	-.049959	.2863844
Iocc_25	.2296013	.0842698	2.725	0.006	.0644184	.3947843
Iocc_26	.2239001	.0789818	2.835	0.005	.0690826	.3787176
Iocc_27	.1919116	.0799604	2.400	0.016	.0351759	.3486473
Iocc_28	.2173009	.0801099	2.713	0.007	.060272	.3743297
Iocc_29	.1354322	.0829487	1.633	0.103	-.0271613	.2980256
Iocc_30	.2333876	.0810873	2.878	0.004	.074443	.3923323
Iocc_31	.2201487	.0888209	2.479	0.013	.0460449	.3942525
Iocc_32	.2461149	.0899667	2.736	0.006	.0697651	.4224647
Iocc_33	.2571083	.0930027	2.765	0.006	.0748074	.4394092
Iocc_34	.2366875	.0816418	2.899	0.004	.0766558	.3967192
Iocc_35	.2772055	.0828177	3.347	0.001	.114869	.439542
Iocc_36	.2549663	.0866289	2.943	0.003	.0851591	.4247734
Iocc_37	.2615628	.0914387	2.861	0.004	.0823277	.4407979
Iocc_38	.2681276	.0917728	2.922	0.003	.0882375	.4480177
Iocc_39	.2760778	.0891877	3.095	0.002	.1012549	.4509006

Iocc_40	.2687465	.0869904	3.089	0.002	.0982307	.4392624
Iocc_41	.2420571	.0854779	2.832	0.005	.074506	.4096081
Iocc_42	.2567827	.0806047	3.186	0.001	.0987839	.4147815
Iocc_43	.2407831	.0848957	2.836	0.005	.0743732	.407193
Iocc_44	.2349454	.0902702	2.603	0.009	.0580007	.4118901
Iocc_45	.2151617	.0796393	2.702	0.007	.0590552	.3712682
Iocc_46	.2341708	.0800759	2.924	0.003	.0772086	.3911329
Iocc_47	.239658	.0944165	2.538	0.011	.0545858	.4247301
Iocc_48	.2197607	.0817303	2.689	0.007	.0595556	.3799657
Iocc_49	.178386	.087541	2.038	0.042	.006791	.349981
Iocc_50	.2498545	.0924815	2.702	0.007	.0685752	.4311337
Iocc_51	.2154076	.090751	2.374	0.018	.0375205	.3932947
Iworkf_1	-.016965	.0487405	-0.348	0.728	-.1125045	.0785746
Iworkf_2	-.0263318	.0497435	-0.529	0.597	-.1238373	.0711737
Iincom_2	.0234786	.0309912	0.758	0.449	-.0372695	.0842266
Iincom_3	-.0132025	.0304854	-0.433	0.665	-.0729589	.046554
Iincom_4	-.0272315	.0306804	-0.888	0.375	-.0873703	.0329073
Iincom_5	-.0287829	.0302633	-0.951	0.342	-.088104	.0305382
Iincom_6	-.0435395	.0303769	-1.433	0.152	-.1030833	.0160043
Iincom_7	-.0660699	.0308741	-2.140	0.032	-.1265884	-.0055515
Iincom_8	-.0941562	.0332722	-2.830	0.005	-.1593752	-.0289371
Iownki_1	.0214991	.0197669	1.088	0.277	-.0172474	.0602456
Iownki_2	.0611728	.035278	1.734	0.083	-.0079779	.1303235
Iownki_3	.1279186	.1838012	0.696	0.486	-.2323626	.4881997
Iownkia1	.0231921	.0161418	1.437	0.151	-.0084485	.0548328
Iownkia2	.0314435	.0385624	0.815	0.415	-.0441454	.1070323
Iownkia3	-.8290114	.408347	-2.030	0.042	-1.62944	-.0285827
Iownkib1	.0303737	.0209106	1.453	0.146	-.0106147	.071362

```
Iownkib2 |   .0015182   .0236757      0.064   0.949      -.0448901
.0479266
Iownkib3 |   .0282264   .0329243      0.857   0.391      -.0363108
.0927636
Iownkib4 |   .0398367   .0578995      0.688   0.491       -.073656
.1533295
Iownkib5 |   .2130335   .1304133      1.634   0.102      -.0425985
.4686655
Icitys_2 |  -.0097056   .0135915     -0.714   0.475      -.0363471
.016936
Icitys_3 |  -.0100951   .0132317     -0.763   0.446      -.0360314
.0158412
Icitys_4 |  -.0342598   .0119751     -2.861   0.004       -.057733   -
.0107866
Icitys_5 |  -.0114432   .0122245     -0.936   0.349      -.0354053
.0125189
     eng |   .0376487    .032592      1.155   0.248       -.026237
.1015345
    fren |     .03357   .0350352      0.958   0.338      -.0351049
.1022449
   other |  -.0805608   .0311613     -2.585   0.010      -.1416422   -
.0194794
Irelig_2 |  -.0201449   .0138633     -1.453   0.146      -.0473193
.0070295
Irelig_3 |  -.0448581   .0132844     -3.377   0.001      -.0708978   -
.0188184
Irelig_4 |  -.0330662   .0168624     -1.961   0.050      -.0661194   -
.0000131
Iregio_2 |    .055946     .01827      3.062   0.002       .0201336
.0917583
Iregio_3 |   .0223716   .0125656      1.780   0.075       -.002259
.0470021
Iregio_4 |  -.0071283   .0116024     -0.614   0.539       -.029871
.0156144
Iregio_5 |  -.0226089   .0151371     -1.494   0.135      -.0522802
.0070624
 workhrs |   .0000736   .0002665      0.276   0.782      -.0004488
.000596
    kids |  -.0231496   .0222215     -1.042   0.298      -.0667074
.0204083
   _cons |   1.856079   .1831479     10.134   0.000       1.497078
2.21508
------------------------------------------------------------------------
------

. hausman, save;

. hettest;

Cook-Weisberg test for heteroscedasticity using fitted values of vol
     Ho: Constant variance
         chi2(1)      =    652.47
         Prob > chi2  =      0.0000

. /*
> Without occupation
> */
```

```
>
> xi: regress vol sex i.agegroup i.educ i.marital i.workf i.income
> i.ownkids1 i.ownkids2 i.ownkids3 i.citysize eng fren other i.religion
>  i.region workhrs kids;
i.agegroup          Iagegr_1-9   (naturally coded; Iagegr_1 omitted)
i.educ              Ieduc_1-5    (naturally coded; Ieduc_1 omitted)
i.marital           Imarit_1-3   (naturally coded; Imarit_1 omitted)
i.workf             Iworkf_0-2   (naturally coded; Iworkf_0 omitted)
i.income            Iincom_1-8   (naturally coded; Iincom_1 omitted)
i.ownkids1          Iownki_0-3   (naturally coded; Iownki_0 omitted)
i.ownkids2          Iownkia0-3   (naturally coded; Iownkia0 omitted)
i.ownkids3          Iownkib0-5   (naturally coded; Iownkib0 omitted)
i.citysize          Icitys_1-5   (naturally coded; Icitys_1 omitted)
i.religion          Irelig_1-4   (naturally coded; Irelig_1 omitted)
i.region            Iregio_1-5   (naturally coded; Iregio_1 omitted)

  Source |       SS       df       MS                  Number of obs =
11747
---------+------------------------------               F( 51, 11695) =
16.76
   Model |  142.054731    51  2.78538688               Prob > F      =
0.0000
Residual |  1943.83273 11695   .16621058               R-squared     =
0.0681
---------+------------------------------               Adj R-squared =
0.0640
   Total |  2085.88746 11746  .177582791               Root MSE      =
.40769

------------------------------------------------------------------------------
------
     vol |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------------
------
     sex |   .0405276   .0082545      4.910   0.000       .0243474
.0567079
Iagegr_2 |  -.0583382   .0247773     -2.354   0.019      -.1069059   -
.0097705
Iagegr_3 |  -.0593026   .0239401     -2.477   0.013      -.1062293    -
.012376
Iagegr_4 |  -.0679166   .0237091     -2.865   0.004      -.1143904   -
.0214428
Iagegr_5 |  -.1293923    .024323     -5.320   0.000      -.1770694   -
.0817151
Iagegr_6 |  -.1755496   .0261359     -6.717   0.000      -.2267803   -
.1243189
Iagegr_7 |  -.1742744   .0264913     -6.579   0.000      -.2262017   -
.1223471
Iagegr_8 |  -.2328848   .0295452     -7.882   0.000      -.2907984   -
.1749713
Iagegr_9 |    -.28094   .0280146    -10.028   0.000      -.3358534   -
.2260267
 Ieduc_2 |  -.0836043   .0106478     -7.852   0.000      -.1044758   -
.0627328
 Ieduc_3 |  -.1643923   .0172241     -9.544   0.000      -.1981545   -
.1306302
```

Ieduc_4	|	-.1726939	.0152758	-11.305	0.000	-.2026369	-.1427508
Ieduc_5	|	-.2168058	.0185935	-11.660	0.000	-.2532522	-.1803594
Imarit_2	|	.0277819	.012544	2.215	0.027	.0031936	.0523701
Imarit_3	|	.0126913	.0128263	0.989	0.322	-.0124503	.0378329
Iworkf_1	|	-.0314512	.0132805	-2.368	0.018	-.0574831	-.0054193
Iworkf_2	|	-.052245	.0147457	-3.543	0.000	-.0811491	-.023341
Iincom_2	|	.0254872	.030997	0.822	0.411	-.0352721	.0862466
Iincom_3	|	-.0103709	.0304775	-0.340	0.734	-.070112	.0493702
Iincom_4	|	-.023211	.0306585	-0.757	0.449	-.0833068	.0368848
Iincom_5	|	-.0246281	.0302198	-0.815	0.415	-.083864	.0346079
Iincom_6	|	-.0405551	.0303246	-1.337	0.181	-.0999965	.0188862
Iincom_7	|	-.0664538	.0307865	-2.159	0.031	-.1268005	-.006107
Iincom_8	|	-.1010766	.0331834	-3.046	0.002	-.1661215	-.0360316
Iownki_1	|	.0196732	.0197549	0.996	0.319	-.0190497	.058396
Iownki_2	|	.0569584	.0352585	1.615	0.106	-.0121542	.1260709
Iownki_3	|	.1444977	.1838286	0.786	0.432	-.215837	.5048325
Iownkia1	|	.0245538	.0161424	1.521	0.128	-.0070881	.0561957
Iownkia2	|	.0321625	.0385166	0.835	0.404	-.0433366	.1076615
Iownkia3	|	-.7971523	.4083187	-1.952	0.051	-1.597525	.0032204
Iownkib1	|	.0302185	.0208788	1.447	0.148	-.0107074	.0711444
Iownkib2	|	.0009545	.0236408	0.040	0.968	-.0453854	.0472945
Iownkib3	|	.0245316	.0328908	0.746	0.456	-.0399398	.0890029
Iownkib4	|	.0436285	.0578452	0.754	0.451	-.0697578	.1570147
Iownkib5	|	.2278861	.1302579	1.749	0.080	-.0274411	.4832134
Icitys_2	|	-.0075218	.0135481	-0.555	0.579	-.0340783	.0190347
Icitys_3	|	-.0072281	.0131799	-0.548	0.583	-.0330629	.0186068
Icitys_4	|	-.0313599	.0118658	-2.643	0.008	-.0546187	-.008101
Icitys_5	|	-.0095384	.0118876	-0.802	0.422	-.0328401	.0137632

```
      eng |   .0374744   .0325494      1.151   0.250       -.0263279
.1012767
     fren |   .0317823   .0350128      0.908   0.364       -.0368485
.1004132
    other |  -.0893659   .0311147     -2.872   0.004       -.1503558   -
.0283759
 Irelig_2 |  -.0197579   .0138424     -1.427   0.154       -.0468914
.0073756
 Irelig_3 |   -.043877   .0132666     -3.307   0.001       -.0698818   -
.0178721
 Irelig_4 |  -.0307739   .0168424     -1.827   0.068       -.0637877
.00224
 Iregio_2 |    .053969   .0182279      2.961   0.003        .0182392
.0896987
 Iregio_3 |   .0204132   .0124418      1.641   0.101       -.0039747
.0448012
 Iregio_4 |  -.0143542   .0114885     -1.249   0.212       -.0368736
.0081652
 Iregio_5 |  -.0238437   .0150786     -1.581   0.114       -.0534003
.0057129
  workhrs |  -.0001109   .0002599     -0.427   0.670       -.0006203
.0003985
     kids |  -.0222843   .0221982     -1.004   0.315       -.0657964
.0212278
    _cons |    2.11405   .1584272     13.344   0.000        1.803506
2.424593
-------------------------------------------------------------------------
------

. hausman;

            ---- Coefficients ----
         |      (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |     Prior       Current       Difference          S.E.
---------+-------------------------------------------------------------
     sex |   .0321131     .0405276       -.0084146     .0052493
Iagegr_2 |  -.0555099    -.0583382        .0028283     .0041325
Iagegr_3 |  -.0576734    -.0593026        .0016293     .0051513
Iagegr_4 |  -.0625111    -.0679166        .0054056     .0057129
Iagegr_5 |  -.1191445    -.1293923        .0102478     .0061378
Iagegr_6 |  -.1612816    -.1755496        .0142681     .0062028
Iagegr_7 |  -.1606298    -.1742744        .0136445     .0067739
Iagegr_8 |  -.2133149    -.2328848        .0195699     .0076367
Iagegr_9 |  -.2600447      -.28094        .0208954     .0084865
 Ieduc_2 |  -.0786713    -.0836043         .004933     .0013681
 Ieduc_3 |  -.1512455    -.1643923        .0131468      .002878
 Ieduc_4 |  -.1587024    -.1726939        .0139915     .0046771
 Ieduc_5 |  -.1836109    -.2168058         .033195     .0092052
Imarit_2 |   .0273076     .0277819       -.0004743     .0011415
Imarit_3 |   .0098108     .0126913       -.0028805     .0006243
Iworkf_1 |   -.016965    -.0314512        .0144862     .0468963
Iworkf_2 |  -.0263318     -.052245        .0259132     .0475076
Iincom_2 |   .0234786     .0254872       -.0020087            .
Iincom_3 |  -.0132025    -.0103709       -.0028316      .000691
Iincom_4 |  -.0272315     -.023211       -.0040205     .0011596
Iincom_5 |  -.0287829    -.0246281       -.0041548     .0016206
Iincom_6 |  -.0435395    -.0405551       -.0029844     .0017813
```

```
Iincom_7 |  -.0660699    -.0664538         .0003838     .0023238
Iincom_8 |  -.0941562    -.1010766         .0069204     .0024292
Iownki_1 |   .0214991     .0196732         .0018259     .0006908
Iownki_2 |   .0611728     .0569584         .0042144     .0011705
Iownki_3 |   .1279186     .1444977        -.0165792            .
Iownkia1 |   .0231921     .0245538        -.0013617            .
Iownkia2 |   .0314435     .0321625         -.000719      .001879
Iownkia3 |  -.8290114    -.7971523        -.0318591      .004815
Iownkib1 |   .0303737     .0302185         .0001552      .001154
Iownkib2 |   .0015182     .0009545         .0005637     .0012847
Iownkib3 |   .0282264     .0245316         .0036948     .0014864
Iownkib4 |   .0398367     .0436285        -.0037917     .0025063
Iownkib5 |   .2130335     .2278861        -.0148526      .006365
Icitys_2 |  -.0097056    -.0075218        -.0021838     .0010855
Icitys_3 |  -.0100951    -.0072281        -.0028671     .0011691
Icitys_4 |  -.0342598    -.0313599        -.0028999     .0016145
Icitys_5 |  -.0114432    -.0095384        -.0019048     .0028503
     eng |   .0376487     .0374744         .0001743     .0016648
    fren |     .03357     .0317823         .0017876     .0012536
   other |  -.0805608    -.0893659         .0088051      .001705
Irelig_2 |  -.0201449    -.0197579         -.000387     .0007603
Irelig_3 |  -.0448581     -.043877        -.0009811     .0006877
Irelig_4 |  -.0330662    -.0307739        -.0022924     .0008224
Iregio_2 |    .055946      .053969          .001977       .00124
Iregio_3 |   .0223716     .0204132         .0019583     .0017593
Iregio_4 |  -.0071283    -.0143542         .0072259     .0016217
Iregio_5 |  -.0226089    -.0238437         .0012348     .0013295
 workhrs |   .0000736    -.0001109         .0001845     .0000591
    kids |  -.0231496    -.0222843        -.0008653      .001018
---------+-----------------------------------------------------------
            b = less efficient estimates obtained previously from
regress.
            B = fully efficient estimates obtained from regress.

Test:  Ho:  difference in coefficients not systematic

            chi2( 51) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    16.20
            Prob>chi2 =     1.0000

. hausman, clear;

. /*
> Now recoding occupation as in the paper
> */
> gen occcat=occ;

. recode occcat 1/9=1 10/16=2 17/22=3 23/28=4 29/33=5 34=4 35/49=5
50/51=6;
(26625 changes made)

. xi: regress vol sex i.agegroup i.educ i.marital i.occcat i.workf
i.income
> i.ownkids1 i.ownkids2 i.ownkids3 i.citysize eng fren other i.religion
>  i.region workhrs kids;
i.agegroup          Iagegr_1-9   (naturally coded; Iagegr_1 omitted)
i.educ              Ieduc_1-5    (naturally coded; Ieduc_1 omitted)
```

```
i.marital            Imarit_1-3   (naturally coded; Imarit_1 omitted)
i.occcat             Ioccca_1-6   (naturally coded; Ioccca_1 omitted)
i.workf              Iworkf_0-2   (naturally coded; Iworkf_0 omitted)
i.income             Iincom_1-8   (naturally coded; Iincom_1 omitted)
i.ownkids1           Iownki_0-3   (naturally coded; Iownki_0 omitted)
i.ownkids2           Iownkia0-3   (naturally coded; Iownkia0 omitted)
i.ownkids3           Iownkib0-5   (naturally coded; Iownkib0 omitted)
i.citysize           Icitys_1-5   (naturally coded; Icitys_1 omitted)
i.religion           Irelig_1-4   (naturally coded; Irelig_1 omitted)
i.region             Iregio_1-5   (naturally coded; Iregio_1 omitted)

  Source |       SS       df       MS                  Number of obs =
11747
---------+------------------------------               F( 56, 11690) =
15.76
   Model |  146.436402    56  2.61493575               Prob > F      =
0.0000
Residual |  1939.45106 11690  .165906848               R-squared     =
0.0702
---------+------------------------------               Adj R-squared =
0.0657
   Total |  2085.88746 11746  .177582791               Root MSE      =
.40732

------------------------------------------------------------------------
------
     vol |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
------
     sex |    .032243   .0091152      3.537   0.000       .0143756
.0501104
Iagegr_2 |  -.0598032   .0247695     -2.414   0.016      -.1083556   -
.0112508
Iagegr_3 |  -.0621361   .0239249     -2.597   0.009      -.1090328   -
.0152393
Iagegr_4 |  -.0697037    .023704     -2.941   0.003      -.1161675   -
.0232399
Iagegr_5 |  -.1281327   .0243247     -5.268   0.000      -.1758133   -
.0804522
Iagegr_6 |  -.1745722   .0261257     -6.682   0.000       -.225783   -
.1233614
Iagegr_7 |  -.1740813    .026502     -6.569   0.000      -.2260297   -
.1221329
Iagegr_8 |  -.2319032   .0295306     -7.853   0.000      -.2897882   -
.1740183
Iagegr_9 |  -.2800779    .028008    -10.000   0.000      -.3349783   -
.2251776
 Ieduc_2 |  -.0814517   .0106716     -7.633   0.000      -.1023699   -
.0605335
 Ieduc_3 |  -.1564391   .0173438     -9.020   0.000      -.1904358   -
.1224425
 Ieduc_4 |  -.1605266   .0157006    -10.224   0.000      -.1913024   -
.1297508
 Ieduc_5 |  -.1880216   .0198376     -9.478   0.000      -.2269065   -
.1491366
```

Imarit_2	.0274571	.012538	2.190	0.029	.0028806	.0520336
Imarit_3	.0119733	.0128187	0.934	0.350	-.0131534	.0371
Ioccca_2	.0200087	.020536	0.974	0.330	-.0202453	.0602628
Ioccca_3	.0585842	.01792	3.269	0.001	.0234579	.0937105
Ioccca_4	.0375701	.0164053	2.290	0.022	.0054129	.0697273
Ioccca_5	.0734052	.0159403	4.605	0.000	.0421596	.1046508
Ioccca_6	.0558995	.0499737	1.119	0.263	-.0420572	.1538562
Iworkf_1	-.0265508	.0485122	-0.547	0.584	-.1216429	.0685413
Iworkf_2	-.0431997	.0493786	-0.875	0.382	-.13999	.0535906
Iincom_2	.0257635	.030974	0.832	0.406	-.0349508	.0864777
Iincom_3	-.0103823	.0304613	-0.341	0.733	-.0700916	.049327
Iincom_4	-.0246997	.0306469	-0.806	0.420	-.0847727	.0353732
Iincom_5	-.0257862	.0302249	-0.853	0.394	-.0850321	.0334598
Iincom_6	-.0404306	.030331	-1.333	0.183	-.0998845	.0190232
Iincom_7	-.0649689	.0308089	-2.109	0.035	-.1253594	-.0045784
Iincom_8	-.0972097	.0332227	-2.926	0.003	-.1623317	-.0320876
Iownki_1	.0200002	.0197386	1.013	0.311	-.0186907	.0586911
Iownki_2	.0557946	.0352317	1.584	0.113	-.0132654	.1248546
Iownki_3	.1369765	.1836815	0.746	0.456	-.2230699	.497023
Iownkia1	.0231812	.0161345	1.437	0.151	-.0084451	.0548076
Iownkia2	.031204	.0384928	0.811	0.418	-.0442482	.1066563
Iownkia3	-.7862258	.4079738	-1.927	0.054	-1.585922	.0134709
Iownkib1	.0275453	.0208729	1.320	0.187	-.0133691	.0684597
Iownkib2	-.0016679	.0236381	-0.071	0.944	-.0480025	.0446667
Iownkib3	.024184	.0328809	0.736	0.462	-.0402681	.0886361
Iownkib4	.0366583	.0578189	0.634	0.526	-.0766764	.1499931
Iownkib5	.2131587	.1301871	1.637	0.102	-.0420297	.4683472
Icitys_2	-.0087611	.0135527	-0.646	0.518	-.0353268	.0178045

```
Icitys_3 |  -.0092045   .0131924     -0.698   0.485      -.0350638
.0166547
Icitys_4 |  -.0332083   .0118965     -2.791   0.005      -.0565275    -
.009889
Icitys_5 |  -.0145645   .0119912     -1.215   0.225      -.0380693
.0089402
     eng |    .038327   .0325245      1.178   0.239      -.0254264
.1020804
    fren |   .0321667   .0349838      0.919   0.358      -.0364075
.1007408
   other |   -.085165   .0311008     -2.738   0.006      -.1461279   -
.0242022
Irelig_2 |  -.0199384   .0138339     -1.441   0.150      -.0470552
.0071784
Irelig_3 |  -.0442059   .0132585     -3.334   0.001      -.0701947   -
.0182172
Irelig_4 |  -.0302815   .0168286     -1.799   0.072      -.0632683
.0027053
Iregio_2 |   .0545934   .0182229      2.996   0.003       .0188734
.0903134
Iregio_3 |   .0227345   .0124532      1.826   0.068      -.0016757
.0471448
Iregio_4 |  -.0122635   .0114942     -1.067   0.286      -.0347941
.0102671
Iregio_5 |  -.0230834   .0150683     -1.532   0.126      -.0526198
.006453
 workhrs |  -.0000635   .0002617     -0.243   0.808      -.0005764
.0004494
    kids |  -.0203219   .0221845     -0.916   0.360      -.0638071
.0231633
   _cons |   2.049875   .1657797     12.365   0.000        1.72492
2.374831
------------------------------------------------------------------------
------

. hausman, save;

. hettest;

Cook-Weisberg test for heteroscedasticity using fitted values of vol
     Ho: Constant variance
         chi2(1)      =    627.62
         Prob > chi2  =      0.0000

. /*
> Testing to see if occupational categories are significant.
> Answer, yes as a group they seem to be.
> Then cross-tab with workf.  We get results that are contradictory
within the
> categories, so despite the Ftest, we choose to drop occupation
> */
> test Ioccca_2 Ioccca_3 Ioccca_4 Ioccca_5 Ioccca_6;

 ( 1)  Ioccca_2 = 0.0
 ( 2)  Ioccca_3 = 0.0
 ( 3)  Ioccca_4 = 0.0
 ( 4)  Ioccca_5 = 0.0
```

```
 ( 5)  Ioccca_6 = 0.0

       F(  5, 11690) =    5.28
            Prob > F =    0.0001


. tab occcat workf;

           |           type of job
    occcat |         0  full-time  part-time |     Total
-----------+---------------------------------+----------
         1 |         0       2889        202 |      3091
         2 |         0       1719        700 |      2419
         3 |         0       2380        906 |      3286
         4 |         0       3547       1842 |      5389
         5 |         0       5832        735 |      6567
         6 |      5880        110         15 |      6005
-----------+---------------------------------+----------
     Total |      5880      16477       4400 |     26757


. /*
> Without occupation
> */
>
> xi: regress vol sex i.agegroup i.educ i.marital i.workf i.income
> i.ownkids1 i.ownkids2 i.ownkids3 i.citysize eng fren other i.religion
>  i.region workhrs kids;
i.agegroup          Iagegr_1-9   (naturally coded; Iagegr_1 omitted)
i.educ              Ieduc_1-5    (naturally coded; Ieduc_1 omitted)
i.marital           Imarit_1-3   (naturally coded; Imarit_1 omitted)
i.workf             Iworkf_0-2   (naturally coded; Iworkf_0 omitted)
i.income            Iincom_1-8   (naturally coded; Iincom_1 omitted)
i.ownkids1          Iownki_0-3   (naturally coded; Iownki_0 omitted)
i.ownkids2          Iownkia0-3   (naturally coded; Iownkia0 omitted)
i.ownkids3          Iownkib0-5   (naturally coded; Iownkib0 omitted)
i.citysize          Icitys_1-5   (naturally coded; Icitys_1 omitted)
i.religion          Irelig_1-4   (naturally coded; Irelig_1 omitted)
i.region            Iregio_1-5   (naturally coded; Iregio_1 omitted)

  Source |       SS       df       MS                  Number of obs =
11747
---------+------------------------------               F( 51, 11695) =
16.76
   Model |  142.054731    51  2.78538688               Prob > F      =
0.0000
Residual |  1943.83273 11695   .16621058               R-squared     =
0.0681
---------+------------------------------               Adj R-squared =
0.0640
   Total |  2085.88746 11746  .177582791               Root MSE      =
.40769

------------------------------------------------------------------------
------
     vol |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
```

```
---------+--------------------------------------------------------------
------
     sex |   .0405276   .0082545      4.910   0.000       .0243474
.0567079
Iagegr_2 |  -.0583382   .0247773     -2.354   0.019      -.1069059   -
.0097705
Iagegr_3 |  -.0593026   .0239401     -2.477   0.013      -.1062293    -
.012376
Iagegr_4 |  -.0679166   .0237091     -2.865   0.004      -.1143904   -
.0214428
Iagegr_5 |  -.1293923    .024323     -5.320   0.000      -.1770694   -
.0817151
Iagegr_6 |  -.1755496   .0261359     -6.717   0.000      -.2267803   -
.1243189
Iagegr_7 |  -.1742744   .0264913     -6.579   0.000      -.2262017   -
.1223471
Iagegr_8 |  -.2328848   .0295452     -7.882   0.000      -.2907984   -
.1749713
Iagegr_9 |    -.28094   .0280146    -10.028   0.000      -.3358534   -
.2260267
 Ieduc_2 |  -.0836043   .0106478     -7.852   0.000      -.1044758   -
.0627328
 Ieduc_3 |  -.1643923   .0172241     -9.544   0.000      -.1981545   -
.1306302
 Ieduc_4 |  -.1726939   .0152758    -11.305   0.000      -.2026369   -
.1427508
 Ieduc_5 |  -.2168058   .0185935    -11.660   0.000      -.2532522   -
.1803594
Imarit_2 |   .0277819    .012544      2.215   0.027       .0031936
.0523701
Imarit_3 |   .0126913   .0128263      0.989   0.322      -.0124503
.0378329
Iworkf_1 |  -.0314512   .0132805     -2.368   0.018      -.0574831   -
.0054193
Iworkf_2 |   -.052245   .0147457     -3.543   0.000      -.0811491    -
.023341
Iincom_2 |   .0254872    .030997      0.822   0.411      -.0352721
.0862466
Iincom_3 |  -.0103709   .0304775     -0.340   0.734       -.070112
.0493702
Iincom_4 |   -.023211   .0306585     -0.757   0.449      -.0833068
.0368848
Iincom_5 |  -.0246281   .0302198     -0.815   0.415       -.083864
.0346079
Iincom_6 |  -.0405551   .0303246     -1.337   0.181      -.0999965
.0188862
Iincom_7 |  -.0664538   .0307865     -2.159   0.031      -.1268005    -
.006107
Iincom_8 |  -.1010766   .0331834     -3.046   0.002      -.1661215   -
.0360316
Iownki_1 |   .0196732   .0197549      0.996   0.319      -.0190497
.058396
Iownki_2 |   .0569584   .0352585      1.615   0.106      -.0121542
.1260709
Iownki_3 |   .1444977   .1838286      0.786   0.432       -.215837
.5048325
```

```
Iownkia1 |   .0245538   .0161424      1.521   0.128       -.0070881
.0561957
Iownkia2 |   .0321625   .0385166      0.835   0.404       -.0433366
.1076615
Iownkia3 |  -.7971523   .4083187     -1.952   0.051       -1.597525
.0032204
Iownkib1 |   .0302185   .0208788      1.447   0.148       -.0107074
.0711444
Iownkib2 |   .0009545   .0236408      0.040   0.968       -.0453854
.0472945
Iownkib3 |   .0245316   .0328908      0.746   0.456       -.0399398
.0890029
Iownkib4 |   .0436285   .0578452      0.754   0.451       -.0697578
.1570147
Iownkib5 |   .2278861   .1302579      1.749   0.080       -.0274411
.4832134
Icitys_2 |  -.0075218   .0135481     -0.555   0.579       -.0340783
.0190347
Icitys_3 |  -.0072281   .0131799     -0.548   0.583       -.0330629
.0186068
Icitys_4 |  -.0313599   .0118658     -2.643   0.008       -.0546187    -
.008101
Icitys_5 |  -.0095384   .0118876     -0.802   0.422       -.0328401
.0137632
     eng |   .0374744   .0325494      1.151   0.250       -.0263279
.1012767
    fren |   .0317823   .0350128      0.908   0.364       -.0368485
.1004132
   other |  -.0893659   .0311147     -2.872   0.004       -.1503558   -
.0283759
Irelig_2 |  -.0197579   .0138424     -1.427   0.154       -.0468914
.0073756
Irelig_3 |   -.043877   .0132666     -3.307   0.001       -.0698818   -
.0178721
Irelig_4 |  -.0307739   .0168424     -1.827   0.068       -.0637877
.00224
Iregio_2 |    .053969   .0182279      2.961   0.003        .0182392
.0896987
Iregio_3 |   .0204132   .0124418      1.641   0.101       -.0039747
.0448012
Iregio_4 |  -.0143542   .0114885     -1.249   0.212       -.0368736
.0081652
Iregio_5 |  -.0238437   .0150786     -1.581   0.114       -.0534003
.0057129
 workhrs |  -.0001109   .0002599     -0.427   0.670       -.0006203
.0003985
    kids |  -.0222843   .0221982     -1.004   0.315       -.0657964
.0212278
   _cons |    2.11405   .1584272     13.344   0.000        1.803506
2.424593
--------------------------------------------------------------------------
------

. /*
> Residuals, Specification Errors & Heteroskedasticity
> */
>
```

```
> predict res, residuals;
(15010 missing values generated)

. gen ressq=res^2;
(15010 missing values generated)

. graph ressq vol;

. sort vol;

. gen ressqcat=ressq;
(15010 missing values generated)

. recode ressqcat 0/0.1=1 0.1/0.2=2 0.2/0.3=3 0.3/0.4=4 0.4/0.5=5
>  0.5/0.6=6 0.6/0.7=7 0.7/0.8=8 0.8/0.9=9 0.9/1=10 1/1.1=11;
(11747 changes made)

. by vol: tab ressqcat;

-> vol=     yes
   ressqcat |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          1        0.04        0.04
          2 |          9        0.33        0.37
          3 |        134        4.94        5.31
          4 |        424       15.63       20.94
          5 |        683       25.18       46.13
          6 |        649       23.93       70.06
          7 |        473       17.44       87.50
          8 |        236        8.70       96.20
          9 |         71        2.62       98.82
         10 |         22        0.81       99.63
         11 |         10        0.37      100.00
------------+-----------------------------------
      Total |       2712      100.00

-> vol=      no
   ressqcat |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       7514       83.17       83.17
          2 |       1383       15.31       98.47
          3 |        134        1.48       99.96
          4 |          4        0.04      100.00
------------+-----------------------------------
      Total |       9035      100.00

-> vol=       .  no observations


. by vol: sum ressqcat;

-> vol=     yes
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
ressqcat |    2712        5.75    1.51745          1         11

-> vol=      no
```

```
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
ressqcat |    9035    1.184062   .4272261          1          4

-> vol=       .
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
ressqcat |       0


. hausman;

            ---- Coefficients ----
         |      (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |     Prior       Current       Difference        S.E.
---------+-------------------------------------------------------------
     sex |    .032243     .0405276        -.0082846      .0038665
Iagegr_2 |  -.0598032    -.0583382         -.001465             .
Iagegr_3 |  -.0621361    -.0593026        -.0028334             .
Iagegr_4 |  -.0697037    -.0679166        -.0017871             .
Iagegr_5 |  -.1281327    -.1293923         .0012595      .0002897
Iagegr_6 |  -.1745722    -.1755496         .0009774             .
Iagegr_7 |  -.1740813    -.1742744         .0001931      .0007545
Iagegr_8 |  -.2319032    -.2328848         .0009816             .
Iagegr_9 |  -.2800779      -.28094         .0008621             .
 Ieduc_2 |  -.0814517    -.0836043         .0021526      .0007126
 Ieduc_3 |  -.1564391    -.1643923         .0079532      .0020335
 Ieduc_4 |  -.1605266    -.1726939         .0121673      .0036277
 Ieduc_5 |  -.1880216    -.2168058         .0287842      .0069145
Imarit_2 |   .0274571     .0277819        -.0003248             .
Imarit_3 |   .0119733     .0126913         -.000718             .
Iworkf_1 |  -.0265508    -.0314512         .0049004       .046659
Iworkf_2 |  -.0431997     -.052245         .0090453      .0471255
Iincom_2 |   .0257635     .0254872         .0002762             .
Iincom_3 |  -.0103823    -.0103709        -.0000114             .
Iincom_4 |  -.0246997     -.023211        -.0014888             .
Iincom_5 |  -.0257862    -.0246281        -.0011581      .0005554
Iincom_6 |  -.0404306    -.0405551         .0001245      .0006233
Iincom_7 |  -.0649689    -.0664538         .0014848      .0011727
Iincom_8 |  -.0972097    -.1010766         .0038669      .0016162
Iownki_1 |   .0200002     .0196732         .0003271             .
Iownki_2 |   .0557946     .0569584        -.0011638             .
Iownki_3 |   .1369765     .1444977        -.0075212             .
Iownkia1 |   .0231812     .0245538        -.0013726             .
Iownkia2 |    .031204     .0321625        -.0009584             .
Iownkia3 |  -.7862258    -.7971523         .0109265             .
Iownkib1 |   .0275453     .0302185        -.0026732             .
Iownkib2 |  -.0016679     .0009545        -.0026225             .
Iownkib3 |    .024184     .0245316        -.0003475             .
Iownkib4 |   .0366583     .0436285        -.0069701             .
Iownkib5 |   .2131587     .2278861        -.0147274             .
Icitys_2 |  -.0087611    -.0075218        -.0012393       .000356
Icitys_3 |  -.0092045    -.0072281        -.0019765      .0005731
Icitys_4 |  -.0332083    -.0313599        -.0018484      .0008552
Icitys_5 |  -.0145645    -.0095384        -.0050261       .001573
     eng |    .038327     .0374744         .0008526             .
    fren |   .0321667     .0317823         .0003843             .
```

```
   other |   -.085165    -.0893659         .0042008            .
Irelig_2 |  -.0199384    -.0197579        -.0001805            .
Irelig_3 |  -.0442059     -.043877         -.000329            .
Irelig_4 |  -.0302815    -.0307739         .0004924            .
Iregio_2 |   .0545934      .053969         .0006244            .
Iregio_3 |   .0227345     .0204132         .0023213     .0005322
Iregio_4 |  -.0122635    -.0143542         .0020907      .000362
Iregio_5 |  -.0230834    -.0238437         .0007603            .
 workhrs |  -.0000635    -.0001109         .0000474     .0000307
    kids |  -.0203219    -.0222843         .0019624            .
---------+-----------------------------------------------------------
            b = less efficient estimates obtained previously from
regress.
            B = fully efficient estimates obtained from regress.

Test:  Ho:  difference in coefficients not systematic

            chi2( 50) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    27.65
            Prob>chi2 =     0.9957

. hettest;

Cook-Weisberg test for heteroscedasticity using fitted values of vol
     Ho: Constant variance
         chi2(1)      =    600.67
         Prob > chi2  =      0.0000

. /*
> F-tests
> */
> test Iagegr_2 Iagegr_3 Iagegr_4 Iagegr_5 Iagegr_6 Iagegr_7 Iagegr_8
Iagegr_9;

 ( 1)  Iagegr_2 = 0.0
 ( 2)  Iagegr_3 = 0.0
 ( 3)  Iagegr_4 = 0.0
 ( 4)  Iagegr_5 = 0.0
 ( 5)  Iagegr_6 = 0.0
 ( 6)  Iagegr_7 = 0.0
 ( 7)  Iagegr_8 = 0.0
 ( 8)  Iagegr_9 = 0.0

       F(  8, 11695) =   20.46
            Prob > F =    0.0000


. test Ieduc_2 Ieduc_3 Ieduc_4 Ieduc_5;

 ( 1)  Ieduc_2 = 0.0
 ( 2)  Ieduc_3 = 0.0
 ( 3)  Ieduc_4 = 0.0
 ( 4)  Ieduc_5 = 0.0

       F(  4, 11695) =   53.45
            Prob > F =    0.0000
```

```
. test Imarit_2 Imarit_3;

 ( 1)  Imarit_2 = 0.0
 ( 2)  Imarit_3 = 0.0

       F(  2, 11695) =    2.66
            Prob > F =    0.0700


. test Iworkf_1 Iworkf_2;

 ( 1)  Iworkf_1 = 0.0
 ( 2)  Iworkf_2 = 0.0

       F(  2, 11695) =    6.34
            Prob > F =    0.0018


. test Iincom_2 Iincom_3 Iincom_4 Iincom_5 Iincom_6 Iincom_7 Iincom_8;

 ( 1)  Iincom_2 = 0.0
 ( 2)  Iincom_3 = 0.0
 ( 3)  Iincom_4 = 0.0
 ( 4)  Iincom_5 = 0.0
 ( 5)  Iincom_6 = 0.0
 ( 6)  Iincom_7 = 0.0
 ( 7)  Iincom_8 = 0.0

       F(  7, 11695) =    7.06
            Prob > F =    0.0000


. test Iownki_1 Iownki_2 Iownki_3;

 ( 1)  Iownki_1 = 0.0
 ( 2)  Iownki_2 = 0.0
 ( 3)  Iownki_3 = 0.0

       F(  3, 11695) =    1.09
            Prob > F =    0.3537


. test Iownkia1 Iownkia2 Iownkia3;

 ( 1)  Iownkia1 = 0.0
 ( 2)  Iownkia2 = 0.0
 ( 3)  Iownkia3 = 0.0

       F(  3, 11695) =    2.18
            Prob > F =    0.0878


. test Iownkib1 Iownkib2 Iownkib3 Iownkib4 Iownkib5;

 ( 1)  Iownkib1 = 0.0
 ( 2)  Iownkib2 = 0.0
```

```
 ( 3)  Iownkib3 = 0.0
 ( 4)  Iownkib4 = 0.0
 ( 5)  Iownkib5 = 0.0

       F(  5, 11695) =    1.49
            Prob > F =    0.1897


. test Icitys_2 Icitys_3 Icitys_4 Icitys_5;

 ( 1)  Icitys_2 = 0.0
 ( 2)  Icitys_3 = 0.0
 ( 3)  Icitys_4 = 0.0
 ( 4)  Icitys_5 = 0.0

       F(  4, 11695) =    2.06
            Prob > F =    0.0831


. test eng fren;

 ( 1)  eng = 0.0
 ( 2)  fren = 0.0

       F(  2, 11695) =    0.68
            Prob > F =    0.5048


. test eng fren other;

 ( 1)  eng = 0.0
 ( 2)  fren = 0.0
 ( 3)  other = 0.0

       F(  3, 11695) =   15.88
            Prob > F =    0.0000


. test Irelig_2 Irelig_3 Irelig_4;

 ( 1)  Irelig_2 = 0.0
 ( 2)  Irelig_3 = 0.0
 ( 3)  Irelig_4 = 0.0

       F(  3, 11695) =    4.27
            Prob > F =    0.0051


. test Iregio_2 Iregio_3 Iregio_4 Iregio_5;

 ( 1)  Iregio_2 = 0.0
 ( 2)  Iregio_3 = 0.0
 ( 3)  Iregio_4 = 0.0
 ( 4)  Iregio_5 = 0.0

       F(  4, 11695) =    5.29
            Prob > F =    0.0003
```

```
.
end of do-file
-> . log close


.
```

ECON 452

Project 3

Health Services Utilization of Canada's Immigrant and
Non-Immigrant Populations

Authors:

Data Set:

Name: General Social Survey (GSS)

File #: 35

Paper:

"Health Status and Health Services Utilization of Canada's Immigrant and Non-Immigrant Populations" by **Mireille Laroche**. *Canadian Public Policy, Vol. XXVI, No. 1, 2000*.

1. Introduction

Under Canada's *Immigration Act*, every immigrant applicant needs to undergo successfully a medical examination in order to immigrate to Canada. Applicants are judged inadmissible to immigrate if they are likely to be a danger to public health or safety, or if their admission could generate excessive demands on health or social services. While the immigration legislation ensures a satisfactory health condition for those entering the country, it cannot guarantee the maintenance of such a condition through time. The objective of the current study is to assess possible differences in the utilization rates of health services between immigrants and those of the Canadian-born population. Pursuing this line of inquiry will help policymakers evaluate the impact of immigration on the health-care system as well as the efficiency of the health-screening policy in place.

Our analysis is built on a recent study by Laroche (2000)[1]. She compared the health status of immigrants and their utilization rates of health services to those of the Canadian-born population using data contained within two cycles (1985 and 1991) of the General Social Survey (GSS). Her results showed that neither the health status of immigrants nor their utilization rates of health services differ significantly from those of the Canadian-born population.

Laroche used three self-reported measures of health status. The first measure describes the health status of the respondents by subjectively qualifying it in categories that range from "excellent" to "poor". The second measure is a binary variable, taking the value of one if the respondent has at least one health problem and zero otherwise. Finally, the third measure takes the value one if the respondent suffers from any long-term activity limitations and zero otherwise. All three measures of health status were regressed on an extensive number of explanatory variables. The ordered probit estimation method was preferred to OLS since with the former approach, the categorical dependent variables could be evaluated on a non-linear scale.

Health services utilization was estimated using four different dependent variables: the length of stay in hospital and the number of consultations with a general practitioner, a specialist or a nurse the respondents have had during the 12 months preceding the time of the interview[2]. Since a large number of observations for these variables were clustered at zero, Ordinary Least Squares (OLS) estimation method could not be used as the resulting estimates could have been subject to bias[3]. Instead, a two-limit tobit estimation procedure was used to regress the length of stay in a hospital and the number of consultations on the explanatory variables.

The explanatory variables for the utilization of health services regressions were immigration status, time of arrival, sex, weight, type of smoker, age, age-squared, marital status, number of children, household income, education, occupation and mother tongue. Interaction variables were created between immigration status and household income, education, age, age-squared and mother tongue.

The estimated coefficients on the dummy variables related to the immigration status and interaction variables between immigration status and various socio-economic characteristics were generally not , when tested jointly, significantly different from zero, indicating that immigrants' and non-immigrants' use of health services is not significantly different. Laroche concluded that Canada's immigrant population is more or less as healthy as the average native-born Canadian is and will use, on average, similar amounts of health-care services. Her findings reflect the fact that the medical screening process that immigrant applicants must pass before immigrating to Canada efficiently 'filters out' those with severe medical conditions.

Our aim in the current study is to assess possible differences in the utilization rates of health services by both populations using the 1991 cycle of the GSS. In particular, for the sub-sample of respondents who have spent some time in a hospital during the 12 months prior to the time of the interview, we test whether the length of stay differs for immigrants and non-immigrants. This modification of Laroche's model allows us to use the OLS method of estimation. Our results show that the utilization rates of health services of immigrants do not differ significantly from those of the Canadian-born population and thus support Laroche's findings.

The study is divided into four sections. The second part outlines our manipulation of the data and introduces a regression model similar to the one used by Laroche. Section 3 presents the regression results and a final section discusses our main conclusions.

2. Data

The 1991 cycle of the GSS provides information on the health condition, usage of medical services in the prior 12 months, and socio-economic characteristics of 11,924 respondents of which approximately 1,700 are immigrants. Residents of the Yukon and Northwest Territories were excluded. The survey involves interviews with non-institutionalized Canadians 15 years of age or older. Since invididuals in the survey were not selected using random sampling, weights developed by Statistics Canada were used to adjust the quantitative estimates.

Laroche used a broad range of explanatory variables to verify the importance of age, education, and several other socio-economic characteristics in the determination of a person's utilization rate of health services, regardless of the respondent's immigration status. However, our focus is on possible differences in utilization rates between the two groups. Consequently, only relevant variables were selected (refer to **Table 1**).

Several transformations of the extracted variables were required to make them compatible to our OLS approach. We eliminate those respondents who did not spend any time in hospital, which meant deleting observations that have missing values for these variables. As a result 10,420 of the total 11,924 observations are removed leaving 1,504. A dummy variable called *imm* is generated to indicate whether a person is an immigrant or not. It takes the value of 0 if the person is Canadian-born and 1 if born outside Canada. Observations that had missing values for *imm* are dropped as there would be no basis on which to differentiate between immigrants and non-immigrants if that variable had a missing value. This eliminated a further 34 observations leaving 1,470. Respondents born outside of Canada and for whom no 'age at immigration' was reported were also eliminated from our sub-sample. Hence, we are left with 1,460 observations to conduct our regression. Of these, 14.2% (207 observations) are immigrants.

The effect of age on the length of stay in hospital was incorporated in two explanatory variables. Since the age variable provided in the GSS survey, *dvagegr*, is a range variable, we use the range midpoint to provide single values for age. This is stored in a variable called *midage*. A variable *agesq* is then generated, which is simply the square of *midage*. The intuition behind the use of this variable is that the relationship between the number of nights spent in hospital and age is unlikely to be linear. One drawback associated with the midpoint method is the presence of unequal ranges for the age variable.

Laroche's use of dummy variables for the cohort of arrival of immigrants was a source of confusion. She included these variables in order to identify possible changes in health services utilization by the immigrant population over time. This line of inquiry is not the focus of our paper. Consequently we exclude cohort dummy variables entirely from our regression as they have no bearing on our analysis. Dummy variables related to education are also singled out under the assumption that the level of education is partially reflected in household income.

A further transformation creates a variable that reflects the fraction of their life a respondent has spent in Canada. The idea is that immigrants who have spent most of their lives in Canada are assumed to enjoy a level of health similar to native-born Canadians. The variable *immper* captures this information. It takes a value of 1 if the respondent was born in Canada and a value given by the formula (*midage* – *dvageimc*)/*midage* otherwise. This formula calculates the fraction of an immigrant's life spent in Canada.

Finally, with regards to household income, dummy variables like the ones used by Laroche are generated. These are *inc1* (income less that $10000), *inc2* (income greater than or equal to $10000 and less than $20,000), *inc3* (income greater than or equal to $20000 and less than $40000), and *inc4* (income greater than $40,000) [4].

Our regression thus takes the form[5]:

$$c2a = b_1 + b_2 imm + b_3 immper + b_4 midage + b_5 agesq + b_6 inc2 + b_7 inc3 + b_8 inc4$$

3. Results

The regression was run using the OLS method of estimation. Coefficient estimates and their corresponding p-value are reported in **Table 2**. The key parameter is the one related to the immigration status, *imm*. A t-test was conducted to verify if the coefficient for immigration status is significant. The following null and alternative hypotheses were formulated:

$$H_0: b2 = 0$$

$$H_A: b2 \mathrel{!}= 0$$

The resulting p-value of 0.487 strongly retains the null, agreeing with the general findings of Laroche that there is no significant difference in health services utilization between immigrants and non-immigrants. Interestingly, the parameter related to the fraction of life spent in Canada (*immper*) shows up negative. It is however not significant and is therefore not given further consideration.

As expected, the coefficient for age (*midage*) is not statistically significant, whereas the one related to age-squared (*agesq*) is highly significant (p-value = 0.004). This indicates that the relationship between the time spent in hospital and age is not linear - a plausible result as hospital use is expected to increase exponentially as an individual gets older.

The coefficients related to the dummy variables for income are significant at the 5% level for low and high levels of income. This suggests that as income rises, the time spent in hospital decreases.

4. Summary

Laroche's paper addresses some concerns about immigration and immigration policy in Canada. For one, it allays the fear that the increased number of immigrants in recent years places a disproportionately large burden on the healthcare system. In addition, the screening process employed by immigration Canada is vindicated from being too lax.

Our results demonstrate that there is no significant differences in the rates of utilization of health services between immigrants and non-immigrants in Canada and thus support Laroche's findings. The explanatory variables age-squared and household income explain most of the variation in the dependent variable.

However, the poor quality of our model is reflected in the low value of the R^2 statistic ($R^2 = 0.0559$). Recognizing the fact that only one variable is statistically significant at the 1% level, it is not surprising that such a low R^2 was obtained. Moreover, even if most of the variables were found to be significant, it is unlikely that we would obtain a high R^2. It is to be noted that there are many factors that might affect the time spent in hospital that were not accounted for in our regression model. In particular, as mentioned earlier, the study fails to capture the unmet needs of individuals. Hence this shortcoming in the data has yielded a low R^2. This inadequacy was prevalent in Laroche's paper as well because her coeffcient estimates for most variables were also statistically insignificant.

In spite of our limited understanding of the probit and tobit estimation methods, our modified model yielded the same results as Laroche. Based on the conclusion derived from both Laroche's and our results, it was interesting to learn that both our studies collided with the misconception that immigrants would have a higher level of health service utilization than non-immigrants. In essence, our study, which employed a simplified model, reaffirmed the conclusions arrived at by Laroche.

5. Tables

Table 1

Description of Selected Variables from GSS

GSS Variable	Description
c2a	Number of nights spent in hospital, nursing home, etc… during the past 12 months
q11	Country of birth 1 – If country of birth is Canada 2 – If born outside Canada
Dvageimc	Age at time of immigration

Dvhhinc	Household income (coded value ranges) 1 – no income 2 - < 5000 3 – 5000-9999 4 – 10000-14999 5 – 15000-19999 6 – 20000-29999 7 – 30000-39999 8 – 40000-59999 9 – 60000-79999 10 - >=80000
Dvagegr	Age of respondent (coded value ranges) 1 – 15-17 2 – 18-19 3 – 20-24 4 – 25-29 5 – 30-34 6 – 35-39 7 – 40-44 8 – 45-49 9 – 50-54 10 – 55-59 11 – 60-64 12 – 65-69 13 – 70-74 14 – 75-79 15 – 80-99

Table 2 Regression Results	
Variable	Estimated Coefficient (p-values in parentheses)
Immigration Status	-2.61 (0.487)
Percentage of life spent in Canada	-8.14 (0.278)
Estimated age	-0.322 (0.116)
Estimated age-squared	0.0055 (0.004)
Household income greater than or equal to $10000 and less than $20,000	-4.52 (0.036)
Household income greater than or equal to $20000 and less than $40000	-3.15 (0.150)
Household income greater than $40,000	-5.08 (0.024)
Constant	22.0 (0.016)

Appendix

Table 3 Summary Statistics for respondents who have spent time in hospital	
Variable	Means (std. Errors in parentheses)
Time Spent in Hospital (days)	12.1 (30.3)

Country of birth (1=Canada; 2=Outside)	1.14 (0.349)
Age at time of immigration	21.7 (14.96)
Household income	6.06 (2.02)
Age of respondent	8.86 (4.41)
Fraction of life spent in Canada	0.944 (0.174)
Estimated Age of respondent using median of ranges	52.2 (23.0)
Age-squared	3253.8 (2485.6)
Number of observations	1460

LOG FILE

```
-> . Qextract
getting information about file 35 ...
loading variables from 35 (gss6_91) only (no data yet)...  done
-> . summarize c2a

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     c2a |    1504    12.65891   32.49902          1        365
-> . summarize c2a if dvageimc > 0 & dvageimc < 99

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     c2a |     200       15.25   42.76108          1        365
-> . gen imm=1 if q11 == 2
```

```
(10052 missing values generated)
-> . replace imm=0 if q11 == 1
(9863 real changes made)
-> . drop if c2a==.
(10420 observations deleted)
-> . drop if imm==.
(34 observations deleted)
-> . summarize dvagegr if dvagegr == 15

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
 dvagegr |     153          15           0         15         15
-> . browse
-> . gen midage=16 if dvagegr == 1
(1434 missing values generated)
-> . run "C:\WINDOWS\TEMP\STD000000.tmp"
-> . summarize midage

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
  midage |    1470    52.28129    23.01062         16       89.5
-> . gen immper = 1 if q11 == 1
(217 missing values generated)
-> . replace immper = (midage - dvageimc)/midage if q11 == 2
(207 real changes made)
-> . sort q11
-> . gen inc1 = (dvhhinc >= 1 & dvhhinc <=3)
-> . gen inc2 = (dvhhinc >= 4 & dvhhinc <=5)
-> . gen inc3 = (dvhhinc >= 6 & dvhhinc <=7)
-> . gen inc4 = (dvhhinc >= 8 & dvhhinc <= 10)
```

```
-> . summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     c2a |    1470    12.16122   30.31187          1        365
     q11 |    1470    1.147619   .3548427          1          2
dvageimc |     207    21.65217   14.95932          0         50
 dvhhinc |    1176    6.061224   2.024445          1         10
 dvagegr |    1470     8.87483   4.408434          1         15
     imm |    1470     .147619   .3548427          0          1
  midage |    1470    52.28129   23.01062         16       89.5
  immper |    1460    .9436651   .1736334  -.0454545          1
    inc1 |    1470    .0952381   .2936434          0          1
    inc2 |    1470    .2326531   .4226667          0          1
    inc3 |    1470    .2387755   .4264804          0          1
    inc4 |    1470    .2333333   .4230965          0          1
-> . drop if immper==.
(10 observations deleted)
. regress c2a imm immper midage agesq inc2 inc3 inc4

  Source |       SS       df       MS                  Number of obs =    1460
---------+------------------------------               F(  7,  1452) =   12.28
   Model |  74763.8708     7   10680.553               Prob > F      =  0.0000
Residual |  1263019.39  1452  869.848066               R-squared     =  0.0559
---------+------------------------------               Adj R-squared =  0.0513
   Total |  1337783.26  1459  916.917932               Root MSE      =  29.493

------------------------------------------------------------------------------
     c2a |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
```

```
     imm |  -2.611017   3.754138    -0.696   0.487     -9.975131    4.753096
  immper |  -8.135771   7.493506    -1.086   0.278     -22.83502    6.563483
  midage |  -.3222281   .2047106    -1.574   0.116     -.7237882     .079332
   agesq |   .0055148   .0019112     2.886   0.004      .0017658    .0092638
    inc2 |  -4.524167   2.155772    -2.099   0.036     -8.752927   -.2954076
    inc3 |  -3.145445   2.185488    -1.439   0.150     -7.432496    1.141605
    inc4 |  -5.082878   2.252384    -2.257   0.024     -9.501153   -.6646025
   _cons |   22.03957   9.110446     2.419   0.016      4.168525    39.91061
------------------------------------------------------------------------------
-> . summm

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     c2a |    1460    12.11918   30.28065          1        365
     q11 |    1460    1.141781   .3489447          1          2
dvageimc |     207    21.65217   14.95932          0         50
 dvhhinc |    1170    6.061538   2.024987          1         10
 dvagegr |    1460    8.860959   4.409223          1         15
     imm |    1460    .1417808   .3489447          0          1
  midage |    1460    52.20274   23.00105         16       89.5
  immper |    1460    .9436651   .1736334  -.0454545          1
    inc1 |    1460    .0952055   .2935991          0          1
    inc2 |    1460    .2335616   .4232414          0          1
    inc3 |    1460    .2390411    .426644          0          1
    inc4 |    1460    .2335616   .4232414          0          1
   agesq |    1460    3253.812   2485.558        256    8010.25
 immper2 |    1460    94.36651   17.36334  -4.545455        100
```

---

[1] "Health Status and Health Services Utilization of Canada's Immigrant and Non-Immigrant Populations" Canadian Public Policy Vol.XXVI no.1 2000

[2] Since this study only assesses health services utilization through *actual* hospitalization/consultations rather than desired or required medical attention; it fails to capture the unmet needs of individuals.

[3] For both surveys, approximately 70 percent of respondents did not consult a specialist, while about 89 percent of them reported having no consultations with a nurse. In 1985, 20 percent of respondents reported having no consultations with a general practitioner. This proportion fell to 16 percent in 1991. Finally, in both survey years, approximately 87 percent of the respondents did not stay overnight in a hospital.

[4] The base dummy variable for household income was *inc1*. Hence it is excluded from the regression

[5] All betas are estimates

Have Trade Unions altered the Gender Gap?
Canadian Evidence

Department of Economics
Queen's University

May 25, 2001

Data Set:

1984 Survey of Union Membership
QED DLI Archive File Number: 374

Model Paper:

Maki, D. and Ng, I. (1990), "Effects of trade unions on the earnings differential between male and females: Canadian evidence", *Canadian Journal of Economics*, **2**, 305 – 311.

I. Introduction

Earnings differential between males and females has long been an interest of many researchers. However, at the start of the 1990s, relatively few studies had examined the effects of trade unions on the male-female wage gap. The negligible effects found in the United States, does not immediately imply the same is true in Canada. A study has shown that the male-female wage gap is larger in the union sectors than in the non-union sectors in Canada, suggesting trade unions may have enlarged the wage difference between male and female workers. This motivates Maki and Ng to perform their empirical study, “Effects of trade unions on the earnings differential between males and females: Canadian evidence”.

The purpose of this paper is to first give an overview of Maki and Ng’s study. Section II therefore begins with an outline of their objective, and follows by a description of their data, theoretical considerations, methods, and conclusion. Section III is the beginning of our attempt to create a model that is similar to those used by Maki and Ng in their study. It includes a description of our data set, along with a comparison that is made between our data set and the data set used by Maki and Ng. Section IV is a depiction of our model, along with estimation and hypothesis testing. Section V concludes.

II. Maki and Ng's study

As implicitly stated in the title, the authors' objective is to see whether trade unions have an impact on the earnings gap between male and female workers in Canada. If such an impact has existed, the next response is to find out its direction and magnitude. The micro data, 1984 Survey of Union Membership (SUM), which involved approximately 75,000 Canadian residents, who are age fifteen or older, is used to examine this issue. Given this relatively large sample size, the authors decided to limit their analysis. This is done by firstly, selecting 10 percent random sample of all cases and secondly, eliminating all individuals who are not classified as employed. This leaves a sample of 4,093 individuals, with 2,293 males and 1,800 females.

The theoretical considerations made by the authors are explanations for why the effect of trade unions on male-female earnings differential is an empirical issue. The authors broke down the overall effect of trade unions on male-female differential into three parts: male-female membership effect, the extent which unions differently affected wages of unionised male and female workers, and the extent which unions differently affected wages of non-unionised male and female workers. The effect of the first component is deterministic, in the sense that the male union density was greater than the one of female in 1981. Thus, for any given level of unionisation, unions will increase the male-female wage gap. However, the latter two components, the wage effect, may widen or reduce the male-female wage gap. Thus, an examination of empirical data is needed to draw a conclusion.

The main model used by the authors in their analysis is a lin-lin model that takes the form as follows:

$$
\begin{aligned}
WAGE = a_0 &+ a_1 A2534 + a_3 A4554 + a_4 A55 \\
&+ a_5 EDHS + a_6 EDSPS + a_7 EDDIP + a_8 EDDEG \\
&+ a_9 MARRD + a_{10} PTIME + a_{11} UNION + a_{12} MINES \\
&+ a_{13} MFG + a_{14} CONS + a_{15} TRANS + a_{16} TRADE \\
&+ a_{17} COMM + a_{18} PUBLIC \qquad (1)
\end{aligned}
$$

where: WAGE is the dependent variable, and all independent variables are indicator variables for age, education, marital status, types of job (i.e. part-time or full-time), and industries. Estimations of equation (1) are performed separately for males and females and moreover, separately for public and private sector.

The authors draw two main conclusions from their regression analysis. Firstly, unions have widened the male-female wage gap, even when differences in the level of unionisation are taken into account. Secondly, the effects of trade unions on male-female wage gaps are different between the public and private sectors, with the wage gap increasing in the private sector, but decreasing in the public sector.

III. Data

The micro data, 1984 Survey of Union Memberships (SUM) is also used in the present analysis and obtained though the Queen's Economics Department's Data Liberation Initiative (DLI) Archive. In an attempt to create a similar data set, we generated variables that are used by Maki and Ng in their study. At the initial stage, we discovered there are observations with dv1, the hourly wage rate or the dependent variable in equation (1), classified as missing. Given the importance of dv1 in our analysis, we eliminated these observations. We then followed the sample selection criteria that are outlined by Maki and Ng in their paper. We first generated a 10 percent random sample and then, removed individuals that are not classified as employed. This leaves a sample of 3416 individuals of which 1823 are males and 1593 are females. As noticed, our data set is different from Maki and Ng's data set, which contains a sample of 4,093 individuals, with 2,293 males and 1,800 females. Consequently, different results will be obtained, leading to different interpretations.

Table 1 provides a summary of variable definitions and their sample means used in the present analysis. Given that there are differences in the sample size, variable definitions, and other uncertainties arising from the sample selection process, the sample means are not identical. However, they are similar to the ones shown in Maki and Ng's paper, in the sense that there are consistent social trends observed. For instance, the mean hourly wage rate is higher for males than for females. The reason for this may be because on average, there are higher proportions of male workers with a university degree than females. Another observable trend is that females are on average, more likely to work part-time than males. Consequently, this may explain why the degree of unionisation is higher for males, since part-time workers tend not to be unionised. Lastly, male workers prefer to work in the manufacturing sectors, while female workers more likely to work in the areas of community, business and personal service. For the public sector, the proportions of male and female workers are approximately the same on average.

IV. Estimation and Hypothesis Testing

Instead of following a separate regression approach, as in Maki and Ng's case, we chose females as the base group and used a pooled (interactive) regression function, which is illustrated as follows:

$$
\begin{aligned}
WAGE = \beta_0 + \beta_1 A2534 + \beta_3 A4554 + \beta_4 A55 \\
+ \beta_5 EDHS + \beta_6 EDSPS + \beta_7 EDDIP + \beta_8 EDDEG \\
+ \beta_9 MARRD + \beta_{10} PTIME + \beta_{11} UNION \\
+ \beta_{12} MINES + \beta_{13} MFG + \beta_{14} CONS + \beta_{15} TRANS \\
+ \beta_{16} TRADE + \beta_{17} COMM + \beta_{18} PUBLIC \\
+ \beta_{19} MALE + \beta_{20} MA2534 + \beta_{21} MA4554 + \beta_{22} MA55 \\
+ \beta_{23} MEDHS + \beta_{24} MEDSPS + \beta_{25} MEDDIP + \beta_{26} MEDDEG \\
+ \beta_{27} MMARRD + \beta_{28} MPTIME + \beta_{29} MUNION \\
+ \beta_{30} MMINES + \beta_{31} MMFG + \beta_{32} MCONS + \beta_{33} MTRANS \\
+ \beta_{34} MTRADE + \beta_{35} MCOMM + \beta_{36} MPUBLIC \qquad (2)
\end{aligned}
$$

Coefficient estimates are provided in Table 2, along with t-ratios.

To establish whether a conditional mean male-female wage differential exits, we test the following hypothesis:

$$H_0: \beta_j = 0 \qquad \text{for } j = 19, \ldots, 36$$

$$H_1: \beta_j \neq 0 \qquad \text{for } j = 19, \ldots, 36$$

Given the F-statistic is equal to 12.71, with a p-value of 0.0000, the null hypothesis is rejected in favour of the alternative hypothesis at the one percent significance level. Thus, a male-female wage gap exists. General F-tests are also applied to test whether the gender gap differs across age groups or across education levels, with all else remaining constant in both cases. Provided the test statistics are 1.84 and 1.08, respectively, with the corresponding p-values of 0.1188 and 0.3660, the null hypotheses are retained at the one percent significance level, indicating the wage gap neither differs across age groups nor across education levels.

A t-test is used to test the relevance of MUNION, an interaction term which is a product of two dummy variables, MALE and UNION. Since the test statistics equals -2.863, with a p-value of 0.004, the null hypothesis is rejected at the one percent significance level, indicating trade unions have affected males' earnings differently from females' earnings. Lastly, to determine whether the gender gap differs across industries, a general F-test is applied. The test-statistic is equal to 8.64, with a p-value of 0.0000, and consequently, the null hypothesis test is rejected at the one percent significance level, indicating there is strong evidence that the gender gap is industries-dependent.

V. Conclusion

The issue of whether trade unions have an impact on the conditional mean wage differential between male and female workers in Canada is the central theme of the present analysis. This empirical topic is not new, in the sense that Maki and Ng have examined this issue with the Survey of Union Memberships (SUM) conducted in 1984. They found that trade unions have enlarged the gender gap, even when the differences in the degree of unionization are taken into considerations. Moreover, the effects of unions are different in the private sector than in the public sector, with the gender gap widening in the private sector, while the reverse is true in the public sector.

We attempted to replicate a similar data set used by Maki and Ng to examine this issue. However, the lack of descriptions in the model paper created difficulties in yielding very similar results. For instance, the authors did not describe as to how they generated a 10 percent random sample of all cases, or how they grouped various industries into few major categories. Uncertainties also arose in the process of handing observations having the hourly wage rate reported as missing. Consequently, we made an assumption and a modification. We assumed that the authors eliminated these missing observations prior to generating a 10 percent random sample. Moreover, we redefined the dummy variable for marital status, with a value of one referring to an individual who is married, and a value of zero corresponding to a person who is not currently married (i.e. single, widow or others).

Instead of using two separate regression functions for males and females, as in Maki and Ng's study, we followed a more informative and flexible approach, the pooled (interactive) regression, and chose females as the base group. After estimation, we performed various hypotheses tests, with the first about whether a conditional mean wage gap exists between males and females. Test results indicate there is a gender gap, which varies across industries, but not across age groups or education levels. Moreover, trade unions have affected males' earnings differently from females' earnings.

We suggested future research to find explanations for the opposing effects of trade unions on the wage gaps found in the public and private sectors and moreover, to investigate the dynamics of trade unions. The latter means to examine the effects of trade unions on the gender gap, in terms of directions and magnitudes over time. However, given the usual trade-off between cost-effectiveness and accuracy, this may have to be done using not micro, but aggregate data.

We gained many insights in the present analysis. In the classroom, everything is always assumed to be in control, but there are many factors, whether endogenous or exogenous, in the outside world. As recalled from previous experiences, there are no uncertainties or guesses arising from the data set. For example, there are no missing values, and variables used in assignments are always well defined. This was not the case here. For example, there are observations with the hourly wage rate reported as missing. Moreover, variable definitions are not clearly defined in the model paper. Consequently, adjustments have to be made, and sometimes, this process could be frustrating and time-consuming. Another interesting point arising from the analytical process was that the complexity of the models increases as more parameters are added. This was illustrated by moving from a separate to a pooled (interactive) regression approach. Although the latter approach is more flexible and informative, additional parameters made interpretations of regression coefficient estimates more difficult.

VI. Appendix

A. Tables

TABLE 1: Variable Definitions and Sample Means

Variable	Definition	Sample Means Males	Females	Average
WAGE	Hourly Earnings	1108¢	846¢	986¢
Age				
A2534	Age 25 to 34 = 1; other = 0	30%	30%	30%
A3544	Age 35 to 44 = 1; other = 0	25%	23%	24%
A4554	Age 45 to 54 = 1; other = 0	14%	14%	14%
A55	Age 55 and over = 1; other = 0	10%	8%	9%
Education				
EDHS	High school completion = 1; other = 0	52%	54%	53%
EDSPS	Some post-secondary education = 1; other = 0	9%	9%	9%
EDDIP	Post-secondary diploma = 1; other = 0	11%	17%	14%
EDDEG	University degree = 1; other = 0	14%	12%	13%
Marital Status				
MARRD	Married = 1; other = 0	70%	63%	67%
Type of Work				
PTIME	Part-time worker = 1; full-time worker = 0	8%	27%	17%
Member of Union				
UNION	Union member = 1; non-union member = 0	39%	34%	37%
Recoded Industries				
MINES	Mines, Quarries, and Oil Wells = 1; other = 0	4%	0.4%	2%
MFG	Manufacturing = 1; other = 0	23%	10%	17%
CONS	Construction = 1; other = 0	7%	1%	4%
TRANS	Transportation, Communication and utilities = 1; other = 0	12%	4%	8%
TRADE	Wholesale and Retail Trade; Finance, Insurance and Real Estate = 1; other = 0	19%	27%	23%
COMM	Community, Business and Personal Service = 1; other = 0	22%	48%	34%
PUBLIC	Public Administration = 1; other = 0	10%	8%	9%

SOURCE: 1984 Survey of Union Membership

TABLE 2: Determinants of Hourly Earnings

Independent Variable	Description	Coefficient	t-value
Age			
A2534	Age 25 to 34 = 1; other = 0	151.80	5.27
A2544	Age 35 to 44 = 1; other = 0	258.81	8.26
A4554	Age 45 to 54 = 1; other = 0	245.33	7.00
A55	Age 55 and over = 1; other = 0	198.91	4.72
Education			
EDHS	High school completion = 1; other = 0	152.11	3.78
EDSPS	Some post-secondary education = 1; other = 0	275.65	5.47
EDDIP	Post-secondary diploma = 1; other = 0	398.03	8.73
EDDEG	University degree = 1; other = 0	650.97	13.19
Marital Status			
MARRD	Married = 1; other = 0	28.69	1.29
Type of Work			
PTIME	Part-time worker = 1; full-time worker = 0	-24.05	-1.05
Member of Union			
UNION	Union member = 1; non-union member = 0	235.98	10.12
Recoded Industries			
MINES	Mines, Quarries, and Oil Wells = 1; other = 0	-141.58	-0.79
MFG	Manufacturing = 1; other = 0	-329.33	-3.75
CONS	Construction = 1; other = 0	-248.20	-1.91
TRANS	Transportation, Communication, Utilities = 1; other = 0	-132.98	-1.40
TRADE	Wholesale and Retail Trade; Finance, Insurance and Real Estate	-339.28	-4.05
COMM	Community, Business, and Personal Service = 1; other = 0	-314.33	-3.77
PUBLIC	Public Administration = 1; other = 0	-268.77	-2.99
Gender			
MALE	MALE = 1; FEMALE = 0	-226.69	-2.11
Interactions with Male (M)			
Age			
MA2534	Age 25 to 34	31.89	42.39
MA3544	Age 35 to 44	60.13	1.31
MA4554	Age 45 to 54	120.63	2.35
MA55	Age 55 and over	109.41	1.86
Education			
MEDHS	High school completion	28.00	0.57
MEDSPS	Some post-secondary education	70.67	1.09
MEDDIP	Post-secondary diploma	-25.63	-0.43
MEDDEG	University degree	64.17	1.04

Marital Status			
MMARRD	Married	66.68	2.00
Type of Work			
MPTIME	Part-time worker	-145.63	-3.33
Member of Union			
MUNION	Union member	-88.64	-2.86
Recoded Industries			
MMINES	Mines, Quarries, and Oil Wells	561.04	2.93
MMFG	Manufacturing	459.70	4.49
MCONS	Construction	466.50	3.27
MTRANS	Transportation, Communication and Utilties	372.97	3.38
MTRADE	Wholesale and Retail Trade; Finance, Insurance and Real Estate	343.02	3.47
MCOMM	Community; Business and Personal Service	213.09	2.16
MPUBLIC	Public Administration	487.85	4.56
CONSTANT		652.16	7.10
F (37, 3378)		69.67	
n	Number of Observations	3416	

SOURCE: 1984 Survey of Union Membership

B. Log File

```
-> . Qextract
getting information about file 374 ...
loading variables from 374 (sum84) only (no data yet)...  done
-> . *THIS IS A LIST OF STATA COMMANDS FOR HW1.
-> .
-> . count if dv1 == .
40510
-> . count if dv1 != .
44166
-> . do "C:\windows\TEMP\STD050000.tmp"

.
. /*
> Given the importance of dv1, the hourly wage rate (the dependent
>  variable chosen), we ELIMINATE OBSERVATIONS WITH dv1
> CLASSIFIED AS MISSING.
> */
.
. drop if dv1 == .
(40510 observations deleted)

.
. /*
> GENERATE A 10 PERCENT RANDOM SAMPLES
> */
.
. gen u = uniform()

. sort u

. drop if _n > 0.1*_N
(39750 observations deleted)

. drop u

.
. #delimit
delimiter now ;
. /*
> 36. lfstatus
> KEEP OBSERVATIONS only IF classified as EMPLOYED
> */
>
> tab lfstatus;

labour force status |      Freq.     Percent        Cum.
--------------------+-----------------------------------
           employed |       3416       77.36       77.36
         unemployed |        405        9.17       86.53
not in labour force |        595       13.47      100.00
--------------------+-----------------------------------
              Total |       4416      100.00

. tab lfstatus, nolabel;

     labour |
      force |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
```

```
         1 |       3416       77.36       77.36
         2 |        405        9.17       86.53
         3 |        595       13.47      100.00
------------+-----------------------------------
      Total |       4416      100.00

. drop if lfstatus != 1;
(1000 observations deleted)

. /*
> GENERATE DUMMIES.
> */

. /*
> 72. dv1
> generate a variable, WAGE, equal to dv1
> */
>
> gen WAGE = dv1;

. /*
> 5. sex
> check: number of males and females same as in the paper?
> paper:
> */
> display 2293 + 1800;
4093

. tab sex;

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |       1823       53.37       53.37
     female |       1593       46.63      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab sex, nolabel;

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       1823       53.37       53.37
          2 |       1593       46.63      100.00
------------+-----------------------------------
      Total |       3416      100.00

. /*
> total number of males and females: need not to be the same as in the article
> reason: number of employed may differ in the two random samples
> */
>
> /*
> generate a dummy for male
> */
>
> gen MALE = sex == 1;

.  /*
> 8. age
> generate dummies for each age group
> */
>
> tab age;
```

```
        age group |      Freq.     Percent        Cum.
------------------+-----------------------------------
      15-16 years |         62        1.81        1.81
      17-19 years |        222        6.50        8.31
      20-24 years |        501       14.67       22.98
      25-34 years |       1027       30.06       53.04
      35-44 years |        810       23.71       76.76
      45-54 years |        480       14.05       90.81
      55-64 years |        287        8.40       99.21
      65-69 years |         17        0.50       99.71
70 years and over |         10        0.29      100.00
------------------+-----------------------------------
            Total |       3416      100.00

. tab age, nolabel;

  age group |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         62        1.81        1.81
          2 |        222        6.50        8.31
          3 |        501       14.67       22.98
          4 |       1027       30.06       53.04
          5 |        810       23.71       76.76
          6 |        480       14.05       90.81
          7 |        287        8.40       99.21
          8 |         17        0.50       99.71
          9 |         10        0.29      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab age, gen (dage);

        age group |      Freq.     Percent        Cum.
------------------+-----------------------------------
      15-16 years |         62        1.81        1.81
      17-19 years |        222        6.50        8.31
      20-24 years |        501       14.67       22.98
      25-34 years |       1027       30.06       53.04
      35-44 years |        810       23.71       76.76
      45-54 years |        480       14.05       90.81
      55-64 years |        287        8.40       99.21
      65-69 years |         17        0.50       99.71
70 years and over |         10        0.29      100.00
------------------+-----------------------------------
            Total |       3416      100.00

. /*
> create new age group dummies as in the paper + labelling
> */
>
> gen A2534 = dage4;

. gen A3544 = dage5;

. gen A4554 = dage6;

. gen A55 = dage7 == 1 | dage8 == 1 | dage9 == 1;

. /*
> check: A55, works?
> */
>
```

```
> tab age A55;

                  |          A55
        age group |         0          1 |     Total
------------------+----------------------+----------
      15-16 years |        62          0 |        62
      17-19 years |       222          0 |       222
      20-24 years |       501          0 |       501
      25-34 years |      1027          0 |      1027
      35-44 years |       810          0 |       810
      45-54 years |       480          0 |       480
      55-64 years |         0        287 |       287
      65-69 years |         0         17 |        17
70 years and over |         0         10 |        10
------------------+----------------------+----------
            Total |      3102        314 |      3416


. label var A2534 "age 25 to 34 = 1; other = 0";

. label var A3544 "age 35 to 44 = 1; other = 0";

. label var A4554 "age 45 to 54 = 1; other = 0";

. label var A55 "age 55 and over = 1; other = 0";

. /*
> 6. marstat
> generate a dummy called MARRD + labelling
> */
>
> tab marstat;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
    married |       2281       66.77       66.77
     single |        892       26.11       92.89
      other |        243        7.11      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab marstat, nolabel;

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       2281       66.77       66.77
          2 |        892       26.11       92.89
          3 |        243        7.11      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab marstat, gen(dmarsta);

    marital |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
    married |       2281       66.77       66.77
     single |        892       26.11       92.89
      other |        243        7.11      100.00
------------+-----------------------------------
      Total |       3416      100.00
```

```
. gen MARRD = dmarsta1 == 1;

. label var MARRD "married = 1; other = 0";

. /*
> 9. educ
> generate dummies for different levels of education attained + labelling
> */
>
> tab educ;

                           education |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
                  none or elementary |        381       11.15       11.15
                         high school |       1798       52.63       63.79
                 some post-secondary |        310        9.07       72.86
post-secondary certificate or diploma |        486       14.23       87.09
                   university degree |        441       12.91      100.00
-------------------------------------+-----------------------------------
                               Total |       3416      100.00

. tab educ, gen(deduc);

                           education |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
                  none or elementary |        381       11.15       11.15
                         high school |       1798       52.63       63.79
                 some post-secondary |        310        9.07       72.86
post-secondary certificate or diploma |        486       14.23       87.09
                   university degree |        441       12.91      100.00
-------------------------------------+-----------------------------------
                               Total |       3416      100.00

. gen EDHS = deduc2;

. gen EDSPS = deduc3;

. gen EDDIP = deduc4;

. gen EDDEG = deduc5;

. label var EDHS "high school completion = 1; other = 0";

. label var EDSPS "some post-secondary education = 1; other = 0";

. label var EDDIP "post-secondary diploma = 1; other = 0";

. label var EDDEG "university degree = 1; other = 0";

. /*
> 35. typjob
> generate a dummy called PTIME + labelling
> */
>
> tab typjob;

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
  full-time |       2827       82.76       82.76
  part-time |        589       17.24      100.00
------------+-----------------------------------
      Total |       3416      100.00
```

```
.  tab typjob, nolabel;

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       2827       82.76       82.76
          2 |        589       17.24      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab typjob, gen(dtypjob);

type of job |      Freq.     Percent        Cum.
------------+-----------------------------------
  full-time |       2827       82.76       82.76
  part-time |        589       17.24      100.00
------------+-----------------------------------
      Total |       3416      100.00

. gen PTIME = dtypjob2 == 1;

. label var PTIME "part-time worker = 1; full-time worker = 0";

. /*
> 66. q13_20
> generate a dummy called UNION + labelling
> */
>
> tab q13_20;

member of a |
   union or |
group which |
    bargain |
collectivel |
          y |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       1252       36.65       36.65
         no |       2164       63.35      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab q13_20, nolabel;

member of a |
   union or |
group which |
    bargain |
collectivel |
          y |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |       1252       36.65       36.65
          2 |       2164       63.35      100.00
------------+-----------------------------------
      Total |       3416      100.00

. tab q13_20, gen(dunion);

member of a |
   union or |
group which |
    bargain |
collectivel |
```

```
          y |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |       1252       36.65       36.65
         no |       2164       63.35      100.00
------------+-----------------------------------
      Total |       3416      100.00

. gen UNION = dunion1 == 1;

. label var UNION "union member = 1; non-union member = 0";

. /*
> 38. ind52
> generate industry dummies
> MINES, MFG, CONS, TRANS, TRADE, COMM, PUBLIC + labelling
> */
>
> tab ind52;

                       recoded industry |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
                            agriculture |         52        1.52        1.52
                               forestry |         30        0.88        2.40
                   fishing and trapping |          7        0.20        2.61
                            metal mines |         17        0.50        3.10
                          mineral fuels |         26        0.76        3.86
                        non-metal mines |         13        0.38        4.24
                quarries and sand pits |          2        0.06        4.30
          services incidental to mining |         21        0.61        4.92
           food and beverage industries |        101        2.96        7.87
                       tobacco products |          1        0.03        7.90
            rubber and plastic products |         23        0.67        8.58
                     leather industries |         10        0.29        8.87
                     textile industries |         14        0.41        9.28
                    clothing industries |         25        0.73       10.01
                        wood industries |         51        1.49       11.50
      furniture and fixture industries |         16        0.47       11.97
            paper and allied industries |         45        1.32       13.29
printing-publishing and allied industri |         39        1.14       14.43
               primary metal industries |         32        0.94       15.37
           metal fabricating industries |         36        1.05       16.42
                   machinery industries |         21        0.61       17.04
    transportation equipment industries |         60        1.76       18.79
         electrical products industries |         40        1.17       19.96
non-metallic mineral product industries |         20        0.59       20.55
 petroleum and coal products industries |          5        0.15       20.70
chemical and chemical products industri |         22        0.64       21.34
 miscellaneous manufacturing industries |         16        0.47       21.81
                    general contractors |         67        1.96       23.77
              special-trade contractors |         76        2.22       26.00
                         transportation |        147        4.30       30.30
                                storage |          7        0.20       30.50
                          communication |         86        2.52       33.02
 electric power,gas and water utilities |         42        1.23       34.25
                        wholesale trade |        147        4.30       38.55
                           retail trade |        468       13.70       52.25
                     finance industries |         85        2.49       54.74
                     insurance carriers |         32        0.94       55.68
insurance agencies and real estate indu |         41        1.20       56.88
         education and related services |        300        8.78       65.66
            health and welfare services |        371       10.86       76.52
                religious organizations |         26        0.76       77.28
      amusement and recreation services |         33        0.97       78.25
```

services to business management	114	3.34	81.59
personal services	53	1.55	83.14
accommodation and food services	223	6.53	89.67
miscellaneous services	50	1.46	91.13
federal administration	110	3.22	94.35
provincial administration	126	3.69	98.04
local administration	67	1.96	100.00
Total	3416	100.00	

```
. tab ind52, nolabel;
```

recoded industry	Freq.	Percent	Cum.
1	52	1.52	1.52
2	30	0.88	2.40
3	7	0.20	2.61
4	17	0.50	3.10
5	26	0.76	3.86
6	13	0.38	4.24
7	2	0.06	4.30
8	21	0.61	4.92
9	101	2.96	7.87
10	1	0.03	7.90
11	23	0.67	8.58
12	10	0.29	8.87
13	14	0.41	9.28
15	25	0.73	10.01
16	51	1.49	11.50
17	16	0.47	11.97
18	45	1.32	13.29
19	39	1.14	14.43
20	32	0.94	15.37
21	36	1.05	16.42
22	21	0.61	17.04
23	60	1.76	18.79
24	40	1.17	19.96
25	20	0.59	20.55
26	5	0.15	20.70
27	22	0.64	21.34
28	16	0.47	21.81
29	67	1.96	23.77
30	76	2.22	26.00
31	147	4.30	30.30
32	7	0.20	30.50
33	86	2.52	33.02
34	42	1.23	34.25
35	147	4.30	38.55
36	468	13.70	52.25
37	85	2.49	54.74
38	32	0.94	55.68
39	41	1.20	56.88
40	300	8.78	65.66
41	371	10.86	76.52
42	26	0.76	77.28
43	33	0.97	78.25
44	114	3.34	81.59
45	53	1.55	83.14
46	223	6.53	89.67
47	50	1.46	91.13
48	110	3.22	94.35
49	126	3.69	98.04

```
         50 |         67        1.96      100.00
------------+-----------------------------------
      Total |       3416      100.00
```

```
. tab ind52, gen(ind);

                      recoded industry |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
                           agriculture |         52        1.52        1.52
                              forestry |         30        0.88        2.40
                  fishing and trapping |          7        0.20        2.61
                           metal mines |         17        0.50        3.10
                         mineral fuels |         26        0.76        3.86
                       non-metal mines |         13        0.38        4.24
                quarries and sand pits |          2        0.06        4.30
         services incidental to mining |         21        0.61        4.92
          food and beverage industries |        101        2.96        7.87
                      tobacco products |          1        0.03        7.90
          rubber and plastic products |         23        0.67        8.58
                    leather industries |         10        0.29        8.87
                    textile industries |         14        0.41        9.28
                   clothing industries |         25        0.73       10.01
                       wood industries |         51        1.49       11.50
      furniture and fixture industries |         16        0.47       11.97
           paper and allied industries |         45        1.32       13.29
printing-publishing and allied industri |         39        1.14       14.43
              primary metal industries |         32        0.94       15.37
          metal fabricating industries |         36        1.05       16.42
                  machinery industries |         21        0.61       17.04
    transportation equipment industries |         60        1.76       18.79
        electrical products industries |         40        1.17       19.96
non-metallic mineral product industries |         20        0.59       20.55
 petroleum and coal products industries |          5        0.15       20.70
chemical and chemical products industri |         22        0.64       21.34
 miscellaneous manufacturing industries |         16        0.47       21.81
                   general contractors |         67        1.96       23.77
              special-trade contractors |         76        2.22       26.00
                        transportation |        147        4.30       30.30
                               storage |          7        0.20       30.50
                         communication |         86        2.52       33.02
 electric power,gas and water utilities |         42        1.23       34.25
                       wholesale trade |        147        4.30       38.55
                          retail trade |        468       13.70       52.25
                    finance industries |         85        2.49       54.74
                    insurance carriers |         32        0.94       55.68
insurance agencies and real estate indu |         41        1.20       56.88
          education and related services |        300        8.78       65.66
           health and welfare services |        371       10.86       76.52
               religious organizations |         26        0.76       77.28
      amusement and recreation services |         33        0.97       78.25
        services to business management |        114        3.34       81.59
                     personal services |         53        1.55       83.14
        accommodation and food services |        223        6.53       89.67
                miscellaneous services |         50        1.46       91.13
               federal administration |        110        3.22       94.35
            provincial administration |        126        3.69       98.04
                  local administration |         67        1.96      100.00
---------------------------------------+-----------------------------------
                                 Total |       3416      100.00

. gen MINES = ind4 == 1 | ind5 == 1 | ind6 == 1 | ind7 == 1 | ind8 == 1;
```

```
.  gen MFG = ind9 == 1 | ind10 == 1 | ind11 == 1 | ind12 == 1 | ind13 == 1 |
ind14 == 1 |
>  ind15 == 1 | ind16 == 1 | ind17 == 1 | ind18 == 1 | ind19 == 1 | ind20 == 1
| ind21 == 1 |
>  ind22 == 1 | ind23 == 1 | ind24 == 1 | ind25 == 1 | ind26 == 1 | ind27 == 1
;

.  gen CONS = ind28 == 1 | ind29 == 1  ;

. gen TRANS = ind30 == 1 | ind31 == 1 | ind32 == 1 | ind33 == 1 ;

. gen TRADE = ind34 == 1 | ind35 == 1 | ind36 == 1 | ind37 == 1 | ind38 == 1 ;

. gen COMM = ind39 == 1 | ind40 == 1 | ind41 == 1 | ind42 == 1 | ind43 == 1 |
ind44 == 1 |
>  ind45 == 1 | ind46 == 1 ;

. gen PUBLIC = ind47 == 1 | ind48 == 1 | ind49 == 1 ;

. label var MINES "mines, quarries, and oil wells = 1; other = 0";

. label var MFG "manufacturing = 1; other = 0";

. label var CONS "construction = 1; other = 0";

. label var TRANS "transportation, communication, and utilities = 1; other =
0";

. label var TRADE "wholesale and retail trade, finance, insurance, and real
estate = 1; other = 0";

. label var COMM "community, business & personal service = 1; other = 0";

. label var PUBLIC "public administration = 1; other = 0";

. /*
> check: generating the right dummies?
> */
>
> tab ind52 MINES;

                      | mines, quarries, and
                      | oil wells = 1; other
                      |          = 0
     recoded industry |         0          1 |     Total
----------------------+----------------------+----------
          agriculture |        52          0 |        52
             forestry |        30          0 |        30
 fishing and trapping |         7          0 |         7
          metal mines |         0         17 |        17
        mineral fuels |         0         26 |        26
      non-metal mines |         0         13 |        13
quarries and sand pit |         0          2 |         2
services incidental t |         0         21 |        21
food and beverage ind |       101          0 |       101
     tobacco products |         1          0 |         1
rubber and plastic pr |        23          0 |        23
   leather industries |        10          0 |        10
   textile industries |        14          0 |        14
  clothing industries |        25          0 |        25
      wood industries |        51          0 |        51
furniture and fixture |        16          0 |        16
paper and allied indu |        45          0 |        45
```

```
printing-publishing a |        39          0 |        39
primary metal industr |        32          0 |        32
metal fabricating ind |        36          0 |        36
 machinery industries |        21          0 |        21
transportation equipm |        60          0 |        60
electrical products i |        40          0 |        40
non-metallic mineral  |        20          0 |        20
petroleum and coal pr |         5          0 |         5
chemical and chemical |        22          0 |        22
miscellaneous manufac |        16          0 |        16
  general contractors |        67          0 |        67
special-trade contrac |        76          0 |        76
       transportation |       147          0 |       147
              storage |         7          0 |         7
        communication |        86          0 |        86
electric power,gas an |        42          0 |        42
      wholesale trade |       147          0 |       147
         retail trade |       468          0 |       468
   finance industries |        85          0 |        85
   insurance carriers |        32          0 |        32
insurance agencies an |        41          0 |        41
education and related |       300          0 |       300
health and welfare se |       371          0 |       371
religious organizatio |        26          0 |        26
amusement and recreat |        33          0 |        33
services to business  |       114          0 |       114
    personal services |        53          0 |        53
accommodation and foo |       223          0 |       223
miscellaneous service |        50          0 |        50
federal administratio |       110          0 |       110
provincial administra |       126          0 |       126
 local administration |        67          0 |        67
----------------------+----------------------+----------
                Total |      3337         79 |      3416

.  tab ind52 MFG;

                      |  manufacturing = 1;
                      |       other = 0
     recoded industry |         0          1 |     Total
----------------------+----------------------+----------
          agriculture |        52          0 |        52
             forestry |        30          0 |        30
 fishing and trapping |         7          0 |         7
          metal mines |        17          0 |        17
        mineral fuels |        26          0 |        26
      non-metal mines |        13          0 |        13
quarries and sand pit |         2          0 |         2
services incidental t |        21          0 |        21
food and beverage ind |         0        101 |       101
     tobacco products |         0          1 |         1
rubber and plastic pr |         0         23 |        23
   leather industries |         0         10 |        10
   textile industries |         0         14 |        14
  clothing industries |         0         25 |        25
      wood industries |         0         51 |        51
furniture and fixture |         0         16 |        16
paper and allied indu |         0         45 |        45
printing-publishing a |         0         39 |        39
primary metal industr |         0         32 |        32
metal fabricating ind |         0         36 |        36
 machinery industries |         0         21 |        21
```

transportation equipm	0	60	60
electrical products i	0	40	40
non-metallic mineral	0	20	20
petroleum and coal pr	0	5	5
chemical and chemical	0	22	22
miscellaneous manufac	0	16	16
general contractors	67	0	67
special-trade contrac	76	0	76
transportation	147	0	147
storage	7	0	7
communication	86	0	86
electric power,gas an	42	0	42
wholesale trade	147	0	147
retail trade	468	0	468
finance industries	85	0	85
insurance carriers	32	0	32
insurance agencies an	41	0	41
education and related	300	0	300
health and welfare se	371	0	371
religious organizatio	26	0	26
amusement and recreat	33	0	33
services to business	114	0	114
personal services	53	0	53
accommodation and foo	223	0	223
miscellaneous service	50	0	50
federal administratio	110	0	110
provincial administra	126	0	126
local administration	67	0	67
Total	2839	577	3416

. tab ind52 CONS;

	construction = 1; other = 0		
recoded industry	0	1	Total
agriculture	52	0	52
forestry	30	0	30
fishing and trapping	7	0	7
metal mines	17	0	17
mineral fuels	26	0	26
non-metal mines	13	0	13
quarries and sand pit	2	0	2
services incidental t	21	0	21
food and beverage ind	101	0	101
tobacco products	1	0	1
rubber and plastic pr	23	0	23
leather industries	10	0	10
textile industries	14	0	14
clothing industries	25	0	25
wood industries	51	0	51
furniture and fixture	16	0	16
paper and allied indu	45	0	45
printing-publishing a	39	0	39
primary metal industr	32	0	32
metal fabricating ind	36	0	36
machinery industries	21	0	21
transportation equipm	60	0	60
electrical products i	40	0	40
non-metallic mineral	20	0	20
petroleum and coal pr	5	0	5

chemical and chemical	22	0	22
miscellaneous manufac	16	0	16
general contractors	0	67	67
special-trade contrac	0	76	76
transportation	147	0	147
storage	7	0	7
communication	86	0	86
electric power,gas an	42	0	42
wholesale trade	147	0	147
retail trade	468	0	468
finance industries	85	0	85
insurance carriers	32	0	32
insurance agencies an	41	0	41
education and related	300	0	300
health and welfare se	371	0	371
religious organizatio	26	0	26
amusement and recreat	33	0	33
services to business	114	0	114
personal services	53	0	53
accommodation and foo	223	0	223
miscellaneous service	50	0	50
federal administratio	110	0	110
provincial administra	126	0	126
local administration	67	0	67
Total	3273	143	3416

. tab ind52 TRANS;

	transportation, communication, and utilities = 1; other = 0		
recoded industry	0	1	Total
agriculture	52	0	52
forestry	30	0	30
fishing and trapping	7	0	7
metal mines	17	0	17
mineral fuels	26	0	26
non-metal mines	13	0	13
quarries and sand pit	2	0	2
services incidental t	21	0	21
food and beverage ind	101	0	101
tobacco products	1	0	1
rubber and plastic pr	23	0	23
leather industries	10	0	10
textile industries	14	0	14
clothing industries	25	0	25
wood industries	51	0	51
furniture and fixture	16	0	16
paper and allied indu	45	0	45
printing-publishing a	39	0	39
primary metal industr	32	0	32
metal fabricating ind	36	0	36
machinery industries	21	0	21
transportation equipm	60	0	60
electrical products i	40	0	40
non-metallic mineral	20	0	20
petroleum and coal pr	5	0	5
chemical and chemical	22	0	22
miscellaneous manufac	16	0	16

general contractors	67	0	67
special-trade contrac	76	0	76
transportation	0	147	147
storage	0	7	7
communication	0	86	86
electric power,gas an	0	42	42
wholesale trade	147	0	147
retail trade	468	0	468
finance industries	85	0	85
insurance carriers	32	0	32
insurance agencies an	41	0	41
education and related	300	0	300
health and welfare se	371	0	371
religious organizatio	26	0	26
amusement and recreat	33	0	33
services to business	114	0	114
personal services	53	0	53
accommodation and foo	223	0	223
miscellaneous service	50	0	50
federal administratio	110	0	110
provincial administra	126	0	126
local administration	67	0	67
Total	3134	282	3416

. tab ind52 TRADE;

	wholesale and retail trade, finance, insurance, and real estate = 1; other = 0		
recoded industry	0	1	Total
agriculture	52	0	52
forestry	30	0	30
fishing and trapping	7	0	7
metal mines	17	0	17
mineral fuels	26	0	26
non-metal mines	13	0	13
quarries and sand pit	2	0	2
services incidental t	21	0	21
food and beverage ind	101	0	101
tobacco products	1	0	1
rubber and plastic pr	23	0	23
leather industries	10	0	10
textile industries	14	0	14
clothing industries	25	0	25
wood industries	51	0	51
furniture and fixture	16	0	16
paper and allied indu	45	0	45
printing-publishing a	39	0	39
primary metal industr	32	0	32
metal fabricating ind	36	0	36
machinery industries	21	0	21
transportation equipm	60	0	60
electrical products i	40	0	40
non-metallic mineral	20	0	20
petroleum and coal pr	5	0	5
chemical and chemical	22	0	22
miscellaneous manufac	16	0	16
general contractors	67	0	67
special-trade contrac	76	0	76

transportation	147	0	147
storage	7	0	7
communication	86	0	86
electric power,gas an	42	0	42
wholesale trade	0	147	147
retail trade	0	468	468
finance industries	0	85	85
insurance carriers	0	32	32
insurance agencies an	0	41	41
education and related	300	0	300
health and welfare se	371	0	371
religious organizatio	26	0	26
amusement and recreat	33	0	33
services to business	114	0	114
personal services	53	0	53
accommodation and foo	223	0	223
miscellaneous service	50	0	50
federal administratio	110	0	110
provincial administra	126	0	126
local administration	67	0	67
Total	2643	773	3416

. tab ind52 COMM;

	community, business & personal service = 1; other = 0		
recoded industry	0	1	Total
agriculture	52	0	52
forestry	30	0	30
fishing and trapping	7	0	7
metal mines	17	0	17
mineral fuels	26	0	26
non-metal mines	13	0	13
quarries and sand pit	2	0	2
services incidental t	21	0	21
food and beverage ind	101	0	101
tobacco products	1	0	1
rubber and plastic pr	23	0	23
leather industries	10	0	10
textile industries	14	0	14
clothing industries	25	0	25
wood industries	51	0	51
furniture and fixture	16	0	16
paper and allied indu	45	0	45
printing-publishing a	39	0	39
primary metal industr	32	0	32
metal fabricating ind	36	0	36
machinery industries	21	0	21
transportation equipm	60	0	60
electrical products i	40	0	40
non-metallic mineral	20	0	20
petroleum and coal pr	5	0	5
chemical and chemical	22	0	22
miscellaneous manufac	16	0	16
general contractors	67	0	67
special-trade contrac	76	0	76
transportation	147	0	147
storage	7	0	7
communication	86	0	86

```
electric power,gas an |        42          0 |        42 
     wholesale trade |       147          0 |       147 
        retail trade |       468          0 |       468 
  finance industries |        85          0 |        85 
  insurance carriers |        32          0 |        32 
insurance agencies an |        41          0 |        41 
education and related |         0        300 |       300 
health and welfare se |         0        371 |       371 
religious organizatio |         0         26 |        26 
amusement and recreat |         0         33 |        33 
services to business  |         0        114 |       114 
   personal services |         0         53 |        53 
accommodation and foo |         0        223 |       223 
miscellaneous service |         0         50 |        50 
federal administratio |       110          0 |       110 
provincial administra |       126          0 |       126 
 local administration |        67          0 |        67 
----------------------+----------------------+----------
               Total |      2246       1170 |      3416 


. tab ind52 PUBLIC;

                      | public administration
                      |    = 1; other = 0
     recoded industry |         0          1 |     Total
----------------------+----------------------+----------
          agriculture |        52          0 |        52 
             forestry |        30          0 |        30 
 fishing and trapping |         7          0 |         7 
          metal mines |        17          0 |        17 
        mineral fuels |        26          0 |        26 
      non-metal mines |        13          0 |        13 
quarries and sand pit |         2          0 |         2 
services incidental t |        21          0 |        21 
food and beverage ind |       101          0 |       101 
     tobacco products |         1          0 |         1 
rubber and plastic pr |        23          0 |        23 
   leather industries |        10          0 |        10 
   textile industries |        14          0 |        14 
  clothing industries |        25          0 |        25 
      wood industries |        51          0 |        51 
furniture and fixture |        16          0 |        16 
paper and allied indu |        45          0 |        45 
printing-publishing a |        39          0 |        39 
primary metal industr |        32          0 |        32 
metal fabricating ind |        36          0 |        36 
 machinery industries |        21          0 |        21 
transportation equipm |        60          0 |        60 
electrical products i |        40          0 |        40 
non-metallic mineral  |        20          0 |        20 
petroleum and coal pr |         5          0 |         5 
chemical and chemical |        22          0 |        22 
miscellaneous manufac |        16          0 |        16 
  general contractors |        67          0 |        67 
special-trade contrac |        76          0 |        76 
       transportation |       147          0 |       147 
              storage |         7          0 |         7 
        communication |        86          0 |        86 
electric power,gas an |        42          0 |        42 
      wholesale trade |       147          0 |       147 
         retail trade |       468          0 |       468 
   finance industries |        85          0 |        85 
```

```
   insurance carriers |        32          0 |        32
insurance agencies an |        41          0 |        41
education and related |       300          0 |       300
health and welfare se |       371          0 |       371
religious organizatio |        26          0 |        26
amusement and recreat |        33          0 |        33
services to business  |       114          0 |       114
    personal services |        53          0 |        53
accommodation and foo |       223          0 |       223
miscellaneous service |        50          0 |        50
federal administratio |         0        110 |       110
provincial administra |         0        126 |       126
 local administration |         0         67 |        67
----------------------+----------------------+----------
                Total |      3113        303 |      3416
```

```
. /*
> GENERATING INTERACTION VARIABLES
> */
>
> gen MA2534 = MALE*A2534;

. gen MA3544 = MALE*A3544;

. gen MA4554 = MALE*A4554;

. gen MA55 = MALE*A55;

. gen MEDHS = MALE*EDHS;

. gen MEDSPS = MALE*EDSPS;

. gen MEDDIP = MALE*EDDIP;

. gen MEDDEG = MALE*EDDEG;

. gen MMARRD = MALE*MARRD;

. gen MPTIME = MALE*PTIME;

. gen MUNION = MALE*UNION;

. gen MMINES = MALE*MINES;

. gen MMFG = MALE*MFG;

. gen MCONS = MALE*CONS;

. gen MTRANS = MALE*TRANS;

. gen MTRADE = MALE*TRADE;

. gen MCOMM = MALE*COMM;

. gen MPUBLIC = MALE*PUBLIC;

. /*
> SUMMARY STATISTICS
> */
>
```

```
> summarize WAGE A2534 A3544 A4554 A55 EDHS EDSPS EDDIP EDDEG MARRD PTIME UNION
MINES MFG CONS
>  TRANS TRADE COMM PUBLIC MALE MA2534 MA3544 MA4554 MA55 MEDHS MEDSPS MEDDIP
MEDDEG MMARRD MPTIME
>  MUNION MMINES MMFG MCONS MTRANS MTRADE MCOMM MPUBLIC;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    WAGE |    3416    986.0872    514.178         50       4500
   A2534 |    3416     .300644   .4586052          0          1
   A3544 |    3416    .2371194   .4253784          0          1
   A4554 |    3416    .1405152   .3475717          0          1
     A55 |    3416    .0919204   .2889558          0          1
    EDHS |    3416    .5263466   .4993785          0          1
   EDSPS |    3416    .0907494   .2872945          0          1
   EDDIP |    3416    .1422717   .3493797          0          1
   EDDEG |    3416    .1290984   .3353579          0          1
   MARRD |    3416      .66774   .4710926          0          1
   PTIME |    3416    .1724239   .3778038          0          1
   UNION |    3416    .3665105   .4819217          0          1
   MINES |    3416    .0231265   .1503271          0          1
     MFG |    3416     .168911   .3747281          0          1
    CONS |    3416    .0418618   .2003027          0          1
   TRANS |    3416    .0825527   .2752452          0          1
   TRADE |    3416    .2262881    .418489          0          1
    COMM |    3416    .3425059   .4746173          0          1
  PUBLIC |    3416    .0887002   .2843522          0          1
    MALE |    3416    .5336651   .4989384          0          1
  MA2534 |    3416     .161007   .3675912          0          1
  MA3544 |    3416    .1311475   .3376111          0          1
  MA4554 |    3416     .074356   .2623877          0          1
    MA55 |    3416    .0541569   .2263602          0          1
   MEDHS |    3416    .2754684   .4468154          0          1
  MEDSPS |    3416    .0471311   .2119504          0          1
  MEDDIP |    3416    .0611827   .2397002          0          1
  MEDDEG |    3416     .074356   .2623877          0          1
  MMARRD |    3416    .3744145   .4840422          0          1
  MPTIME |    3416    .0444965   .2062256          0          1
  MUNION |    3416    .2093091    .406875          0          1
  MMINES |    3416      .02137   .1446356          0          1
    MMFG |    3416    .1229508   .3284288          0          1
   MCONS |    3416    .0374707   .1899401          0          1
  MTRANS |    3416    .0626464   .2423613          0          1
  MTRADE |    3416    .1012881   .3017539          0          1
   MCOMM |    3416    .1165105   .3208831          0          1
 MPUBLIC |    3416    .0521077   .2222768          0          1

. summarize WAGE A2534 A3544 A4554 A55 EDHS EDSPS EDDIP EDDEG MARRD PTIME UNION
>  MINES MFG CONS TRANS TRADE COMM PUBLIC if MALE == 1;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    WAGE |    1823    1108.318   527.2083         50       4324
   A2534 |    1823    .3017005   .4591219          0          1
   A3544 |    1823    .2457488   .4306484          0          1
   A4554 |    1823    .1393308   .3463864          0          1
     A55 |    1823    .1014811   .3020475          0          1
    EDHS |    1823    .5161821   .4998752          0          1
   EDSPS |    1823     .088316   .2838317          0          1
   EDDIP |    1823    .1146462   .3186819          0          1
   EDDEG |    1823    .1393308   .3463864          0          1
   MARRD |    1823    .7015908   .4576856          0          1
   PTIME |    1823     .083379   .2765302          0          1
```

```
   UNION |    1823    .3922106   .4883772          0          1
   MINES |    1823    .0400439    .196116          0          1
     MFG |    1823    .2303895   .4211977          0          1
    CONS |    1823    .0702139   .2555773          0          1
   TRANS |    1823    .1173889   .3219715          0          1
   TRADE |    1823     .189797    .392248          0          1
    COMM |    1823    .2183214   .4132201          0          1
  PUBLIC |    1823    .0976413   .2969104          0          1

. summarize WAGE A2534 A3544 A4554 A55 EDHS EDSPS EDDIP EDDEG MARRD PTIME UNION
>  MINES MFG CONS TRANS TRADE COMM PUBLIC if MALE == 0;

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    WAGE |    1593    846.2084   460.7912         70       4500
   A2534 |    1593     .299435   .4581544          0          1
   A3544 |    1593    .2272442   .4191832          0          1
   A4554 |    1593    .1418707    .349027          0          1
     A55 |    1593    .0809793    .272889          0          1
    EDHS |    1593    .5379787   .4987121          0          1
   EDSPS |    1593    .0935342    .291271          0          1
   EDDIP |    1593    .1738858   .3791302          0          1
   EDDEG |    1593    .1173886   .3219838          0          1
   MARRD |    1593    .6290019   .4832237          0          1
   PTIME |    1593    .2743252   .4463137          0          1
   UNION |    1593    .3370998   .4728677          0          1
   MINES |    1593    .0037665   .0612752          0          1
     MFG |    1593    .0985562   .2981588          0          1
    CONS |    1593    .0094162   .0966095          0          1
   TRANS |    1593    .0426868   .2022134          0          1
   TRADE |    1593    .2680477   .4430817          0          1
    COMM |    1593    .4846202   .4999203          0          1
  PUBLIC |    1593    .0784683   .2689915          0          1

. /*
```

> **ESIMATE A POOLED (INTERACTIVE) REGRESSION FUNCTION, WITH FEMALES AS THE BASE GROUP**

```
> */

. /*
> aside: agriculture, forestry, and fishery is the base group for industries
> */
>
> regress WAGE A2534 A3544 A4554 A55 EDHS EDSPS EDDIP EDDEG MARRD PTIME UNION
MINES MFG CONS
>  TRANS TRADE COMM PUBLIC MALE MA2534 MA3544 MA4554 MA55 MEDHS MEDSPS MEDDIP
MEDDEG MMARRD MPTIME
>  MUNION MMINES MMFG MCONS MTRANS MTRADE MCOMM MPUBLIC;

  Source |       SS       df       MS               Number of obs =    3416
---------+------------------------------            F( 37,  3378) =   69.67
   Model |  390774355    37  10561469.0             Prob > F      =  0.0000
Residual |  512080063  3378  151592.677             R-squared     =  0.4328
---------+------------------------------            Adj R-squared =  0.4266
   Total |  902854418  3415  264379.039             Root MSE      =  389.35

------------------------------------------------------------------------------
    WAGE |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   A2534 |   151.7977     28.781      5.274   0.000       95.36772    208.2276
   A3544 |   258.8091   31.34382      8.257   0.000       197.3543    320.2638
   A4554 |   245.3292   35.04772      7.000   0.000       176.6123    314.0461
```

```
         A55 |   198.9053   42.14581      4.719   0.000      116.2714    281.5392
        EDHS |   152.1056   40.27967      3.776   0.000      73.13062    231.0806
       EDSPS |     275.65   50.42427      5.467   0.000      176.7848    374.5151
       EDDIP |   398.0278   45.59036      8.731   0.000      308.6404    487.4153
       EDDEG |    650.973   49.36549     13.187   0.000      554.1837    747.7622
       MARRD |   28.68707   22.32961      1.285   0.199     -15.09385      72.468
       PTIME |  -24.05411   22.86524     -1.052   0.293     -68.88522      20.777
       UNION |   235.9846   23.33026     10.115   0.000      190.2418    281.7275
       MINES |  -141.5771   179.1581     -0.790   0.429     -492.8464    209.6922
         MFG |  -329.3264   87.87035     -3.748   0.000     -501.6109    -157.042
        CONS |  -248.1974   129.8061     -1.912   0.056     -502.7039    6.309177
       TRANS |  -132.9791   95.11746     -1.398   0.162     -319.4728    53.51446
       TRADE |  -339.2768   83.86236     -4.046   0.000     -503.7029   -174.8507
        COMM |  -314.3265   83.29835     -3.774   0.000     -477.6468   -151.0062
      PUBLIC |  -268.7665   89.83932     -2.992   0.003     -444.9114   -92.62152
        MALE |    -226.69    107.683     -2.105   0.035     -437.8206   -15.55952
      MA2534 |   31.89238   42.39426      0.752   0.452     -51.22863    115.0134
      MA3544 |   60.12963   45.96213      1.308   0.191     -29.98678     150.246
      MA4554 |    120.634   51.45322      2.345   0.019      19.75143    221.5166
        MA55 |   109.4056   58.93722      1.856   0.063      -6.15059    224.9618
       MEDHS |   28.00126   49.52714      0.565   0.572     -69.10494    125.1075
      MEDSPS |   70.67227   64.92599      1.089   0.276     -56.62593    197.9705
      MEDDIP |  -25.63337   59.02053     -0.434   0.664     -141.3529    90.08619
      MEDDEG |    64.1716    61.9993      1.035   0.301     -57.38834    185.7315
      MMARRD |   66.68036   33.30172      2.002   0.045      1.386795    131.9739
      MPTIME |   -145.627   43.69666     -3.333   0.001     -231.3016   -59.95244
      MUNION |  -88.64374   30.96567     -2.863   0.004     -149.3571   -27.93039
      MMINES |   561.0432   191.2428      2.934   0.003      186.0799    936.0066
        MMFG |   459.7009   102.3161      4.493   0.000      259.0932    660.3085
       MCONS |   466.5005   142.7727      3.267   0.001      186.5709      746.43
      MTRANS |   372.9716   110.4725      3.376   0.001      156.3718    589.5714
      MTRADE |   343.0229   98.97202      3.466   0.001      148.9717     537.074
       MCOMM |   213.0908   98.86554      2.155   0.031      19.24849    406.9332
     MPUBLIC |   487.8536   107.0446      4.557   0.000      277.9748    697.7323
       _cons |   652.1583   91.79664      7.104   0.000      472.1757    832.1409
------------------------------------------------------------------------------

.
end of do-file
```

. ***HYPOTHESIS TESTING**

```
. test  MALE MA2534 MA3544 MA4554 MA55 MEDHS MEDSPS MEDDIP MEDDEG MMARRD MPTIME
>  MUNION MMINES MMFG MCONS MTRANS MTRADE MCOMM MPUBLIC

 ( 1)  MALE = 0.0
 ( 2)  MA2534 = 0.0
 ( 3)  MA3544 = 0.0
 ( 4)  MA4554 = 0.0
 ( 5)  MA55 = 0.0
 ( 6)  MEDHS = 0.0
 ( 7)  MEDSPS = 0.0
 ( 8)  MEDDIP = 0.0
 ( 9)  MEDDEG = 0.0
 (10)  MMARRD = 0.0
 (11)  MPTIME = 0.0
 (12)  MUNION = 0.0
 (13)  MMINES = 0.0
 (14)  MMFG = 0.0
 (15)  MCONS = 0.0
 (16)  MTRANS = 0.0
 (17)  MTRADE = 0.0
 (18)  MCOMM = 0.0
```

```
 (19)  MPUBLIC = 0.0

       F( 19,  3378) =   12.71
            Prob > F =    0.0000


. test MA2534 MA3544 MA4554 MA55

 ( 1)  MA2534 = 0.0
 ( 2)  MA3544 = 0.0
 ( 3)  MA4554 = 0.0
 ( 4)  MA55 = 0.0

       F(  4,  3378) =    1.84
            Prob > F =    0.1188


. test  MEDHS MEDSPS MEDDIP MEDDEG

 ( 1)  MEDHS = 0.0
 ( 2)  MEDSPS = 0.0
 ( 3)  MEDDIP = 0.0
 ( 4)  MEDDEG = 0.0

       F(  4,  3378) =    1.08
            Prob > F =    0.3660


. test MMINES MMFG MCONS MTRANS MTRADE MCOMM MPUBLIC

 ( 1)  MMINES = 0.0
 ( 2)  MMFG = 0.0
 ( 3)  MCONS = 0.0
 ( 4)  MTRANS = 0.0
 ( 5)  MTRADE = 0.0
 ( 6)  MCOMM = 0.0
 ( 7)  MPUBLIC = 0.0

       F(  7,  3378) =    8.64
            Prob > F =    0.0000


-> . save, replace
file sum84_14.dta saved
```

Labor Income and the Disabled: A Comment on Jon Harkness' Article on the Labor Force Participation by Disabled Males in Canada

ECON 452A
____

and

ECON 452B
____

Winter, 2001

Source: QED HALS Survey, Dataset 41
Harkness, Jon. 1993. "Labor Force Participation by Disabled Males in Canada," *Canadian Journal of Economics* (November): 878-89.

1. INTRODUCTION

In his article entitled "Labor Force Participation by Disabled Males in Canada," Harkness (1993) argues that disability benefits discourage labor force participation by the disabled prime-age males in Canada.[1] This is despite the fact that many of the disabled are still employable. Only one quarter of Canadian disabled people report that they are completely unable to work.[2] Furthermore, Harkness finds that, other than health reasons, psychological, social, and economic factors also play a role in the decision of the disabled to work or not to work. Therefore, aside from a complete disability, participation in the labor force is still a matter of choice. It depends on the expected labor income of a disabled individual, his or her expected disability pension, labor-leisure substitution, and other non-wage income opportunities that affect his or her economic well-being. Based on his study on a sample of 6,892 disabled prime-age males in Canada, Harkness concludes that Canadian disability-related insurance schemes discourage the disabled from participating in the job market.

Yet the theme that disability benefits have work disincentive effects remains controversial. For example, Haveman and Wolfe (1984) contend that, in the U.S., the secular decline in labor force participation by older males cannot be adequately explained by the increases in Social Security's disability benefits.[3] Their finding is consistent with that of Bound (1989), who finds that the disincentive effects of such disability insurance have been overstated.[4]

---

[1] Harkness, Jon. 1993. "Labor Force Participation by Disabled Males in Canada," *Canadian Journal of Economics* (November): 878-89.
[2] 13 percent or 1.6 million of working-age Canadian males experienced disability in 1986.
[3] Haveman, R. H. and B. L. Wolfe. 1984. "Disability Transfers and Early Retirement: A Causal Relationship?" *Journal of Public Economics* (24): 47-66.
[4] Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants," *American Economic Review* (79): 482-503.

The purpose of this paper is to explore further the disincentive case for Canadian disability-related insurance schemes. We try to replicate some of the results of Harkness using the Health and Activity Limitation Survey conducted in 1991, as opposed to the 1986 survey used in Harkness' study. The paper is organized as follows. In Section 2, the models and results in Harkness' paper are examined in detail. In Section 3, an attempt is made to replicate Harkness' regression results for expected labor income, which is used in his logit estimate of labor force participation. We are interested in whether expected labor income differs significantly by occupations. Section 4 reports the empirical results. Conclusions are presented in Section 5.

2. MODELS AND RESULTS IN HARKNESS' PAPER

Harkness uses the Health and Activity Limitation Survey (HALS) conducted by Statistics Canada between June and October 1986. In his sample, 6,892 prime-age males are not completely prevented from working. They are also not self-employed and confined to health institutions, jails, and penitentiaries. Harkness models the decision of labor force participation using the models below:

$$Y = \alpha\,[AGE, OCC_1\text{-}OCC_6, YED, MAR, HWY] + u_1 \quad \text{for } Y > 0 \qquad (1)$$

$$D = \beta\,[DIS, AGE, PQ, PRIV, INC, AY, FY] + u_2 \quad \text{for } D > 0 \qquad (2)$$

$$P = \chi\,[DIS, AGE, MAR, PQ, E[Y], E[D], SW, NSE, WE, NWK, MS, NP, PCH, ETH, OWN, FY, AY] + v \qquad (3)$$

where α, β, χ are matrices of parameters. The descriptions of the variables used in each of the above three regression equations are reported in Table 1. The equation (1)

Table 1: Variable Definitions

Variable	Definition	Mean
	Continuous Variables	
Y	Labour income	$11,152.80
D	Pension income	$4,765.59
AY	Asset income	$1,945.91
FY	Rest-of-family income	$20,858.42
DIS	Disability Status	0.065
AGE	Age in years	38.21
YED	Years of formal schooling	10.22
HWY	Hours worked per year	417.44
NP	Number of people in household	3.45
	Dummy Variables	
P	Labour force participation	0.445
MAR	Location = MARITIMES	0.209
PQ	Location = QUEBEC	0.125
OCCi	Occupation	
	1. Managers	0.025
	2. Professionals	0.082
	3. Semi-profs & technicians	0.039
	4. Supervisors & foremen	0.020
	5. Clerical, sales & service workers	0.334
	6. Skilled craftsmen & tradesmen	0.007
PCH	Presence of Children	0.47
MS	Married (including comomnlaw)	0.67
OWN	Home Owener	0.56
WE	Work experience (no work in last 3 years)	0.34
ETH	Ethnicity (not British nor French)	0.34
NWK	Never worked	0.107
NSE	No suitable employment	0.104
SW	Sheltered workshop	0.013
PRIV	Private pension	0.469

Source: Harkness (p.885)

estimates the income of disabled and employed workers. This is the model estimated in Section 3 in which an attempt is made to replicate Harkness' results. The second equation estimates the disabled pension income for workers who collect disability pensions[5]. In equation (3), since the dependent variable is not continuous, probit is used to analyze the determinants of a choice between working or not working. In other words, it models the decision of labor force participation. The dependent variable, *P*, is equal to one if a disabled person works and zero otherwise. Presumably, the decision to participate in the labor force depends on the expected utilities received from labor income, U_w, and those resulting from collecting a disability pension and not working, U_{nw}. Therefore, $P = 1$ if $U_w > U_{nw}$, and $P = 0$ if $U_w < U_{nw}$. For this reason, the equations (1) and (2) are used to compute the expected labor income, *E[Y]*, and the expected disability pension income, *E[D]*. The results are then used in regression equation (3). Finally, u_1, u_2, v are random error terms.

The Ordinary Least Square (OLS) estimation is used to calculate the two income equations, (1) and (2). The first income equation that estimates the labor income of disabled working people who are not collecting pension is analyzed in this paper. Harkness finds a negative relationship between pension and participation. However, the incentive from the extra income is almost three times higher than the disincentive that the pension provides. The pension elasticity of labour force participation is –2.03, and the income elasticity of participation is 6.33 (pp. 885, 886). The result implies that the common practice of reducing one's disability income one to one for every dollar earned in the market is a work disincentive.

Table 2 reports the logit estimates of equation (3) for selected variables. Residency in Quebec does not have an additional affect on participation rate, implying that Quebec Pension Plan and Canada Pension Plan do not affect potential workers'

---

[5] Part time workers who collect disability pensions was eliminated from the first income equation, and were included in the other set of the sample which used the second equation for the estimation of its income.

Table 2: Logit Estimate of Male Labour Force Participation (Selected Variables)

Dependent Variable		P=	
Independent Variable	Definition	Coefficient	T-value
MAR	Location = MARITIMES	-0.148	-1.48 [c]
PQ	Location = QUEBEC	0.010	0.07
FY	Rest-of-family income	0.758×10^{-5}	-1.88 [b]
AY	Asset income	0.557×10^{-4}	-2.38 [a]
OWN	Home Owener	-0.185	-2.06 [a]
WE	No work in last 3 years	-1.862	-20.9 [a]
NSE	No suitable employment	-2.149	-8.08 [a]

Source and Notes: Harkness (p. 887)
[a] Significant at the 99 per cent level
[b] Significant at the 95 per cent level
[c] Significant at the 90 per cent level on a one-tail test

choice of participation differently. However, Maritime residency has a negative effect on participation, which is to be expected because of the physical nature of work in Atlantic Provinces, such as fishery. A negative relationship between wealth (rest of family income) and participation rate also exists. Similarly, the same negative relationship holds between asset income (or home ownership) and participation rate. Finally, it shows that if one has not worked for several years, or never, participation decreases.

3. Replicating Harkness' Results

In this section, we evaluate the robustness of Harkness' results for his labor income equation (1) using the same type of survey in a different year. The QED archive provides the Health and Activity Limitation Survey (HALS) conducted by Statistics Canada in 1991. Harkness' article uses the survey conducted in 1986. For ease of illustration, the regression equation for labor income of the disabled is reproduced below:

Table 3: Summary of the Original Variables from HALS Archive

Code	agegrp	eeocc91	hlosr	newprov	hours	empin
1	15-34	Senior Managers	less than grade 5	9 = NF, PEI	Continuous	less than 0
2	35-54	Middle and Other Managers	grades 5-8	12 = NS	98,99 =	0
3	55-64	Professionals	grades 9-13	13 = NB	missiong values	-999
4	65+	Semi-professionals	secondary grad	24 = QC		-2,999
5		Supervisors/Clerical	trades certificate	35 = ON		-4,999
6		Foremen	non-univ w/o cert.	46 = MN		-6,999
7		Administration/Service	non-univ w trade cert.	47 = SAS		-9,999
8		Sales and Service	non-univ w diploma	48 = ALT		-14,999
9		Skilled Craftsmen	univ w/o cert/diploma	59 = BC		-19,999
10		Clearical Workers	univ w cert.	60 = Terr.		-24,999
11		Sales and Service	univ w bachelor			-29,999
12		Semi-skilled men	univ w diploma			-34,999
13		Sales and Service	masters			35000+
14		Other Manual Workers	Ph.D.			

$$Y = \alpha\,[AGE, OCC_1\text{-}OCC_6, YED, MAR, HWY] + u_1 \qquad \text{for } Y > 0 \qquad (1)$$

The equation includes occupation dummies (*OCC*i), location dummy(*MAR* – Maritime), age dummies(*AGE*), and dummy variables for educational attainment (*YED*). In addition, the hours worked per year (*HWY*) would obviously contribute to the size of one's income.

Table 3 summarizes the variable that we have used to replicate the regression, and Table 4 reports the manipulations we have made to each variable. Harkness lists age (*AGE*), years of formal schooling (*YED*), and employment income (*Y*) as continuous variables, while agegrp, hlosr, and empin are all categorical variables, as Table 3 shows. Therefore, we assign the mean values of each category to the variable, as Table 4 illustrates. For agegrp, 70 is assigned to 65 or older, and 40,000 is assigned to $35,000 or larger for empin.

Table 4: The Manipulations of the Corresponding Data from the HALS Survey

Harkness' Variable	Description	Corresponding Variable from HALS Dataset	Manipulations
Y	Labour Income	empin	1,2 = 0, 3 = 500
			4 = 2000, 5 = 4000, 6 = 6000 7 = 8500 8 = 12500 9 = 17500 10 = 22500
			11 = 27500
			12 = 32500
			13 = 40000
AGE	Age in years	agegrp	1 = 25, 2 = 45
			3 = 60, 4 = 70
OCC_1	1. Managers	eeocc91	1 = 1,2
OCC_2	2. Professionals		2 = 3
OCC_3	3. Semi-profs & technicians		3 = 4
OCC_4	4. Supervisors & foremen		4 = 5,6
OCC_5	5. Clerical, sales & service workers		5 = 7,8,10,11,13
OCC_6	6. Skilled craftsmen & tradesmen		6 = 9,12
OCC_7	(Reference Dummy)		7 = 14
YED	Years of formal schooling	Hlosr [a]	1 = 3.001
			2 = 7.001
			3 = 11.001
			4 = 12.501
			5 = 13.001
			6 = 14.001
			7,8,9 = 15.001
			10 = 16.001
			11 = 17.501
			12 = 17.001
			13 = 19.001
			14 = 24.001
MAR	Location = MARITIME	newprov	9,12,13 = MAR
HWY	Hours worked per year	hours	hours * 52 [b]

Sources and Notes: HALS dataset, Harkness (1993)

[a] Recoding category number to any whole number between 1 to 14 would cause a problem because it would be a different category number, and will eventually be modified by subsequent recoding command. Therefore, 0.001 was added to each assigned number to eliminate the problem.

[b] No adjustment was made to correct the missing values of 98 and 99 because the numbers were already treated as missing values in the original dataset. The maximum value was 66.

Harkness also categorizes occupation into seven different brackets including the reference. This is different from the census, which reports 14 different categories. Therefore, as Table 4 summarizes, categories 1 through 13 are newly assigned to the 6 categories specified by Harkness, and 14, the manual labor category, is assigned to the seventh reference category. In addition, we categorize the provinces of NF, NB, PEI, and NS together to create the dummy variable for *MAR*, the Maritime Provinces. Finally, hours worked are weekly hours and therefore they are multiplied by 52 in order to estimate annual working hours. It is worth emphasizing that, for the present purpose, the assumption of 52 weeks per year does not change our basic results in the study.

Since Harkness only includes males who are physically able to work after being disabled and not getting any pension, an effort is made to eliminate females, people who cannot work, and pension collectors. However, because the labels for pension collectors and people who cannot work are vaguely defined in the HALS survey, we have to include them in our regression. With this in mind, we turn to the next section on empirical testing.

4. Empirical Results

Table 5 reports our regression results using the Health and Activity Limitation Survey in 1991 and those of Harkness based on the 1986 survey. There are two major discrepancies to explain. First, the results are not identical in numbers. Second, the sign of the coefficient of OCC_5 is different, while the signs for all other coefficients are identical. The Harkness estimate of the coefficient is –655.04, whereas we obtain an

Table 5: The Comparison of Regression Results of Harkness and Kim & Yu [a]

	Harkness	Kim & Yu
Constant	-6,065.45	-9,337.13
	(6.73) [b]	(-14.82) [b]
AGE	131.19	286.39
	(7.81) [b]	(38.33) [b]
MAR	-567.33	-2,851.22
	(1.64)	(-13.70) [b]
OCC_1	4,293.12	8,723.77
	(6.11) [b]	(20.62) [b]
OCC_2	5,029.17	8,421.10
	(10.22) [b]	(19.13) [b]
OCC_3	1,044.67	6,598.17
	(1.70) [b]	(13.58) [b]
OCC_4	2,863.05	4,576.27
	(3.41) [b]	(9.60) [b]
OCC_5	-655.04	1,667.87
	(1.93) [b]	(4.54) [b]
OCC_6	2,551.16	4,729.87
	(1.87) [b]	(12.90) [b]
YED	648.76	785.95
	(12.72) [b]	(32.48) [b]
HWY	5.19	4.44
	(32.62) [b]	(31.96) [b]

Sources and Notes: Harkness (p.886)
[a] T-values in parentheses
[b] Significant at the 95 per cent level or more

estimate of 1667.87. According to Harkness' result, clerical, sales, and service workers have less income than the reference group on average. Our estimate suggests that workers in clerical, sales, and service sector have more income than workers in the manual labor sector.

The first point to explain is the general difference in numerical values; that is, the numbers are not identical. The differences may be due to the fact that we use an entirely different sample from Harkness' sample. In addition, as explained in the previous section, we estimate values in order to assign to different categories and that could have been different from the exact values Harkness used. Finally, Harkness eliminates pension

Table 6: Difference between Kim & Yu and Harkness Estimates of the OCC_i Coefficients

	Harkness (1)	Kim & Yu (2)	Difference (2) – (1)	Percentage Difference (%)
OCC_1	4,293.12	8,723.77	4,430.65	103.20
OCC_2	5,029.17	8,421.10	3,391.93	67.45
OCC_3	1,044.67	6,598.17	5,553.50	531.60
OCC_4	2,863.05	4,576.27	1,713.22	59.84
OCC_5	-655.04	1,667.87	2,322.91	354.62
OCC_6	2,551.16	4,729.87	2,178.71	85.40

collectors and people who could not work from the regression sample, whereas we include them in our sample. Our estimate of the OCC_5 coefficient is different from that of Harkness, not only in magnitude but also in their signs. Furthermore, the estimates of the OCC_i, $\forall i = 1,...,6$, coefficients are greater than Harkness' estimates, as Table 6 indicates. This implies that the reference group in our sample has much less income than Harkness' reference group. Therefore, our assumption that only manual labor category of occupation is in the reference group of dummy variables seems to have been different from Harkness' categorization of reference group of dummy variables. To test whether expected labor income differs significantly by occupations, we conduct an F-test of the joint significance of the coefficient estimates of the OCC_i, $\forall i = 1,...,6$. The sample value of the F-statistic is 149.67, which is higher than the critical values of the F[6, 20855]-distribution for the 1%, 5%, and 10% significance levels. Therefore, we have to reject the null hypothesis that expected labor income is the same for all occupations.

5. CONCLUSIONS

Harkness' deterrent argument that disability benefits discourage the disabled from working is consistent with the findings of Slade (1984).[6] In particular, Harkness found that a disability pension is a significant but not an important work deterrent. Our empirical results reaffirm Harkness' findings in one important aspect: expected labor income is not the same for all occupations of the disabled prime-age males in Canada.

The replication for Harkness' regression of labor income equation using the 1991 HALS data has resulted in qualitatively similar coefficient estimates of all the explanatory variables, except for the coefficient estimate of OCC_5. Three factors may be responsible for these numerical discrepancies between the analyses. First, we used 1991 sample while Harkness used 1986 sample. Second, we estimated the values to assign to each category of variables such as labor income, age, and years of schooling, in order to change categorical variables into continuous variables. Third, we retained pension collectors and the disabled who are physically unable to work in our sample. In addition, the large differences between regression coefficients of the occupation dummies may have resulted from our preference in selecting a different reference group from that of Harkness. Overall, Table 5 illustrates that two regression results are coherent when considering the aforementioned differences in analysis.

As a concluding remark, Samuel Johnson's notion that "a decent provision for the poor is a true test of civilization" perhaps also accords to the disabled. The disability benefits may serve as a means of discouraging the disabled from working at unsuitable jobs, increasing the labor market's efficiency. Therefore, our social policies should weigh the societal gains from this increased labor-market efficiency against the deadweight losses resulting from a disability pension based on the incentive arguments in economics.

[6] Slade, F. P. 1984. "Older Men: Disability Insurance and the Incentives to Work," *Industrial Relations*.

APPENDIX: Stata LOG File

```
/* STATA hw1.log file for Econ 452 Project 3 */


-> . Qmerge empin hlosr hours marstl, ds(41)
QEDid     QEDmerge  empin     hlosr     hours     marstl
 Was observation in memory, data set |
                        41, or both? |      Freq.     Percent
Cum.
-------------------------------------+-----------------------------------
obs. from both master and using data |      84845      100.00
100.00
-------------------------------------+-----------------------------------

                               Total |      84845      100.00

-> . recode empin 1=-500 2=0 3=500 4=2000 5=4000 6=6000 7=8500 8=12500
9=17500 10=22500 11=27500 12=32500 13=40000
(80108 changes made)

-> . Qmerge e74x i3, ds(41)
QEDid     QEDmerge  e74x      i3
 Was observation in memory, data set |
                        41, or both? |      Freq.     Percent
Cum.
-------------------------------------+-----------------------------------
obs. from both master and using data |      84845      100.00
100.00
-------------------------------------+-----------------------------------
                               Total |      84845      100.00
-> . recode agegrp 1=25 2=45 3=60 4=70

-> . tab eeocc91, gen(occ)

      employment |
          equity |
      occupation |
  groups  (1991) |      Freq.     Percent        Cum.
-----------------+-----------------------------------
 senior managers |        444        0.71        0.71
middle and other |       4533        7.22        7.93
professionals (s |       7464       11.89       19.81
semi-profess and |       3774        6.01       25.82
supervs -clercl/ |        868        1.38       27.21
 sup -man, proc, |       1832        2.92       30.12
admin. and sr. c |       3631        5.78       35.91
sales and servic |       2642        4.21       40.11
skilled crafts a |       5146        8.20       48.31
clerical workers |       7125       11.35       59.66
sales and servic |       8134       12.95       72.61
semi-skilled man |       6918       11.02       83.63
sales and servic |       7121       11.34       94.97
other manual wor |       3158        5.03      100.00
-----------------+-----------------------------------
           Total |      62790      100.00
```

```
-> . tab newprov, gen(mar)

        province |      Freq.     Percent        Cum.
-----------------+-----------------------------------
newfoundland / p |       9935       11.71       11.71
     nova scotia |       6779        7.99       19.70
   new brunswick |       4043        4.77       24.46
          quebec |       7388        8.71       33.17
         ontario |      17877       21.07       54.24
        manitoba |       6286        7.41       61.65
    saskatchewan |       7464        8.80       70.45
         alberta |      10641       12.54       82.99
british columbia |       6497        7.66       90.65
yukon / northwes |       7935        9.35      100.00
-----------------+-----------------------------------
           Total |      84845      100.00


-> . recode hlosr 1=3.001
(2101 changes made)

-> . recode hlosr 2=7.001 3=11.001 4=12.501 5=13.001 6=14.001 7=15.001
8=15.001 9=15.001 10=16.001 11=17.501 12=17.001 13=19.001 14=24.001
(77854 changes made)


-> . regress empin agegrp occq1 occq2 occq3 occq4 occq5 occq6 hlosr marq
hwy if sex==1

  Source |       SS       df       MS                  Number of obs =
20866
---------+------------------------------               F( 10, 20855) =
664.79
   Model |  1.0274e    10  1.0274e               Prob > F      =  0.0000
Residual |  3.2231e 20855   154546318               R-squared     =
0.2417
---------+------------------------------               Adj R-squared =
0.2414
   Total |  4.2505e 20865   203712945               Root MSE      =
12432

------------------------------------------------------------------------
------
   empin |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+--------------------------------------------------------------
-----
  agegrp |   286.3862   7.472253     38.327   0.000         271.74
301.0324
   occq1 |   8723.772   423.0897     20.619   0.000       7894.483
9553.06
   occq2 |   8421.098   440.2592     19.128   0.000       7558.156
9284.04
   occq3 |   6598.173   485.8573     13.580   0.000       5645.855
7550.491
```

```
   occq4 |   4576.269   476.6986      9.600   0.000       3641.903
5510.636
   occq5 |   1667.866   367.4819      4.539   0.000       947.5731
2388.159
   occq6 |    4729.87   366.8125     12.895   0.000       4010.889
5448.851
   hlosr |   785.9481   32.48009     24.198   0.000       722.2846
849.6116
    marq |  -2851.217   208.0358    -13.705   0.000      -3258.984   -
2443.451
     hwy |   4.444829    .139093     31.956   0.000       4.172196
4.717462
   _cons |  -9337.131   629.8988    -14.823   0.000      -10571.78   -
8102.481
------------------------------------------------------------------------
------

-> . test occq1 occq2 occq3 occq4 occq5 occq6


 ( 1)  occq1 = 0.0
 ( 2)  occq2 = 0.0
 ( 3)  occq3 = 0.0
 ( 4)  occq4 = 0.0
 ( 5)  occq5 = 0.0
 ( 6)  occq6 = 0.0

       F(  6, 20855) =  149.67
            Prob > F =    0.0000


/* this is the end of the hw1.log file */
```

Estimating the Union Effect Using Regression Analysis

and

“Survey of Union Membership” (374)

“Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men”
Francis Vella and Marno Verbeek
Journal of Applied Econometrics, Volume 13, Issue 2

INTRODUCTION

Vella and Verbeek estimate the union effect for men over a period of declining unionization. The data was taken from the National Longitudinal Survey, which is comprised of a sample of full-time working males who completed their schooling by 1980. These males were followed over the period 1980 to 1987. The sample consists of 545 observations. Union membership is based on the question reflecting whether or not the individual had his wage set in a collective bargaining agreement. The goal is to estimate the average increase in wages resulting from union employment. We will use a similar model to that of Verbeek and Vella to estimate the difference in workers' wages in union and non-union employment. This is known as the union effect.

Vella and Verbeek set out to answer three questions. First, what is the impact of unions on wages and how does it vary by worker characteristics? Second, which are the primary forms of worker heterogeneity generating the endogeneity of union status? Finally, with what form of economic sorting behaviour, in terms of union and non-union employment, are the data consistent? It is assumed that individuals locate in union or non-union employment on the basis of wages. Observed and unobserved characteristics and their associated prices determine these. The regression equation takes the form

$$w_{j,it} = \beta_{j,t}X_{it} + \alpha_{j,t} + \varepsilon_{j,it} \qquad t = 1,\ldots,T; \qquad i = 1,\ldots N$$

where $w_{j,it}$ represents the (potential) wage of individual i in sector j (j = 0,1) in time period t, where j = 1 corresponds to the union sector; β is an unknown parameter vector; and X_{it} is a vector of characteristics, including time dummy variables. The α and the ε represent the unobserved random components of the individual's wage. The vector of characteristics includes the variables years of schooling, experience, wage set by collective agreement, marital status, black, hispanic, has health disability, lives in rural area, lives in North East, lives in Northern Central, lives in South, log of hourly wage, hourly wage, and a union differential. There are dummy variables for the type of industry and occupation that the individual is in.

Vella and Verbeek reported many findings. It was found that the union effect is approximately 21 percent. Many of the explanatory variables had a statistically significant impact on the probability of union membership. The time effects displayed an increasingly negative pattern consistent with the data, which indicate sizable decreases in unionization over this period. The coefficients on the time dummies indicate that the time effect on union membership is negative. An estimate of 0.611 for the coefficient on lagged union status indicates a substantial degree of positive state dependence. The estimate for σ^2 of 0.57 indicates that 57 percent of the total variance is due to across individual variation. The coefficients on the dummy variables denoting whether the individual is black or hispanic are both positive and statistically significant, which may be due to the fact that these groups choose to bargain through union membership rather than on an individual basis. The dummy variables used to capture occupational status indicate that it does appear to influence the probability of union membership. Individuals in the blue-collar industries display a higher probability to acquire union membership.

Our paper will be organized as follows. We will begin with a description of our data set and model. This will be followed by a statement of our results and finally a conclusion.

DATA

We will consider a similar model, however our data will differ slightly. Our data was extracted from the QED Data Archive and is entitled Survey on Union Membership, reference number 374. This survey was conducted in 1984. The data set is considerably larger than that used in Vella and Verbeek's paper. The Survey on Union Membership considers 84,676 individuals. Since we are only using male observations, our data set will consist of 34,093 individuals. We extracted the variables wage, province, marital status, age, education, occupation, tenure, pension plans, sex, and union status from this data set. Refer to Table 1 for summary statistics and a description of the data.

The version of the regression equation in matrix notation that we will use is:

$$w_i = \beta_i X_i + \varepsilon_i \qquad i = 1, \ldots N$$

where w_i represents the (potential) wage of individual i; β is an unknown parameter vector; and X_i is a vector of characteristics. The ε represents the unobserved random components of the individual's wage. We will only consider males from our data sample to ensure that our results will resemble those of Vella and Verbeek.

RESULTS

After using Qextract to extract the variables from the data set, we decided that it was necessary to construct dummy variables for sex and union status in order to separate males and females, and union and non-union members. The variable DSEX1 identifies a male individual and the variable DUNION1 identifies a union member. This enables us to look at the effect on wages of males who participate in unions. We then generated a variable for the natural logarithm of the hourly wage rate as our dependent variable. We also constructed dummy variables for the different levels of education that the individuals may possess. DEDUC1 represents little or no education. DEDUC2 represents an individual who completed high school. DEDUC3 represents an individual who has some secondary education. DEDUC4 represents an individual who has a secondary school certificate or diploma. DEDUC5 represents an individual who has a university degree. It was also decided that we would generate an interaction term, UN_SEX = DSEX1*DUNION1, which is the effect on the wage rate of a male individual who is active in a union. Thus, our final regression equation is:

$$\ln w_i = \beta_1 + \beta_2 PROV_i + \beta_3 MARSTAT_i + \beta_4 AGE_i + \beta_5 DEDUC1_i + \beta_6 DEDUC2_i + \beta_7 DEDUC3_i + \beta_8 DEDUC4_i + \beta_9 DEDUC5_i + \beta_{10} OCC_i + \beta_{11} TENURE_i + \beta_{12} PENSION_i + \beta_{13} DSEX1_i + \beta_{14} DUNION1_i + \beta_{15} UN\_SEX_i + \varepsilon_i \qquad \text{for } i = 1 \ldots N$$

We then ran this regression in STATA, which provided for some interesting results. The value for the coefficient on union status was 0.21. This corresponds to a union effect of 21 percent, which is the same value that Vella and Verbeek found in their study. Although the data sets differed, both studies found a similar union effect, which proves that this result is significant. Table 2 illustrates a complete listing of the coefficients and their respective standard errors for the regression.

Many of the coefficients had differing effects on the hourly wage rate. Marital status, occupation, pension plans, and the interaction term for males and union participation all have a negative effect on the hourly wage rate. However, although these coefficients are negative their values are considerably small. The values of the coefficients for marital status, occupation, pension plans and the interaction term are -0.078, -0.002, -0.025, -0.073 respectively. Therefore, occupation has the least impact on the hourly wage rate of males. The dummy variables for level of education have the most significant effect on the dependent variable. As the education level increases, the effect on wage rate increases as well. For instance, a university degree has a 45 percent increase effect on wages, while a secondary school certificate or diploma increases the hourly wage rate by 33 percent. Males who have some secondary school education will experience a wage increase of 24 percent while males who have only completed high school will have a wage increase of 15 percent. The coefficient on the dummy variable for males was 0.267. This demonstrates the fact that the hourly wage rate for males will be 27 percent higher than that of females. The coefficient on the dummy variable for union participation was 0.212. Therefore, males who are union members have a wage that is 21 percent higher than males who are not union members. The value on our interaction term was –0.073. This tells us that the increase in the hourly wage rate of males in a union is less than that for females who are not in a union. The region or province in which the male union members reside has little effect on their hourly wage rate, as the value of this coefficient is only 0.004.

We then decided to conduct an F-test to test the joint significance of the coefficients. The null and alternative hypotheses are:

$$H_o: \beta_i = 0 \qquad \text{for all } i = 2 \ldots 14$$

$$H_A: \beta_i \neq 0 \qquad \text{for all } i = 2 \ldots 14$$

This test reported an F-statistic of 2191.5 and a p-value of zero. The critical values recorded for the one percent, five percent and ten percent significance values were 2.13, 1.72 and 1.52 respectively. Since the F-statistic is greater than the critical value and the p-value is less than the significance level in each case, we reject the null hypothesis at all three levels of significance. We reject the null that the coefficients are jointly equal to zero and thus, our coefficients are significant and therefore valid. We also tested the significance of each coefficient separately and found the same results.

We tested the marginal effect of the variable DUNION1 on the regressand WAGE by taking the partial derivative of the regression function with respect to $DUNION1_i$. The null and alternative hypotheses for this test are:

$$H_o: \beta_i = 0 \qquad \text{for all } i = 12, 14$$

$$H_A: \beta_i \neq 0 \qquad \text{for all } i = 12, 14$$

We used the test command in STATA to perform a joint F-test of the two coefficient restrictions specified by the null hypothesis. This test generated an F-statistic of 576.51 and a p-value of zero. We found the one percent, five percent and ten percent critical values to be 4.606, 2.996, and 2.303 respectively. We reject the null hypothesis that the marginal effect of $DUNION1_i$ equals zero and therefore the effect is significant on the hourly wage rate.

We then tested the marginal effect of the variable DSEX1 on the regressand WAGE by taking the partial derivative of the regression function with respect to $DSEX1_i$. The null and alternative hypotheses for this test are:

$$H_o: \beta_i = 0 \qquad \text{for all } i = 13, 14$$

$$H_A: \beta_i \neq 0 \qquad \text{for all } i = 13, 14$$

We used the test command in STATA to perform a joint F-test of the two coefficient restrictions specified by the null hypothesis. This test generated an F-statistic of 1387.65 and a p-value of zero. The critical values are the same as those of the test conducted above. We reject the null hypothesis that the marginal effect of $DSEX1_i$ equals zero and therefore the effect is significant on the hourly wage rate as well.

CONCLUSION

By comparing our results with those of Vella and Verbeek and by testing the significance of the values of our coefficients, we can be fairly confident that the regression we performed was significant. The regression showed that the effects of being a male and part of a union have a positive effect on the hourly wage rate. We also found the union effect to be 21 percent, which is the same as the one found by Vella and Verbeek in their regression analysis. Vella and Verbeek's paper was quite straightforward and thorough in explaining the model and the analysis used. The fact that our results were similar allowed for a clearer analysis of the regression we performed.

Table 1. Descriptive Statistics

Variable	Definition	Mean	Standard Deviation
Prov	Region and province	33.19	(15.94)
Sex	Sex	1.52	(00.49)
Marstat	Marital Status	1.47	(00.68)
Age	Age group	5.08	(02.13)
Educ	Education	2.29	(01.16)
Occ	Occupation	32.18	(15.49)
Tenure	Job tenure	3.44	(01.55)
q13_20	Member of a union	1.68	(00.47)
q15_23	Covered by pension plan	1.62	(00.49)
dv1	Hourly wage rate	923.77	(510.32)
lnw	Log of wage	6.68	(00.55)
Sex dummies			
dsex1	Male	0.48	(00.50)
dsex2	Female	0.52	(00.50)
Union dummies			
dunion1	Union member	0.32	(00.47)
dunion2	Non-union member	0.68	(00.47)
Education dummies			
deduc1	None or elementary	0.23	(00.42)
deduc2	High School	0.51	(00.50)
deduc3	Some post-secondary	0.08	(00.27)
deduc4	Post-secondary certificate or diploma	0.11	(00.31)
deduc5	University degree	0.08	(00.27)

Table 2. Coefficient estimates and standard errors

Variable	**Estimate**	**Standard Error**
Constant	6.351	(0.091)
Prov	0.004	(0.000)
Marstat	-0.078	(0.004)
Age	0.051	(0.002)
Occ	-0.002	(0.000)
Tenure	0.523	(0.002)
q15_23	-0.249	(0.005)
Sex dummy		
dsex1	0.267	(0.006)
Education Dummies		
deduc1	(dropped)	
deduc2	0.147	(0.007)
deduc3	0.237	(0.010)
deduc4	0.327	(0.010)
deduc5	0.455	(0.010)
Union dummy		
dunion1	0.212	(0.007)
Interaction Term		
un_sex	-0.073	(0.009)
Number of Observations	34093	

Appendix

1. Log File

```
-> . Qextract
getting information about file 374 ...
loading variables from 374 (sum84) only (no data yet)...  done
-> . browse
-> . tab sex, nolabel

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      40420       47.73       47.73
          2 |      44256       52.27      100.00
------------+-----------------------------------
      Total |      84676      100.00
-> . tab sex, gen(dsex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      40420       47.73       47.73
     female |      44256       52.27      100.00
------------+-----------------------------------
      Total |      84676      100.00
-> . tab q13_20, nolabel

member of a |
   union or |
group which |
    bargain |
collectivel |
          y |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      14350       32.49       32.49
          2 |      29816       67.51      100.00
------------+-----------------------------------
      Total |      44166      100.00
-> . tab q13_20, gen(dunion)

member of a |
   union or |
group which |
    bargain |
collectivel |
          y |      Freq.     Percent        Cum.
------------+-----------------------------------
        yes |      14350       32.49       32.49
         no |      29816       67.51      100.00
------------+-----------------------------------
      Total |      44166      100.00
-> . gen lnw = ln(dv1)
(40510 missing values generated)
-> . tab educ, nolabel
```

```
  education |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      19504       23.03       23.03
          2 |      42826       50.58       73.61
          3 |       6720        7.94       81.55
          4 |       8931       10.55       92.09
          5 |       6695        7.91      100.00
------------+-----------------------------------
      Total |      84676      100.00
-> . tab educ, gen(deduc)

                            education |      Freq.     Percent
Cum.
-------------------------------------+-----------------------------------
---
                  none or elementary |      19504       23.03
23.03
                         high school |      42826       50.58
73.61
                 some post-secondary |       6720        7.94
81.55
post-secondary certificate or diploma |       8931       10.55
92.09
                   university degree |       6695        7.91
100.00
-------------------------------------+-----------------------------------
---
                                Total |      84676      100.00
-> . gen un_sex = dsex1*dunion1
(40510 missing values generated)
-> . regress lnw prov marstat age deduc* occ tenure q15_23 dsex1
dunion1 un_sex

  Source |       SS       df       MS                  Number of obs =
34093
---------+------------------------------               F( 13, 34079) =
2191.50
   Model |  4415.58787    13  339.660605               Prob > F      =
0.0000
Residual |  5281.90757 34079  .154990099               R-squared     =
0.4553
---------+------------------------------               Adj R-squared =
0.4551
   Total |  9697.49544 34092  .284450764               Root MSE      =
.39369

-----------------------------------------------------------------------
    lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
    prov |   .0042486   .0001422     29.875   0.000       .0039699
.0045274
 marstat |  -.0779495   .0036069    -21.611   0.000      -.0850192   -
.0708798
     age |   .0513108   .0018141     28.284   0.000       .0477551
.0548666
  deduc1 |  (dropped)
```

```
  deduc2 |   .1467322   .0075076     19.544   0.000       .1320169    
.1614474
  deduc3 |   .2373696   .0099947     23.750   0.000       .2177798    
.2569595
  deduc4 |   .3268106   .0090788     35.997   0.000       .3090157    
.3446054
  deduc5 |   .4549919   .0099649     45.659   0.000       .4354603    
.4745234
     occ |  -.0019174   .0001901    -10.088   0.000      -.0022899   -
.0015449
  tenure |   .0527147   .0017693     29.795   0.000       .0492469    
.0561825
  q15_23 |  -.2485039   .0053801    -46.189   0.000      -.2590492   -
.2379587
   dsex1 |   .2665896   .0055795     47.780   0.000       .2556536    
.2775255
 dunion1 |   .2122263   .0072643     29.215   0.000        .197988    
.2264646
  un_sex |  -.0733089   .0090333     -8.115   0.000      -.0910144   -
.0556034
   _cons |   6.351122   .0191545    331.574   0.000       6.313579    
6.388666
----------------------------------------------------------------------
-------
-> . test prov marstat age deduc1 deduc2 deduc3 deduc4 deduc5 occ 
tenure q15_23 dunion1 dsex1 un_sex

 ( 1)  prov = 0.0
 ( 2)  marstat = 0.0
 ( 3)  age = 0.0
 ( 4)  deduc1 = 0.0
 ( 5)  deduc2 = 0.0
 ( 6)  deduc3 = 0.0
 ( 7)  deduc4 = 0.0
 ( 8)  deduc5 = 0.0
 ( 9)  occ = 0.0
 (10)  tenure = 0.0
 (11)  q15_23 = 0.0
 (12)  dunion1 = 0.0
 (13)  dsex1 = 0.0
 (14)  un_sex = 0.0
       Constraint 4 dropped

       F( 13, 34079) = 2191.50
            Prob > F =    0.0000

-> . display invfprob(13, 34079, 0.01)
2.1303892
-> . display invfprob(13, 34079, 0.05)
1.7204434
-> . display invfprob(13, 34079, 0.10)
1.5241913
-> . test prov

 ( 1)  prov = 0.0

       F(  1, 34079) =  892.51
```

```
             Prob > F =    0.0000

-> . display invfprob(1, 34079, 0.01)
6.6356659
-> . display invfprob(1, 34079, 0.05)
3.8417578
-> . display invfprob(1, 34079, 0.10)
2.7056932
-> . test marstat

 ( 1)  marstat = 0.0

       F(  1, 34079) =  467.04
            Prob > F =    0.0000

-> . display invfprob(1, 34079, 0.01)
6.6356659
-> . display invfprob(1, 34079, 0.05)
3.8417578
-> . display invfprob(1, 34079, 0.10)
2.7056932
-> . test age

 ( 1)  age = 0.0

       F(  1, 34079) =  799.99
            Prob > F =    0.0000

-> . test occ

 ( 1)  occ = 0.0

       F(  1, 34079) =  101.77
            Prob > F =    0.0000

-> . test tenure

 ( 1)  tenure = 0.0

       F(  1, 34079) =  887.72
            Prob > F =    0.0000

-> . test q15_23

 ( 1)  q15_23 = 0.0

       F(  1, 34079) = 2133.44
            Prob > F =    0.0000

-> . test deduc1

 ( 1)  deduc1 = 0.0
       Constraint 1 dropped
-> . test deduc2

 ( 1)  deduc2 = 0.0
```

```
       F(  1, 34079) =  381.98
            Prob > F =    0.0000

-> . test deduc3

 ( 1)  deduc3 = 0.0

       F(  1, 34079) =  564.04
            Prob > F =    0.0000

-> . test deduc4

 ( 1)  deduc4 = 0.0

       F(  1, 34079) = 1295.78
            Prob > F =    0.0000

-> . test deduc5

 ( 1)  deduc5 = 0.0

       F(  1, 34079) = 2084.78
            Prob > F =    0.0000

-> . test un_sex

 ( 1)  un_sex = 0.0

       F(  1, 34079) =   65.86
            Prob > F =    0.0000

-> . test dunion1 un_sex

 ( 1)  dunion1 = 0.0
 ( 2)  un_sex = 0.0

       F(  2, 34079) =  576.51
            Prob > F =    0.0000

-> . display invfprob(2, 34079, 0.01)
4.6057701
-> . display invfprob(2, 34079, 0.05)
2.9959977
-> . display invfprob(2, 34079, 0.10)
2.302736
-> . test dsex1 un_sex

 ( 1)  dsex1 = 0.0
 ( 2)  un_sex = 0.0

       F(  2, 34079) = 1387.65
            Prob > F =    0.0000


-> . summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
```

```
---------+-----------------------------------------------------
   QEDid |   84676     42338.5       24444          1      84676
    prov |   84676    33.18678     15.9384         10         59
     sex |   84676    1.522651    .4994896          1          2
 marstat |   84676    1.473357    .6830946          1          3
     age |   84676    5.082078    2.126991          1          9
    educ |   84676    2.297168    1.163998          1          5
     occ |   84676    32.18029    15.49968          1         51
  tenure |   43588    3.441957    1.549411          1          6
  q13_20 |   44166    1.675089    .4683467          1          2
  q15_23 |   44166    1.620477    .4852737          1          2
     dv1 |   44166    923.7682     510.324         13       4500
   dsex1 |   84676     .477349    .4994896          0          1
   dsex2 |   84676     .522651    .4994896          0          1
 dunion1 |   44166    .3249106    .4683467          0          1
 dunion2 |   44166    .6750894    .4683467          0          1
     lnw |   44166    6.684982    .5485619   2.564949   8.411833
  deduc1 |   84676    .2303368    .4210509          0          1
  deduc2 |   84676    .5057631    .4999697          0          1
  deduc3 |   84676    .0793613    .2703035          0          1
  deduc4 |   84676    .1054726    .3071633          0          1
  deduc5 |   84676    .0790661    .2698435          0          1
  un_sex |   44166    .1965539    .3973966          0          1

-> . save "D:\JBBS\proj2.dta", replace
file D:\JBBS\proj2.dta saved
-> . BREAK
************************************************************************
```

ESTIMATION OF THE ANNUAL EARNINGS EQUATION OF THE RETIREMENT MODEL

Friday March 23, 2001

ECONOMICS 452 PROJECT II.1

Canada Health Survey: 1977, Data set #12

Breslaw and Stelcner, "The Effect of Health on the Labor Force Behavior of Elderly Men in Canada," The Journal of Human Resources, 1987. Pages 490-517.

Contents:

INTRODUCTION

In their paper "The Effect of Health on the Labor Force Behavior of Elderly Men in Canada", Jon A. Breslaw and Morton Stelcner attempt to prove that health is just as important as potential earnings in determining the labor force behavior of elderly men in Canada. Annual earnings is arguably the most important determinant of the labor force behavior of individuals, and public policies intended to regulate the size and composition of the labor force reflect this importance. Using data from the Canada Health Survey in 1979, this study focuses on the role of health in influencing the labor market behavior of men aged 50 and older. “We try to answer the specific question: what is the effect of health on labor supply behavior?” (p.491)

The retirement model is used to determine the labor market behavior of elderly males. The retirement model is essentially characterized by various labor options equations of the following form:

$$\mathbf{PLMS_{ij} = f(X, H, E, I, R, O, \varepsilon)} \qquad \mathbf{(1)}$$

where: $PLMS_{ij}$ is a binary choice probability of being in labor state (i) over market state (j); X is a vector of personal and family characteristics; H is health status; E is potential full-time labor earnings; I is non-labor income; R is region of residence; O is occupation; and ε is a random error term.

The first step in this analysis is to generate a measure of the potential full-time labor earnings across the entire sample of males aged 50 and over. This is done by OLS estimation of the annual earnings equation, where the dependent variable is the natural logarithm of a measure of annual earnings:

$$\mathbf{lnY=\beta_1+REGION_i\beta_{i2}+AGEi\beta_{i3}+LANG_i\beta_{i4}+EDUC_i\beta_{i5}+OCCUP_i\beta_{i6}+HEALTH\beta_7+\varepsilon} \qquad \mathbf{(2)}$$

This estimated equation is then integrated into the retirement model as one of the labor options equations. It is this first step in the analysis that we will attempt to replicate in our paper.

DATA

Breslaw and Stelcner regressed region, age, language, education, occupation and health on the natural logarithm of annual earnings. Table 1 provides a list and description of each of the variables they used and the variables that we extracted from the QED data archive index. Breslaw and Stelcner selected a sample of 1,541 out of 31,668 observations. After retrieving all of the required variables of 31,668 observations, we began to narrow our sample size. First, we eliminated females and candidates less than 50 years of age from our sample. We then filtered out anyone not in the labor force or not employed, and reduced our sample size to males 50 years of age and older who are employed principle earners. Thus, we were left with 1,263 observations for men aged 50 and over.

We extracted the variables needed to construct the Annual Earnings Equation from the data in the Canada Health Survey. We were able to obtain data for all of the explanatory variables except health. “No single item covered in the CHS questionnaire suitably differentiates levels of health along the full length of the spectrum” (p.513). Breslaw and Stelcner used three different variables to define the independent variable health in equation (2). They used activity limitation, the number of chronic health problems, and the health opinion score. Then they ordered these variables on a scale of poorest health to most healthy. The variables were then weighted using an age and sex specific "Relative to an Identified Distribution" (ridit) method.

We could not replicate this independent variable, so we used the subjective variable “hlthscor”, which is a linear, numerical ranking of health by the respondents from the health opinion score. After researching the possibility of using activity limitation, chronic illness or an overall health variable, none of them seemed to be adequate for running regressions. There were either not enough observations, or there were too many separate variables (health problems) within. Although many other labor force behavior studies have used a similar variable to “hlthscor”, Breslaw and Stelcner discourage the use of this measure for a number of reasons. First, since this measure is self-assessed, it is subjective to the psychological state of each individual. Furthermore, the so-called average healthy person is a vague measure that is applicable irrespective of sex and age. Lastly, there is an error in the measurement problem since categorical variables such as healthy/fairly healthy/not healthy are used. This is an error because the underlying health status is a continuous variable that would require more than two or three states (p.500). Despite these concerns, we used the variable “hlthscor” because it is manageable for running regressions, and many other studies have used similar variables.

Our dependent variable is the natural log of income, where the variable we chose to use for income is called income. There were other variables we could have used as our dependent variable such as “indinc” and “efinc”. We decided not to use “indinc” because it was individual income from wages and salary for self-employed individuals. We want all individuals. Similarly, we discarded “efinc” because that was economic family income. Thus, we chose income, even though using the variable income may be problematic because it is income from all sources, not just wages and salary.

We have essentially replicated the variables Breslaw and Stelcner used, and their dummy variables where we had the data to do so. It should be mentioned that with the occupation variable, there were many components, so we recoded the variable to reduce the dummy variables within it from 12 to 3. Essentially we combined certain dummy variables to produce three of the dummy variables they used. We were unable to obtain a measure of people with no occupation or unknown occupation, so it was left out of our regression.

RESULTS

The results of the regression run by Breslaw and Stelcner are summarized in Table 2. It can be seen from Table 2 that region has a positive effect on income, and that all four t-statistics are significant. It is interesting to note that on average, income increases from east to west. It appears that as the population ages, there is increasing negative effects on income. The coefficients on the age variables are increasing in negativity from age 55 to age 70, which is expected, although the t statistics for ages 55 and 60 are not significant. The coefficients on the language variables suggest that on average, people who speak languages other than English make less income. However, the coefficient on the variable for French speaking individuals is not significant. From the regression data it is evident that on average, as education increases, so too does income. The data suggests that using professional occupations as a base group, on average, blue-collar workers earn more income than the other occupations, but earn less than the professional designation. The coefficient on the health variable suggests that on average, poor health has negative effects on income.

We use the "xi" command in stata to regress all of our explanatory variables on the natural logarithm of income. This command creates dummy variables for our variables keeping one group constant. Our regression equation takes the following form:

$$\mathbf{lnY=\beta_1+REGION_i\beta_{i2}+AGEi\beta_{i3}+LANG_i\beta_{i4}+EDUC_i\beta_{i5}+OCCUP_i\beta_{i6}+HEALTH\beta_7+\varepsilon} \qquad \mathbf{(3)}$$

Our regression yields very similar results to that of Breslaw and Stelcner. With a final sample of 719 observations, we test the null hypothesis that none of our explanatory variables effect the natural logarithm of income.

Out of the six explanatory variables we used, we retained the null for age group 55-60, "hlthscor" and both occupation dummies. The insignificance of the health variable we used can be explained by its subjective nature. The variable we used as a proxy for the manufactured variable found in the paper is a personal rating of an individual's perceived health and therefore may not accurately reflect the true health status of the individual. The occupation variable was also very subjective in nature. The original variable contained 12 individual observation categories divided by occupation. We narrowed down the categories to match as closely as possible those used by Breslaw and Stelcner. We were unable to find in survey data any measure for unknown occupation and thus we lack the fourth dummy used in the paper. It is possible that in the reconstruction of the occupation variable, we were unable to properly structure it in a way that mimicked the variable in the paper.

The balance of our analysis very closely reflected the findings in the paper. By using an “xi” regression in stata we regressed “lnincome” on each of the five variables requiring dummies and the variable “hlthscor”. Allowing Stata to generate the dummy variables for each variable, the subsequent regression used the first dummy variable from each set as the reference group. The output and command used to generate our results can be seen in appendix B; our log file. As a result, our regression was not only clear and simple to interpret, but also employed the same format as that run by Breslaw and Stelcner.

The results of the regression discussed below are all in reference to table 3 of our appendix. Our regression found that region has a significant effect on income. Clearly, as one moves west across Canada, income increases. One interesting result in our regression was that locations in Ontario, the Prairies and British Colombia seem to increase earnings by almost exactly the same proportion relative to the Maritimes. This mildly contradicts the results found in the paper that the increase is a steady trend from east to west. One of the major topics in this paper was age, and clearly our results show, as did the paper, that age is strongly, negatively related to earnings. This effect is significantly magnified in the 65-70 and 70+ category, as demonstrated by the jump in coefficients on our “age65” and “age70” dummy variables. An interesting result found both in our regression is the negative relationship between income and those whose primary language is not English. Those who speak French earn about 13% less than those whose primary language is English, but that result is almost double at 24% lower earnings for those who primarily speak other languages. As expected, further education has a strong positive relationship to income. It is interesting to note, that even for a group of men over 50, those with university degrees still earn on average 50% more than those with secondary education. As mentioned, our occupation variables are possibly biased. Not only does the regression suggest that they are insignificant, but the results seem to suggest that statistically, clerical/sales/service and blue-collar workers on average earn slightly more than professionals. This is a very questionable result and in final analysis should be discarded. Similarly, our "healthscor" variable, as described, is questionable. The nature of the variable makes it very subjective and the regression shows it has an insignificant effect on income.

SUMMARY

Our results turned out to be very similar to the results obtained in the paper. Although our results are similar, there were discrepancies in the methods we used to narrow down our sample size, as well as some of the variables we used. In the paper, Breslaw and Stelcner do not specify how they obtained the dummy variable "OCCUPN". We attempted to include all unemployed persons in this variable, but due to difficulties with Stata and in running the regression, we chose to omit this variable all together. It was significant in their regression, but not overly significant with a t-statistic of (2.40). The most notable difference in our variables was the variable we chose to use for health. As previously mentioned, our health variable is a subjective measure of self-assessment, but it was found to be insignificant in our regression. In the paper Breslaw and Stelcner used a modified health variable that is beyond our comprehension to replicate. They did not specify what measure of income they used for the dependent variable, so we used a variable that included income from all sources, for lack of a better variable.

In their results, it seems that income increases moving east to west across provinces, which we thought was interesting to note. However, our results suggest that relative income is consistently greater across Ontario, the Prairies and BC, but that income in Quebec is relatively lower than the Atlantic Provinces. It appears from the data in the paper that on average there is no significant difference between the income earned by anglophone and francophone workers, however, in our paper there is a significant difference. In both sets of results there is a significant decrease in income when individuals speak a language other than English or French. The results in the paper are intuitive, and for the most part were expected, our results produced minimal deviation from the results in the paper irrespective of differences in some of our variables and a discrepancy in sample size. They had a sample size of 610 observations and we had a sample size of 719. Our results are on par with what we expected they would be.

APPENDIX 1 - Tables

Table 1. *Definition of Variables*

Breslaw and Stelcner Variables	Description	Our Variables	Description
HEALTHRD	Ridit value of health status, age/sex corrected	**HLTHSCOR**	Health opinion score
LINCF	Percentile value of individual's fitted earnings	**INCOME**	Individual income from all sources
LOTHINC	Natural logarithm of economic family income less individual's annual earnings	**SEX**	Sex
Region	Reference group is Atlantic Provinces	**Region**	Region
REGIONQ **REGIONO** **REGIONP** **REGIONB**	If region is Quebec If region is Ontario If region is Prairies If region is British Columbia	**REGIONQ** **REGIONO** **REGIONP** **REGIONB**	If region is Quebec If region is Ontario If region is Prairies If region is British Columbia
Age	Reference group is age 50 to 54	**Agegrp**	Age Group
AGE55 **AGE60** **AGE65** **AGE70**	If age 55 to 59 If age 60 to 64 If age 65 to 69 If age is 70 or more	**AGE55** **AGE60** **AGE65** **AGE70**	If age 55 to 59 If age 60 to 64 If age 65 to 69 If age is 70 or more
Marital Status	Reference is single (never married)	**Marstat**	Marital Status
MARSTUSM **MARSTUSW** **MARSTUSD**	If married/common law If widowed If separated/divorced	**MARSTUSM** **MARSTUSW** **MARSTUSD**	If married/common law If widowed If separated/divorced
Family Size	Reference group is unattached individual	**Famsize**	Size of Family
FAMSIZE2 **FAMSIZE4** **FAMSIZE7**	If 2-3 people If 4-6 people If 7 or more people	**FAMSIZE2** **FAMSIZE4** **FAMSIZE7**	If 2-3 people If 4-6 people If 7 or more people
Language	Reference group is English	**Languse**	Language used all or most of the time
LANGF **LANGO**	If language used is French If language used is other	**LANGF** **LANGO**	If language used is French If language used is other
Ethnicity	Reference group is English only mother tongue	**Lang**	Language
MOTHTNGF **MOTHTNGO** **MOTHTNGB**	If mother tongue is French If unilingual other mother If multilingual mother tongue	**MOTHTNGF** **MOTHTNGO** **MOTHTNGB**	If mother tongue is French If unilingual other mother If multilingual mother tongue
Education	Reference group is secondary	**Educ**	Education
EDUCSPS **EDUCDIP** **EDUCDEG**	If some post secondary If post secondary degree diploma If university degree	**EDUCSPS** **EDUCDIP** **EDUCDEG**	If some post secondary If post secondary degree diploma If university degree
Occupation	Reference group is Professional	**Occup**	Occupation
OCCUPW **OCCUPB** **OCCUPN**	If clerical/sales/ service If blue collar If no occupation or unknown	**OCCUPW** **OCCUPB** **OCCUPN**	If clerical/sales/ service If blue collar Did not use

(Breslaw and Stelcner, p.499)

Table 2. *Breslaw and Stelcner's OLS Estimation: Annual Earnings*

Variables	Coefficients	t-statistics
Natural logarithm of Income	9.7348	134.40
REGIONQ	0.1924	2.81
REGION0	0.2455	4.13
REGIONP	0.2854	4.34
REGIONB	0.3493	5.18
AGE55	-0.0540	1.72
AGE60	-0.0831	1.70
AGE65	-0.2469	2.16
AGE70	-0.8946	4.05
LANGF	-0.0661	1.33
LANGO	-0.1771	3.43
EDUCSPS	0.0351	0.64
EDUCDIP	0.1590	2.81
EDUCDEG	0.2926	5.85
OCCUPW	-0.2355	5.62
OCCUPB	-0.1954	4.58
OCCUPN	-0.9194	2.40
HEALTHRD	-0.1625	2.25

(Breslaw and Stelcner, p.505)

Table 3. *Our OLS Estimation: Annual Earnings*

Variables	**Coefficients**	**Standard Errors**	**t-statistics**
Natural logarithm of Income	9.1694	0.1931	47.484
REGIONQ	0.1938	0.0703	2.755
REGION0	0.2333	0.0489	4.769
REGIONP	0.2335	0.0491	4.753
REGIONB	0.2235	0.0609	3.671
AGE55	-0.0348	0.0366	-0.951
AGE60	-0.1345	0.0445	-3.019
AGE65	-0.4066	0.0857	-4.744
AGE70	-0.5199	0.1126	-4.616
LANGF	-0.1294	0.0649	-1.993
LANGO	-0.2342	0.0782	-2.994
EDUCSPS	0.1892	0.0653	2.899
EDUCDIP	0.3067	0.0656	4.669
EDUCDEG	0.4904	0.0579	8.461
OCCUPW	0.0069	0.0675	0.103
OCCUPB	0.0533	0.0634	0.841
HLTHSCOR	0.0068	0.0042	1.631

```
********************************************************************************

This is a Stata log file for a QED session

Course: Econ 452
Students: jfc
Date and time: Sat, 24 Mar 2001, 13:09:20


At the end of the QED session, this file will be copied to:
83_222_Sat_jfc.log
These files will also be uploaded to:
http://edith.econ.queensu.ca/statausr/logfiles/Econ452
Type help QEDstata for a list of QED commands

Student work begins below this line
********************************************************************************


pause:  "Type BREAK to end session started at 24 Mar 2001 13:09:20"
-> . do "A:\process.do"

. Qextract QEDid region sex agegrp marstat famsize languse lang educ occup income
indinc efinc healthp numhlprb hlthscor prinearn lfstat, dset(12)
getting information about file 12 ...
loading variables from 12 (chs77) only (no data yet)...  done

. codebook sex


sex ---------------------------------------------------------------------- sex
                  type:  numeric (byte)
                 label:  sex

                 range:  [1,2]                        units:  1
         unique values:  2                    coded missing:  0 / 31668

            tabulation:  Freq.   Numeric  Label
                         15655         1  male
                         16013         2  female

. drop if sex>1
(16013 observations deleted)

. codebook age


agegrp ---------------------------------------------------------------- age group
                  type:  numeric (byte)
                 label:  agegrp

                 range:  [1,15]                       units:  1
         unique values:  15                   coded missing:  0 / 15655

              examples:  3     10-14
                         5     20-24
                         7     30-34
```

```
                         11    50-54

. drop if agegrp<11
(12081 observations deleted)

. codebook prinearn


prinearn ---------------------------------------------- principal income earner
                  type:  numeric (byte)
                 label:  prinearn

                 range:  [1,2]                        units:  1
         unique values:  2                    coded missing:  0 / 3574

            tabulation:  Freq.   Numeric  Label
                          2936         1  principal income earner of
                                          economic family
                           638         2  not a principal income earner of
                                          eco family

. drop if prinearn>1
(638 observations deleted)

. codebook lfstat


lfstat ---------------------------------------------------- labour force status
                  type:  numeric (byte)
                 label:  lfstat

                 range:  [1,3]                        units:  1
         unique values:  3                    coded missing:  26 / 2936

            tabulation:  Freq.   Numeric  Label
                          1771         1  employed
                            83         2  unemployed
                          1056         3  not in labour force

. drop if lfstat>1
(1165 observations deleted)

. codebook income


income ------------------------------------- individual income from all sources
                  type:  numeric (int)
                 label:  income, but 59 values are not labeled

                 range:  [0,30000]                    units:  100
         unique values:  60                   coded missing:  508 / 1771

              examples:  12000
                         17000
                         25500
                         .
```

```
. mvencode income, mv(999999999)
income:  508 missing values

. drop if income>999999998
(508 observations deleted)

. gen lnincome= ln(income)
(13 missing values generated)

. codebook occup


occup ------------------------------------------------------------- occupation
                  type:  numeric (byte)
                 label:  occup

                 range:  [1,12]                       units:  1
         unique values:  12                   coded missing:  5 / 1263

              examples:  5     clerical
                         7     services
                         9     mining,processing,machining
                         11    construction

. gen occtwo=occup
(5 missing values generated)

.
.
end of do-file
-> . tab occtwo

     occtwo |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        116        9.22        9.22
          2 |         86        6.84       16.06
          3 |         27        2.15       18.20
          4 |         17        1.35       19.55
          5 |         80        6.36       25.91
          6 |        130       10.33       36.25
          7 |        146       11.61       47.85
          8 |        132       10.49       58.35
          9 |        137       10.89       69.24
         10 |        112        8.90       78.14
         11 |        134       10.65       88.79
         12 |        141       11.21      100.00
------------+-----------------------------------
      Total |       1258      100.00
-> . tab occup

                           occupation |      Freq.     Percent        Cum.
--------------------------------------+-----------------------------------
         managerial,administrative |        116        9.22        9.22
                      professional |         86        6.84       16.06
                          teaching |         27        2.15       18.20
                    medecine-health |         17        1.35       19.55
                          clerical |         80        6.36       25.91
```

```
                                  sales |        130       10.33       36.25
                               services |        146       11.61       47.85
                    primary occupations |        132       10.49       58.35
            mining,processing,machining |        137       10.89       69.24
   fabricating,assembling and repairing |        112        8.90       78.14
                           construction |        134       10.65       88.79
transportation,materials handling & oth |        141       11.21      100.00
----------------------------------------+-----------------------------------
                                  Total |       1258      100.00
-> . do "A:\temp.do"

. recode occtwo 2=20
(86 changes made)

. recode occtwo 5=21 6=21 7=21
(356 changes made)

. recode occtwo 1=22 3=22 4=22 8=22 9=22 10=22 11=22 12=22
(816 changes made)

.
end of do-file
-> . tab occtwo

     occtwo |      Freq.     Percent        Cum.
------------+-----------------------------------
         20 |         86        6.84        6.84
         21 |        356       28.30       35.14
         22 |        816       64.86      100.00
------------+-----------------------------------
      Total |       1258      100.00
-> . xi: regress  lnincome i.region i.agegrp i.languse i.educ i.occup hlthscor
i.region          Iregi_10-50  (naturally coded; Iregi_10 omitted)
i.agegrp          Iageg_11-15  (naturally coded; Iageg_11 omitted)
i.languse         Ilangu_1-3   (naturally coded; Ilangu_1 omitted)
i.educ            Ieduc_2-5    (naturally coded; Ieduc_2 omitted)
i.occup           Ioccu_1-12   (naturally coded; Ioccu_1 omitted)

  Source |       SS       df       MS                  Number of obs =     719
---------+------------------------------               F( 25,   693) =   10.85
   Model |  45.9607315    25  1.83842926               Prob > F      =  0.0000
Residual |  117.409215   693  .169421667               R-squared     =  0.2813
---------+------------------------------               Adj R-squared =  0.2554
   Total |  163.369947   718  .227534745               Root MSE      =  .41161

------------------------------------------------------------------------------
lnincome |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iregi_20 |   .1640345   .0692775      2.368   0.018       .0280156    .3000534
Iregi_30 |    .214467   .0483021      4.440   0.000        .119631     .309303
Iregi_40 |   .2119905   .0491463      4.313   0.000        .115497    .3084841
Iregi_50 |   .2100479   .0596338      3.522   0.000       .0929633    .3271324
Iageg_12 |  -.0347583   .0361404     -0.962   0.337      -.1057161    .0361995
Iageg_13 |  -.1301902   .0436781     -2.981   0.003      -.2159475   -.0444328
Iageg_14 |  -.4031546   .0841725     -4.790   0.000      -.5684183   -.2378909
Iageg_15 |  -.4954667   .1109866     -4.464   0.000       -.713377   -.2775563
Ilangu_2 |  -.1027206   .0636664     -1.613   0.107      -.2277228    .0222816
```

```
Ilangu_3 |  -.1939155   .0769744     -2.519   0.012      -.3450466   -.0427844
 Ieduc_3 |   .0964466   .0654931      1.473   0.141      -.0321421    .2250353
 Ieduc_4 |   .2151377   .0665552      3.232   0.001       .0844637    .3458117
 Ieduc_5 |   .3924991    .065941      5.952   0.000        .263031    .5219672
 Ioccu_2 |  -.2093461   .0744672     -2.811   0.005      -.3555544   -.0631378
 Ioccu_3 |  -.0822529    .117147     -0.702   0.483      -.3122586    .1477527
 Ioccu_4 |  -.4661747   .1938759     -2.405   0.016      -.8468294     -.08552
 Ioccu_5 |  -.2101619   .0757038     -2.776   0.006      -.3587982   -.0615256
 Ioccu_6 |   -.126804   .0680276     -1.864   0.063       -.260369     .006761
 Ioccu_7 |  -.4211817   .0712454     -5.912   0.000      -.5610644    -.281299
 Ioccu_8 |  -.1658814   .0769773     -2.155   0.032      -.3170181   -.0147447
 Ioccu_9 |  -.3278384   .0723107     -4.534   0.000      -.4698126   -.1858641
Ioccu_10 |  -.2395731   .0763494     -3.138   0.002      -.3894771   -.0896691
Ioccu_11 |  -.2396629   .0725036     -3.306   0.001       -.382016   -.0973099
Ioccu_12 |  -.2472346   .0749435     -3.299   0.001      -.3943782   -.1000911
hlthscor |   .0067422   .0040982      1.645   0.100      -.0013041    .0147885
   _cons |   9.457608   .1872876     50.498   0.000       9.089888    9.825327
------------------------------------------------------------------------------
-> .
```

An analysis of - The Colour of Money:

Earnings Differentials among Ethnic Groups in Canada.

Pendakur, Pendakur CJE Aug, 1998

Data Set pumfi, 354

Introduction

Canada has always prided itself for having a diverse multicultural makeup. Citizens and foreigners have always known that Canada open its arms to many different cultures and peoples and it is this fact that makes Canada such a rich country. Throughout the past two decades laws have been introduced to give every individual the same rights regardless of race, sex, or creed. So much so that the Canadian Charter of Rights and Freedom (Canada 1981) outlines the basis in which a person's ethnic heritage should not constrain his or her labour market opportunity. These laws were enacted in order to create an environment where people would be judged on their skills not their ethnic origin. These laws should foster a culture of no wage disparities due to these factors.

Although there is a long history of research assessing whether this vision accurately reflects American labour markets, until recently, there has been comparatively little research done in Canada. However in the past five years, a surge of research has documented the wage disparities that exist amongst Canadians. The researchers Howland and Sakellariou 1993; Christofides and Swidinsky 1994; Stelcner and Kyriazis 1995; Baker and Benjamin 1997 have all conducted research in order to determine if there are wage discrepancies in Canada.

These studies all found that there were earning disparities among different ethnic groups that cannot be attributed to differences in observable individual characteristics such as age and education. Although suitably cautious, the authors concluded that discriminatory practices might be having a negative impact on the earnings of these groups. We examine this issue by looking at the paper "The colour of money: earnings differentials among ethnic groups in Canada." by Pendakur and Pendakur in the Canadian Journal of Economics Aug 1998.

The research of Pendakur and Pendakur suggests that wage disparities when looking at different ethnic groups. At the heart of the analysis is the comparison of Canadian born white individuals and Canadians who are of a visible minority.

The paper contained many different regressions exploring many different combinations of factor, such as where a person is born, the ethnic back ground, if they live in a major city. To allow use to examine a regression, we looked at the first, most simple model, which looked at the effect of being Canadian, or Non-Canadian, and being white or a visible minority, or aboriginal.

They later expanded their analysis to look at males and females as well as disparities amongst visible minorities. The analysis for empirical evaluation is derived from the 1991 PUMF for individuals, which was in the QED data archive. This represents a 3 percent sample size from the Canadian population. The independent variable that we will be using will be the "visible minority status" where we could further define every individual as white or visible minority. We define visible minority as all individuals that are not captured in the white category. We will further break down these two categories into sub-categories where individuals are either Canadian born or immigrants. The basic dependent variable that we will be using in this paper is the log of earnings from wages and salaries. The labour market in Canada is by far the largest sector of Canada's labor force comprising of at least 87 percent for men and 93 percent for women respectively (1991 PUMF). With such a large group, we can safely agree that almost all individuals participate in the labor market not limiting to only a small population. Since visible minorities by large are self employed compared to whites, our analysis has the potential to overestimate earnings gap amongst visible minorities. To correct this problem, we will only use data series where individuals primary source of income is from wage labour sources.

The Data

The data frame for our empirical work includes the following; individuals must be a permanent resident in Canada between the ages of 20 and 64, not in school full time, living in provinces outside the Atlantic region (Quebec and Westward) whose primary source of income was from salary. They also must be employed. As the paper mentions this may shift the finding up as it may be found that one or more group has a great deal of unemployment, which effects the mean wage of the group. However, in the first model, the one we used, all personal characteristics were held constant in order to determine the effect that the independent variables would have. These were later relaxed, and other characteristics were looked.

Looking at the data we determined what variables were used in the study. To estimate our model we needed variables for; age, sex, immigrant status, ethnic group, registered Indian, wage, employment status and province. The age, sex, employment and province, variables were needed in order to obtain the correct observations for our estimation. As stated earlier we were only interested in men, who were employed and who did not live in Atlantic Canada. We then created a log of the wage variable in order to follow the paper. With this we determined the mean wage and log wage for each category of man, and the count of those who fell in to each category. This is shown in the table below. We found that the means for our data were very similar to those published in the paper.

Mean Earnings and Sample Counts

From pumf 354, 1991

Immigrant Status	Equity Status	Mean Earnings ($)	Log of Earnings	Difference of Log	Count
Total		33992.71	10.15123		80171
Canadian	White	34075.52	10.16474	0.0135	61668
	Visible Minority	32051.71	10.0261	-0.1251	755
	Aboriginal	17617.84	9.255543	-0.8958	651
Immigrant	White	37367.22	10.25474	0.1035	10730
	Visible Minority	28801.87	9.924678	-0.2265	6304

The Regression

In order to do the regression for this model it was necessary to determine the interactions of the variables. As the immigration status variable stated if a person was born in Canada, we needed to interact this with whether the person was a visible minority or white. We assumed, as the paper did, that aboriginal people were Canadians, so we had to interact the registered Indian variable with the others. We first tried to use the xi: command. The first attempts using this were not successful do to the registered Indian variable. If this variable had been a third option of the visible minority variable it would have been, possible to just use this command. As it was not it was determined that we should use dummy variables to represent the different groups. By determining the possible combinations we drew a matrix that showed them. We then created five dummy variables to represent these interactions, which are stated below.

Dummy Variable matrix

Canadian White	Can Visible	Can Aboriginal	Imm. White	Imm. Visible
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

By regressing these dummy variables on the log of wages and salaries we were able to estimate the effect that each of these interactions had on the log of wage that each group could earn. When the looking at the results it was important to realize that STATA would drop one of the dummy variables, which then became the constant term.

Stata Output

Variable	Coef.	Std. Err.
Canw	0.775	0.0500
Canvis	0.919	0.0368
Canind	Dropped	
Immw	.999	0.0377
Immv	.669	0.0385
Cons	9.256	0.0366

R-squared = 0.0143, n =79457

These results were for men who were employed between the ages of 24-64 in Canada except the Atlantic region. The results were fine, except for the expected effect of being a Canadian visible minority. This had a value of 0.919, meaning that a visible minority will make an expected average log of earning 10.175. This is not the same as the expected result from the study. They found that the male visible minority Canadian would be paid 10% less than a Canadian white. However, the rest of the results were all in line with those from the paper. It stated that white immigrants were expected to have the highest salary. The result was a higher in our regression. One aspect of the paper that made it hard to compare our results was the way that the results were published. They were stated as a percent comparison of the white Canadian log of earnings. They did not actually state the results that they obtained.

One limitation of the predictive possibilities of the regression is the fact that this estimation was done based only on the facts surrounding the ethnic origin of the males. This kept all other factors that would most likely contribute to discrepancies of wage constant. The paper did continue by looking at other factors, as they wanted to explore this question further. This does leave us with a simplistic model that just states the basic trends. The results should also not be looked at as a way of determining the wage that you should be paid. It is only good at comparing the overall mean

What was interesting to see in our results and those from the paper was the fact that white immigrants had the highest average salary. By looking at the mean of the data, it was by far the highest. By the definition of the ethnic origin, these people must gave been from Western Europe or Australia, which are developed countries. People moving from these countries usually do so due to opportunities created by their education, experience or some other factor. This was examined in later regressions.

Conclusion

While the paper “The Colour of Money” explored more of the reasons for the discrepancies of wages between different groups, we only had a chance to look at the first model. While it was possible to see the same ideas it was not exactly the same result. It would have been interesting to find out what was the cause of the discrepancies in the between our model and that of the paper. Unlike many papers this did not write out the model they used so it was up to our interpretation, so their may have been some discrepancies. One aspect of the model that was positive was the fact that they used a Statistics Canada survey, which allowed us to access the same figures.

The findings of our paper and the paper we looked at were that white immigrants have the highest average earnings. However, our regression was different in the aspect of explaining the average wage of either white Canadians or Canadians of a visible minority.

Appendix

Log File

****

This is a Stata log file for a QED session

Course: Econ 452

Students: money

Date and time: Fri, 23 Mar 2001, 11:13:56

At the end of the QED session, this file will be copied to:

82_192_Fri_money.log

These files will also be uploaded to:

http://edith.econ.queensu.ca/statausr/logfiles/Econ452

Type help QEDstata for a list of QED commands

Student work begins below this line

****

pause: "Type BREAK to end session started at 23 Mar 2001 11:13:56"

-> . set mem 48000K

'48000K' found where number expected

r(198);

```
-> . set mem 48000k
(48000k)
-> . Qextract
getting information about file 354 ...
loading variables from 354 (pumf91i) only (no data yet)...  done
-> . browse
-> . drop immiagep
-> . Qmerge
varlist required
r(100);
-> . help Qextract
-> . Qmerge immpopp, dset(354)
QEDid not found
r(111);
-> . Qextract
getting information about file 354 ...
loading variables from 354 (pumf91i) only (no data yet)...  done
-> . browse
-> . browse
-> . gen lnwage=log(wagesp)
(187086 missing values generated)
-> . label lnwage "Ln of Wage"
invalid syntax
r(198);
-> . name lnwage, "Ln of Wage"
unrecognized command:  name
r(199);
-> . label var lnwage "Ln of Wage"
-> . tabstat lnwage if immpop==1

variable |      mean
```

```
---------+----------
  lnwage |   9.56069
---------+----------
-> . tab lnwage if immpop==1
too many values
r(134);
-> . tabsum lnwage if immpop==1
unrecognized command:  tabsum
r(199);
-> . tabsum lnwage
-> . tab lnwage
too many values
r(134);
-> . tabstat lnwage

variable |      mean
---------+----------
  lnwage |  9.857325
---------+----------
-> . tabstat lnwage if immpopp==1

variable |      mean
---------+----------
  lnwage |   9.83331
---------+----------
-> . browse
-> . tabstat lnwage if immpopp==1, visminp==1
visminp invalid
r(198);
-> . tabstat lnwage if immpopp==1  visminp==1
```

```
invalid 'visminp'

r(198);

-> . tabstat lnwage if immpopp==1 and visminp==1

invalid 'and'

r(198);

-> . tabstat lnwage if immpopp==1,  visminp==1

visminp invalid

r(198);

-> . tabstat lnwage if visminp==1


variable |      mean

---------+----------

  lnwage |  9.619292

---------+----------

-> . browse

-> . browse

-> . tab lnwage, by (reginp==1) c(freq mean)

by() invalid

r(198);

-> . tab lnwage, if (reginp==1) c(freq mean)

if() invalid

r(198);

-> . tabstat lnwage if reginp==1


variable |      mean

---------+----------

  lnwage |  8.868345

---------+----------

-> . drop if provp==60

(532 observations deleted)

-> . tabstat lnwage if reginp==1
```

```
variable |      mean
---------+----------
  lnwage |  8.869129
---------+----------
-> . tabstat wagesp

variable |      mean
---------+----------
  wagesp |  29590.32
---------+----------
-> . browse
-> . drop if wagsp=="-"
wagsp not found
r(111);
-> . drop if wagesp=="-"
type mismatch
r(109);
-> . drop if wagesp==-
invalid syntax
r(198);
-> . browse
-> . drop if wagesp==.
(38351 observations deleted)
-> . browse
-> . tabstat wagesp

variable |      mean
---------+----------
  wagesp |  29590.32
---------+----------
```

```
-> . browse

-> . drop if agep<=23

(16287 observations deleted)

-> . browse

-> . drop if agep>=65

(2447 observations deleted)

-> . browse

-> . tabstat wagesp


variable |      mean

---------+----------

  wagesp |  33992.71

---------+----------

-> . tabstat lnwage if reginp==1


variable |      mean

---------+----------

  lnwage |  9.244879

---------+----------

-> . tab lnwage, by (reginp==1) c(freq mean)

by() invalid

r(198);

-> . tab lnwage, if (reginp==1) c(freq mean)

if() invalid

r(198);

-> . tabstat lnwage if visminp==1


variable |      mean

---------+----------

  lnwage |  9.899722

---------+----------
```

```
-> . tabstat lnwage

variable |      mean
---------+----------
  lnwage |  10.15123
---------+----------
-> . tab lnwage, by (immpopp==1) c(freq mean)
by() invalid
r(198);
-> . tab wagesp
too many values
r(134);
-> . table lnwage , c()
c() invalid
r(198);
-> . table lnwage , c(mean)
mean invalid or requires argument
r(198);
-> . tab lnwage
too many values
r(134);
-> . browse
-> . summary statistics
unrecognized command:  summary
r(199);
-> . sum

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   provp |   80171    37.31733   11.45013         24         59
```

```
    agep |   80171    39.93952   10.63796         24         64
    sexp |   80171           2          0          2          2
 immpopp |   80171    1.230283   .4416639          1          3
 visminp |   80171    1.905739   .2921933          1          2
  reginp |   80171    1.991867   .0898141          1          2
  wagesp |   80171    33992.71   22629.34          1     200000
  lnwage |   80171    10.15123   .9456468          0   12.20607
-> . sum lnwage if immpopp=1
invalid syntax
r(198);
-> . sum lnwage

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |   80171    10.15123   .9456468          0   12.20607
-> . sum lnwage by(immpopp==1)
by:  operator invalid
r(198);
-> . sum lnwage, if immpopp==1
if invalid
r(198);
-> . sum lnwage if immpopp==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |   62423    10.16306   .9269535          0   12.20607
-> . browse
-> . sum lnwage if immpopp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
```

```
  lnwage |   17034    10.13259   .9915219          0   12.20607
-> . browse
-> . sum lnwage if immpopp==1 visminp==2
invalid 'visminp'
r(198);
-> . sum lnwage if immpopp==1, visminp==2
visminp invalid
r(198);
-> . sum lnwage if immpopp==1 + visminp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |       0
-> . sum lnwage if immpopp==1 and visminp==2
invalid 'and'
r(198);
-> . sum lnwage if immpopp==1 & visminp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |   61668    10.16474   .9255927          0   12.20607
-> . drop if immpopp==3
(714 observations deleted)
-> . drop if immpopp==8
(0 observations deleted)
-> . sum lnwage & wagesp if immpopp==2 & visminp==2
& invalid name
r(198);
-> . sum lnwage  wagesp if immpopp==2 & visminp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
```

```
---------+-----------------------------------------------------
  lnwage |   10730    10.25474   .9337538          0   12.20607
  wagesp |   10730    37367.22   24816.64          1     200000
-> . browse
-> . sum lnwage  wagesp if immpopp==1 & visminp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |   61668    10.16474   .9255927          0   12.20607
  wagesp |   61668    34075.52   22190.23          1     200000
-> . sum lnwage  wagesp if immpopp==1 & visminp==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |     755     10.0261   1.023494    4.60517   12.20607
  wagesp |     755    32051.71   24186.31        100     200000
-> . sum lnwage  wagesp if immpopp==2 & visminp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |   10730    10.25474   .9337538          0   12.20607
  wagesp |   10730    37367.22   24816.64          1     200000
-> . sum lnwage  wagesp if immpopp==2 & visminp==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |    6304    9.924678   1.050693   1.609438   12.20607
  wagesp |    6304    28801.87   21073.23          5     200000
-> . browse
-> . sum lnwage  wagesp if reginp==1
```

```
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |     651    9.255543   1.282622   1.791759   11.26446
  wagesp |     651    17617.84   14666.07          6      78000
-> . xi: regress lnwagei..immpopp*visminp
lnwagei:  operator invalid
r(198);
-> . xi: regress lnwage  visminp* immpopp

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  2, 79454) =  247.79
   Model |  436.352554     2  218.176277               Prob > F      =  0.0000
Residual |  69957.1331 79454  .880473395               R-squared     =  0.0062
---------+------------------------------               Adj R-squared =  0.0062
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93834


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 visminp |   .2997909   .0136625     21.943   0.000       .2730124    .3265694
 immpopp |   .0768492   .0094719      8.113   0.000       .0582844     .095414
   _cons |   9.490256   .0336952    281.650   0.000       9.424213    9.556298
------------------------------------------------------------------------------
-> . xi: regress lnwage  visminp*immpopp

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  2, 79454) =  247.79
   Model |  436.352554     2  218.176277               Prob > F      =  0.0000
Residual |  69957.1331 79454  .880473395               R-squared     =  0.0062
---------+------------------------------               Adj R-squared =  0.0062
```

```
   Total |  70393.4857 79456  .885942983              Root MSE      =  .93834

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 visminp |   .2997909   .0136625     21.943   0.000       .2730124    .3265694
 immpopp |   .0768492   .0094719      8.113   0.000       .0582844     .095414
   _cons |   9.490256   .0336952    281.650   0.000       9.424213    9.556298
------------------------------------------------------------------------------
-> . xi  immpopp*visminp
-> . xi:  immpopp* visminp
unrecognized command:  immpopp
r(199);
-> . xi: i.immpopp*i.visminp
i.immpopp            Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp            Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
unrecognized command:  Iimmpo_
r(199);
-> . drop  Iimmpo_2 Ivismi_2 IiXv_2_2
-> . xi: i.immpopp*i.visminp
i.immpopp            Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp            Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
unrecognized command:  Iimmpo_
r(199);
-> . drop  Iimmpo_2 Ivismi_2 IiXv_2_2
-> . xi: i.immpopp*visminp
i.immpopp            Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.immpopp*visminp     IiXvis_#     (coded as above)
unrecognized command:  Iimmpo_
```

```
r(199);

-> . drop  Iimmpo_2 IiXvis_2

-> . xi: regress lnwage i.immpopp*i.visminp

i.immpopp          Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)

i.visminp          Ivismi_1-2   (naturally coded; Ivismi_1 omitted)

i.immpopp*i.visminp   IiXv_#-#     (coded as above)


  Source |       SS       df       MS                  Number of obs =   79457

---------+------------------------------               F(  3, 79453) =  173.96

   Model |  459.362516     3  153.120839               Prob > F      =  0.0000

Residual |  69934.1231 79453  .880194872               R-squared     =  0.0065

---------+------------------------------               Adj R-squared =  0.0065

   Total |  70393.4857 79456  .885942983               Root MSE      =  .93819


------------------------------------------------------------------------------

  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]

---------+--------------------------------------------------------------------

Iimmpo_2 |  -.1014265    .036131     -2.807   0.005      -.1722429     -.03061

Ivismi_2 |   .1386337   .0343525      4.036   0.000       .071303    .2059644

IiXv_2_2 |   .1914273     .03744      5.113   0.000       .1180452    .2648094

   _cons |    10.0261   .0341441    293.641   0.000       9.959182    10.09303

------------------------------------------------------------------------------

-> . xi: regress lnwage i.immpopp* visminp

varlist required

r(100);

-> . xi: regress lnwage i.immpopp*visminp

i.immpopp          Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)

i.immpopp*visminp     IiXvis_#     (coded as above)


  Source |       SS       df       MS                  Number of obs =   79457

---------+------------------------------               F(  3, 79453) =  173.96
```

```
   Model |  459.362516     3  153.120839              Prob > F      =  0.0000
Residual |  69934.1231 79453  .880194872              R-squared     =  0.0065
---------+------------------------------              Adj R-squared =  0.0065
   Total |  70393.4857 79456  .885942983              Root MSE      =  .93819

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -.2928538   .0729252     -4.016   0.000      -.4357867   -.1499208
 visminp |   .1386337   .0343525      4.036   0.000       .071303    .2059644
IiXvis_2 |   .1914273     .03744      5.113   0.000      .1180452    .2648094
   _cons |    9.88747   .0683927    144.569   0.000      9.753421    10.02152
------------------------------------------------------------------------------
-> . drop  Iimmpo_2 IiXvis_2
-> . xi: regress lnwage reginp i.immpopp*i.visminp
i.immpopp            Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp            Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------              F(  4, 79452) =  287.15
   Model |  1003.14627     4  250.786567              Prob > F      =  0.0000
Residual |  69390.3394 79452   .87336177              R-squared     =  0.0143
---------+------------------------------              Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983              Root MSE      =  .93454

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  reginp |   .9187424   .0368195     24.953   0.000       .8465764    .9909083
Iimmpo_2 |  -.1038602   .0359906     -2.886   0.004      -.1744015   -.0333189
```

```
Ivismi_2 |   .1458242   .0342201      4.261   0.000       .0787529    .2128954
IiXv_2_2 |   .1844937   .0372954      4.947   0.000        .111395    .2575924
   _cons |   8.191053   .0810254    101.092   0.000       8.032244    8.349862
------------------------------------------------------------------------------
-> . drop i*
-> . drop  Ivismi_2 Iimmpo_2 IiXv_2_2
-> . browse
-> . keep  i*
i not found
r(111);
-> . Qmerge
varlist required
r(100);
-> . un drop i*
unrecognized command:  un
r(199);
-> . undrop i*
unrecognized command:  undrop
r(199);
-> . Qextract
getting information about file 354 ...
loading variables from 354 (pumf91i) only (no data yet)...  done
-> . drop if provp<14
(32997 observations deleted)
-> . drop if sexp==1
(180569 observations deleted)
-> . drop if wagesp==0
(37625 observations deleted)
-> . drop if provp==60
(532 observations deleted)
```

```
-> . drop if wagesp==.
(38351 observations deleted)
-> . drop if agep<=23
(16287 observations deleted)
-> . drop if agep>=65
(2447 observations deleted)
-> . drop if immpopp==3
(714 observations deleted)
-> . drop if immpopp==8
(0 observations deleted)
-> . tabstat
varlist required
r(100);
-> . tabstat wage

variable |      mean
---------+----------
  wagesp |   34082.4
---------+----------
-> . sum wage

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  wagesp |   79457     34082.4   22579.02          1     200000
-> . gen lnwage=log(wagep)
wagep not found
r(111);
-> . gen lnwage=log(wagesp)
-> . label var lnwage "Ln of Wage"
-> . browse
-> . save "pumf91i_10.dta", replace
```

```
file pumf91i_10.dta saved
-> . browse
-> . xi: regress lnwage i.immpopp*i.visminp
i.immpopp          Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp          Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  3, 79453) =  173.96
   Model |  459.362516     3  153.120839               Prob > F      =  0.0000
Residual |  69934.1231 79453  .880194872               R-squared     =  0.0065
---------+------------------------------               Adj R-squared =  0.0065
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93819

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -.1014265    .036131     -2.807   0.005      -.1722429     -.03061
Ivismi_2 |   .1386337   .0343525      4.036   0.000        .071303    .2059644
IiXv_2_2 |   .1914273     .03744      5.113   0.000       .1180452    .2648094
   _cons |    10.0261   .0341441    293.641   0.000       9.959182    10.09303
------------------------------------------------------------------------------
-> . browse
-> . sum reginp

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  reginp |   79457    1.991807   .0901448          1          2
-> . sum lnwage if reginp==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
```

```
---------+-----------------------------------------------------
  lnwage |   78806    10.16397   .9343078          0   12.20607
-> . sum reginp if reginp==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  reginp |     651           1          0          1          1
-> . sum lnwage if reginp==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |     651    9.255543   1.282622   1.791759   11.26446
-> . browse
-> . drop Iimmpo_2 Ivismi_2 IiXv_2_2
-> . xi: regress lnwage i.immpopp*i.visminp i.immpopp*reginp
i.immpopp            Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp            Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
i.immpopp*reginp      IiXreg_#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  5, 79451) =  230.47
   Model |  1006.39293     5  201.278587               Prob > F      =  0.0000
Residual |  69387.0927 79451  .873331899               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -2.189596   1.082357     -2.023   0.043      -4.31101   -.0681831
```

```
Ivismi_2 |   .1457862   .0342195     4.260   0.000      .0787161    .2128562
IiXv_2_2 |   .1848219   .0372951     4.956   0.000      .1117237    .2579201
Iimmpo_2 |  (dropped)
  reginp |   .9138873   .0369049    24.763   0.000       .841554    .9862206
IiXreg_2 |   1.042875   .5408826     1.928   0.054     -.0172521    2.103001
   _cons |    8.20075   .0811799   101.019   0.000      8.041638    8.359862
------------------------------------------------------------------------------
-> . browse
-> . drop Iimmpo_2 Ivismi_2 IiXv_2_2 IiXreg_2
-> . xi: regress lnwage i.immpopp*i.visminp i.reginp*immpopp
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
i.reginp            Iregin_1-2   (naturally coded; Iregin_1 omitted)
i.reginp*immpopp      IrXimm_#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  5, 79451) =  230.47
   Model |  1006.39293     5  201.278587               Prob > F      =  0.0000
Residual |  69387.0927 79451  .873331899               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  (dropped)
Ivismi_2 |   .1457862   .0342195     4.260   0.000      .0787161    .2128562
IiXv_2_2 |   .1848219   .0372951     4.956   0.000      .1117237    .2579201
Iregin_2 |  -.1289871   .5446466    -0.237   0.813     -1.196491    .9385168
```

```
 immpopp |  -1.146722    .542072     -2.115   0.034      -2.20918    -.084264
IrXimm_2 |   1.042874   .5408826      1.928   0.054     -.0172521    2.103001
   _cons |   10.26136   .5489777     18.692   0.000      9.185366    11.33735
------------------------------------------------------------------------------
-> . drop  Iimmpo_2 Ivismi_2 Ivismi_2 IiXv_2_2 Iregin_2 IrXimm_2
-> . xi: regress lnwage i.immpopp*i.visminp i.immpopp|reginp
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
i.immpopp|reginp      IiXreg_#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  5, 79451) =  230.47
   Model |  1006.39293     5  201.278587               Prob > F      =  0.0000
Residual |  69387.0927 79451  .873331899               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -2.189596   1.082357     -2.023   0.043      -4.31101   -.0681831
Ivismi_2 |   .1457862   .0342195      4.260   0.000      .0787161    .2128562
IiXv_2_2 |   .1848219   .0372951      4.956   0.000      .1117237    .2579201
  reginp |   .9138873   .0369049     24.763   0.000       .841554    .9862206
IiXreg_2 |   1.042875   .5408826      1.928   0.054     -.0172521    2.103001
   _cons |    8.20075   .0811799    101.019   0.000      8.041638    8.359862
------------------------------------------------------------------------------
-> . drop Iimmpo_2 Ivismi_2 IiXv_2_2 IiXreg_2
-> . xi: regress lnwage i.immpopp*i.visminp i.reginp| immpopp
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
```

```
i.visminp             Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
varlist required
r(100);
-> . drop Iimmpo_2 Ivismi_2 IiXv_2_2
-> . xi: regress lnwage i.immpopp*i.visminp i.reginp|immpopp
i.immpopp             Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp             Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
i.reginp              Iregin_1-2   (naturally coded; Iregin_1 omitted)
i.reginp|immpopp      IrXimm_#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  4, 79452) =  288.08
   Model |  1006.34395     4  251.585988               Prob > F      =  0.0000
Residual |  69387.1417 79452  .873321524               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -1.018922    .051374    -19.833   0.000      -1.119615   -.9182296
Ivismi_2 |   .1457954   .0342193      4.261   0.000       .0787258     .212865
IiXv_2_2 |   .1847772   .0372944      4.955   0.000       .1116804    .2578741
 immpopp |  (dropped)
IrXimm_2 |   .9150718   .0365642     25.026   0.000       .8434062    .9867374
   _cons |   9.113456   .0498657    182.760   0.000        9.01572    9.211193
------------------------------------------------------------------------------
-> . drop  Iimmpo_2 Iimmpo_2 IiXv_2_2 Iregin_2 IrXimm_2 Ivismi_2
-> . xi: regress lnwage i.immpopp*i.visminp
```

```
i.immpopp         Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp         Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS              Number of obs =   79457
---------+------------------------------           F(  3, 79453) =  173.96
   Model |  459.362516     3  153.120839           Prob > F      =  0.0000
Residual |  69934.1231 79453  .880194872           R-squared     =  0.0065
---------+------------------------------           Adj R-squared =  0.0065
   Total |  70393.4857 79456  .885942983           Root MSE      =  .93819

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iimmpo_2 |  -.1014265    .036131     -2.807   0.005      -.1722429     -.03061
Ivismi_2 |   .1386337   .0343525      4.036   0.000        .071303    .2059644
IiXv_2_2 |   .1914273     .03744      5.113   0.000       .1180452    .2648094
   _cons |    10.0261   .0341441    293.641   0.000       9.959182    10.09303
------------------------------------------------------------------------------
-> . xi: regress lnwage i.immpopp|reginp i.immpopp*i.  visminp
i.immpopp         Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.immpopp|reginp      IiXreg_#     (coded as above)
varlist required
r(100);
-> . drop Iimmpo_2 IiXreg_2
-> . xi: regress lnwage i.immpopp|reginp i.immpopp*i. visminp
i.immpopp         Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.immpopp|reginp      IiXreg_#     (coded as above)
varlist required
r(100);
-> . xi: regress lnwage i.immpopp|reginp i.immpopp*i.visminp
```

```
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.immpopp|reginp     IiXreg_#     (coded as above)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  5, 79451) =  230.47
   Model |  1006.39293     5  201.278587               Prob > F      =  0.0000
Residual |  69387.0927 79451  .873331899               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  reginp |   .9138873   .0369049     24.763   0.000        .841554    .9862206
IiXreg_2 |   1.042875   .5408826      1.928   0.054      -.0172521    2.103001
Iimmpo_2 |  -2.189596   1.082357     -2.023   0.043       -4.31101   -.0681831
Ivismi_2 |   .1457862   .0342195      4.260   0.000       .0787161    .2128562
IiXv_2_2 |   .1848219   .0372951      4.956   0.000       .1117237    .2579201
   _cons |    8.20075   .0811799    101.019   0.000       8.041638    8.359862
------------------------------------------------------------------------------
-> . browse
-> . xi: regress lnwage reginp|i.immpopp i.immpopp*i.visminp
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)
| invalid name
r(198);
-> . xi: regress lnwage i.reginp|immpopp i.immpopp*i.visminp
```

```
i.reginp          Iregin_1-2   (naturally coded; Iregin_1 omitted)
i.reginp|immpopp      IrXimm_#     (coded as above)
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS              Number of obs =   79457
---------+------------------------------           F(  4, 79452) =  288.08
   Model |  1006.34395     4  251.585988           Prob > F      =  0.0000
Residual |  69387.1417 79452  .873321524           R-squared     =  0.0143
---------+------------------------------           Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983           Root MSE      =  .93452


------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 immpopp |  -1.018922    .051374    -19.833   0.000      -1.119615   -.9182296
IrXimm_2 |   .9150718   .0365642     25.026   0.000       .8434062    .9867374
Iimmpo_2 |  (dropped)
Ivismi_2 |   .1457954   .0342193      4.261   0.000       .0787258     .212865
IiXv_2_2 |   .1847772   .0372944      4.955   0.000       .1116804    .2578741
   _cons |   10.13238   .0690322    146.778   0.000       9.997076    10.26768
------------------------------------------------------------------------------
-> . xi: regress lnwage i.reginp|i.immpopp i.immpopp*i.visminp
I.xxx|I.yyy not allowed
r(198);
-> . xi: regress lnwage i.reginp*immpopp i.immpopp*i.visminp
i.reginp          Iregin_1-2   (naturally coded; Iregin_1 omitted)
i.reginp*immpopp      IrXimm_#     (coded as above)
i.immpopp           Iimmpo_1-2   (naturally coded; Iimmpo_1 omitted)
i.visminp           Ivismi_1-2   (naturally coded; Ivismi_1 omitted)
```

```
i.immpopp*i.visminp   IiXv_#-#     (coded as above)

  Source |       SS       df       MS                  Number of obs =   79457
---------+------------------------------               F(  5, 79451) =  230.47
   Model |  1006.39293     5  201.278587               Prob > F      =  0.0000
Residual |  69387.0927 79451  .873331899               R-squared     =  0.0143
---------+------------------------------               Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983               Root MSE      =  .93452

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iregin_2 |  -.1289871   .5446466     -0.237   0.813      -1.196491    .9385168
 immpopp |  -1.146722    .542072     -2.115   0.034       -2.20918    -.084264
IrXimm_2 |   1.042874   .5408826      1.928   0.054      -.0172521    2.103001
Iimmpo_2 |  (dropped)
Ivismi_2 |   .1457862   .0342195      4.260   0.000       .0787161    .2128562
IiXv_2_2 |   .1848219   .0372951      4.956   0.000       .1117237    .2579201
   _cons |   10.26136   .5489777     18.692   0.000       9.185366    11.33735
------------------------------------------------------------------------------
-> . drop  IrXimm_2 IrXimm_2 Iimmpo_2 Ivismi_2 IiXv_2_2
-> . drop Iregin_2
-> . tab  immpopp

    immigrant status indicator |      Freq.     Percent        Cum.
-------------------------------+-----------------------------------
perm. residents: non-immigrant |      62423       78.56       78.56
    perm. residents: immigrant |      17034       21.44      100.00
-------------------------------+-----------------------------------
                         Total |      79457      100.00
-> . tab  visminp
```

```
    visible minority indicator |      Freq.     Percent        Cum.
-------------------------------+-----------------------------------
    member of visible minority |       7059        8.88        8.88
non-member of visible minority |      72398       91.12      100.00
-------------------------------+-----------------------------------
                         Total |      79457      100.00
-> . tab  lnwage
too many values
r(134);
-> . tab reginp

 registered indian indicator |      Freq.     Percent        Cum.
-----------------------------+-----------------------------------
 registered under indian act |        651        0.82        0.82
not regist. under indian act |      78806       99.18      100.00
-----------------------------+-----------------------------------
                       Total |      79457      100.00
-> . edit
- preserve
-> . edit
- preserve
-> . edit
- preserve
-> . gen dummy1=1 if impopp=1
impopp not found
r(111);
-> . gen dummy1=1 if immpopp=1
invalid syntax
r(198);
-> . gen dummy1==1 if immpopp=1
```

== invalid name

r(198);

-> . gen dummy1=1 if immpopp==1

(17034 missing values generated)

-> . replace dummy1=0 if immpopp=2

invalid syntax

r(198);

-> . replace dummy1=0 if immpopp==2

(17034 real changes made)

-> . edit

- preserve

-> . gen dummy2=1 if visminp==1

(72398 missing values generated)

-> . replace dummy2=0 if immpopp==2

(17034 real changes made)

-> . gen dummy3=1 if reginp==1

(78806 missing values generated)

-> . replace dummy3=0 if reginp=2

invalid syntax

r(198);

-> . replace dummy3=0 if reginp==2

(78806 real changes made)

-> . browse

-> . browse

-> . gen dummy2=1 if visminp==1

dummy2 already defined

r(110);

-> . browse

-> . replace dummy2=0 if immpopp==2

(0 real changes made)

```
-> . browse

-> . drop dummy2

-> . gen dummy2=1 if visminp==1

(72398 missing values generated)

-> . browse

-> . replace dummy2=0 if visminp==2

(72398 real changes made)

-> . browse

-> . tabstat visminp


variable |      mean

---------+----------

 visminp |  1.911159

---------+----------

-> . summ visminp


Variable |     Obs        Mean   Std. Dev.       Min        Max

---------+-----------------------------------------------------

 visminp |   79457    1.911159   .2845152          1          2

-> . tab  visminp


    visible minority indicator |      Freq.     Percent        Cum.

-------------------------------+-----------------------------------

    member of visible minority |       7059        8.88        8.88

non-member of visible minority |      72398       91.12      100.00

-------------------------------+-----------------------------------

                         Total |      79457      100.00

-> . regress lnwage dummy1 dummy2


  Source |       SS       df       MS                Number of obs =   79457

---------+------------------------------                F(  2, 79454) =  247.79
```

```
   Model |  436.352554     2  218.176277              Prob > F      =  0.0000
Residual |  69957.1331 79454  .880473395              R-squared     =  0.0062
---------+------------------------------              Adj R-squared =  0.0062
   Total |  70393.4857 79456  .885942983              Root MSE      =  .93834

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  dummy1 |  -.0768492   .0094719     -8.113   0.000       -.095414   -.0582844
  dummy2 |  -.2997909   .0136625    -21.943   0.000      -.3265694   -.2730124
   _cons |   10.24354   .0087895   1165.430   0.000       10.22631    10.26076
------------------------------------------------------------------------------
-> . drop  dummy1 dummy3 dummy2
-> . browse
-> . browse
-> . gen dcan=1 if   visminp==1
(72398 missing values generated)
-> . replace dcan=0 if visminp==2
(72398 real changes made)
-> . browse
-> . browse
-> . drop  dcan
-> . browse
-> . gen Canw=1 if  immpopp==1 &  visminp==2
(17789 missing values generated)
-> . browse
-> . drop  Canw
-> . gen canw=1
-> . browse
-> . drop  Canw
Canw not found
```

r(111);

-> . drop canw

-> . gen canw=1 if immpopp==1

(17034 missing values generated)

-> . replace canw=o if visminp==1

o not found

r(111);

-> . replace canw=0 if visminp==1

(7059 real changes made)

-> . browse

-> . drop canw

-> . gen canw=1 if immpopp==1

(17034 missing values generated)

-> . replace canw=0 if visminp==2

(72398 real changes made)

-> . browse

-> . drop canw

-> . gen canw=1 if immpopp==1 & visminp==2 & reginp==2

(18435 missing values generated)

-> . replace canw=0 if immpopp==2 & visminp==1 & reginp==1

(0 real changes made)

-> . browse

-> . drop canw

-> . gen canw=1 if immpopp==1 & visminp==1 & reginp==2

(78704 missing values generated)

-> . browse

-> . gen canw=0 if immpopp==.

canw already defined

r(110);

-> . replace canw=0 if canw==.

(78704 real changes made)

-> . browse

-> . gen canvis=1 if immpopp=1 & visminp==2 & reginp==2

invalid syntax

r(198);

-> . gen canvis=1 if immpopp==1 & visminp==2 & reginp==2

(18435 missing values generated)

-> . browse

-> . gen canvis=1 if immpopp==1 & visminp==2 & reginp==2

canvis already defined

r(110);

-> . drop canvis

-> . browse

-> . gen canvis=1 if immpopp==1 & visminp==2 & reginp==2

(18435 missing values generated)

-> . browse

-> . replace canvis=0 if canvis==.

(18435 real changes made)

-> . gen canind=1 if   reginp==1

(78806 missing values generated)

-> . browse

-> . replace canind=0 if canind==.

(78806 real changes made)

-> . browse

-> . gen immw=1 if immpopp==2 & visminp==2 & reginp==2

(68730 missing values generated)

-> . browse

-> . replace immw=0 if immw==.

(68730 real changes made)

-> . browse

-> . gen immvis=1 if immpopp==2 & visminp==1 & reginp=2

```
invalid syntax
r(198);
-> . gen immvis=1 if immpopp==2 &  visminp==1 & reginp==2
(73153 missing values generated)
-> . browse
-> . replace  immvis=0 if immvis==.
(73153 real changes made)
-> . browse
-> . browse
-> . regress lnwage  canw canvis canind immw immvis

  Source |       SS       df       MS              Number of obs =   79457
---------+------------------------------           F(  4, 79452) =  287.18
   Model |  1003.23193     4  250.807983           Prob > F      =  0.0000
Residual |  69390.2537 79452  .873360692           R-squared     =  0.0143
---------+------------------------------           Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983           Root MSE      =  .93454

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    canw |   .7746228   .0500141     15.488   0.000       .6765955    .8726501
  canvis |   .9187477   .0368223     24.951   0.000       .8465762    .9909191
  canind |  (dropped)
    immw |   .9997425   .0377225     26.503   0.000       .9258066    1.073678
  immvis |   .6691344   .0384722     17.393   0.000       .5937292    .7445396
   _cons |   9.255543   .0366274    252.694   0.000       9.183754    9.327333
------------------------------------------------------------------------------
-> . gen immind=1 if immpopp==2 &  visminp==1 & reginp==1
(79457 missing values generated)
-> . browse
```

```
-> . replace immind=0 if immind==.
(79457 real changes made)
-> . browse
-> . regress  lnwage canw canvis canind immw immvis immind

  Source |       SS       df       MS              Number of obs =   79457
---------+------------------------------           F(  4, 79452) =  287.18
   Model |  1003.23193     4  250.807983           Prob > F      =  0.0000
Residual |  69390.2537 79452  .873360692           R-squared     =  0.0143
---------+------------------------------           Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983           Root MSE      =  .93454

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    canw |   .7746228   .0500141     15.488   0.000       .6765955    .8726501
  canvis |   .9187477   .0368223     24.951   0.000       .8465762    .9909191
  canind |  (dropped)
    immw |   .9997425   .0377225     26.503   0.000       .9258066    1.073678
  immvis |   .6691344   .0384722     17.393   0.000       .5937292    .7445396
  immind |  (dropped)
   _cons |   9.255543   .0366274    252.694   0.000       9.183754    9.327333
------------------------------------------------------------------------------
-> . regress  lnwage canw canvis canind immw immvis

  Source |       SS       df       MS              Number of obs =   79457
---------+------------------------------           F(  4, 79452) =  287.18
   Model |  1003.23193     4  250.807983           Prob > F      =  0.0000
Residual |  69390.2537 79452  .873360692           R-squared     =  0.0143
---------+------------------------------           Adj R-squared =  0.0142
   Total |  70393.4857 79456  .885942983           Root MSE      =  .93454
```

```
------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    canw |   .7746228   .0500141     15.488   0.000       .6765955    .8726501
  canvis |   .9187477   .0368223     24.951   0.000       .8465762    .9909191
  canind |  (dropped)
    immw |   .9997425   .0377225     26.503   0.000       .9258066    1.073678
  immvis |   .6691344   .0384722     17.393   0.000       .5937292    .7445396
   _cons |   9.255543   .0366274    252.694   0.000       9.183754    9.327333
------------------------------------------------------------------------------
-> . browse
-> . save "C:\courses\Jan.dta"
file C:\courses\Jan.dta saved
-> . tab  lnwage
too many values
r(134);
-> . exit
Session ended at 23 Mar 2001; 16:28:14
********************************************************************************
```

Karoshi

By:

Econ 452 Assignment 1

Professor: Chris Ferrell

National Population Health Survey (NPHS)

Shields, Margot. 1999. "Long working hours and health." *Health Reports* Autumn 1999, 11(2). Ottawa: Statistics Canada.

Margot Shields hypothesizes that long hours bring about unhealthy lifestyle changes in her paper entitled "Long Working Hours and Health". The Karoshi model, developed by Japanese researchers, is examined to determine the influence of long hours on cardiovascular disease in Canada. Karoshi translates as dying from cardiovascular causes "from overwork". Shields uses data from the National Population Health Survey (1994/1995 - 1996/97) to determine the relationship between long working hours, coupled with work-related and socio-demographic variables, on conditions considered to be unhealthy. Such conditions focused on in the paper were depression, weight, smoking, alcohol consumption, and physical activity. These health concerns are considered to be the main underlying causes of death due to cardiovascular disease.

In regressing long hours on any one of the aforementioned health conditions, the control variables (referred to as the work-related and socio-demographic variables) were taken into account to see if long hours still maintained any statistically significant influence. The control variables are listed as follows:

Socio-demographic factors: age, marital status, educational attainment, household income, and the presence of children <12 in the household.

Work-related factors: occupation type (e.g. White collar), self-employment, shift work, multiple jobs, high job strain, high job security, and low supervisor report. All but occupation type was given by (Yes/No) answers.

Table 1 and 2 in the paper illustrate the percentage working long hours by gender aged 25 - 54 who worked 35 hours or more per week throughout 1994/95 by the selected characteristics outlined above. For men only could long hours be associated to age. For both genders, marital status was not associated with long hours, while higher educational attainment and income was. The presence of young children did not effect the proportion of hours worked for women although it was related to higher levels for men. In terms of work characteristics men and women in white collar positions experienced longer work hour as with the shift-work, self-employment and multiple job holder factors. High job-strain, job insecurity and low supervisor support did not show significant influence on the proportions of hours worked.

After performing separate regression analysis for men and women (not presented in the content of the paper) the following conclusions were made regarding each of the health circumstances.

Depression: Women working long hours "had 2.2 times the odds of reporting having experienced a major depressive episode, compared with those who worked standard hours", while no incidence of depression associated with work hours was found for men. High job strain was related to depression for both sexes.

Weight: This was measured in terms of the body mass index (BMI). Men had a higher incidence of being overweight than women for the 1994/95 NPHS data (36% versus 23% in Table 4). "When factors such as age, education, smoking status, occupation, shift work, and work stress were taken into account" it was determined that men had increased odd of 1.4 in having excess body weight. In women no association was found between weight and long hours after taking the above control variables into account.

Smoking: For the 1994/95-year the percentages of male and female workers who were daily smokers were approximately 28 and 25%. In this study no relationship was found between long hours and daily smoking for either of the sexes. However, observing the longitudinal data, increased smoking did occur when a switch was made from standard to long hours. The odds ratio table D in the appendix of Margot Shield's paper is replicated at the end of this section.

Alcohol consumption: Only longitudinal conclusions are reported for this health risk. An increase in weekly hours was not associated with increased alcohol consumption for men, although women and higher odds of higher consumption when switching from standard to long hours.

Physical Activity: For both sexes, there were no significant differences in the average number of time exercising between those who worked standard and long hours. Also the paper reports that increased working hours did not decrease the odds of reported hours exercised.

Adjusted odds ratios relating selected charcteristics to increased daily smoking between 1994/95 and 1996/97 among men and women aged 25 - 54 who worked 35 more hours per week throughout 1994/95, Canada excluding territories. (ctd. next page)

	Men		**Women**	
	Odds ratio	**95% C.I.**	**Odds ratio**	**95% C.I.**

Working hours (1994/95 - 1996/97)				
Long - long	1.1	0.6, 2.0	1.0	0.3, 2.9
Standard - long	2.2*	1.1, 4.5	4.1*	1.4, 11.6
Long - reduced	1.2	0.6, 2.3	1.7	0.8, 4.0
Standard- reduced	1.7	0.7, 4.2	1.3	0.6, 2.8
Standard - standard	1.0	…	1.0	…
Occupation				
White - collar	0.6	0.3, 1.0	0.4*	0.2, 0.8
Self employed	0.5*	0.3, 0.9	0.9	0.3, 2.4
shift worker	1.0	0.6, 1.9	1.3	0.5, 3.1
Multiple job holder	1.5	0.6, 3.9	1.2	0.4, 3.8
Work stress				
high job strain	1.0	0.6, 1.7	0.9	0.5, 1.6
high job insecurity	0.7	0.4, 1.1	1.4	0.8, 2.3
Low super. Support	0.9	0.5, 1.6	1.3	0.7, 2.7
Age				
25 -34	1.0	…	1.0	…
35 -44	0.7	0.4, 1.2	0.9	0.5, 1.8
45 - 54	0.6	0.3, 1.1	0.9	0.4, 2.1
Married	0.9	0.5, 1.6	0.5*	0.3, 0.9
Children <12 years	1.0	0.6, 1.7	1.2	0.6, 2.3
Education				
Sec. Grad or less	1.0	…	1.0	…
Some post-secondary	1.0	0.6, 1.7	0.5	0.3, 1.1
Post-secondary Grad	0.5*	0.3, 0.9	0.4*	0.25, 0.7
Household Income				
Lowest/ Low middle/ Middle	0.9	0.5, 1.7	0.6	0.2, 1.4
Upper-middle	0.9	0.5, 1.6	0.7	0.3, 1.6
Highest	1.0	…	1.0	…
" * " refers to a p-value =< 0.05 and "…" Not appropriate				

Data:

The National Population Health Survey (NPHS) conducted by Statistics Canada collects both cross-sectional and longitudinal data on the physical and mental health of Canadians and their use of health care services. The main objectives are to:

- Measure the health status of the population and its relationship to the use of health care services and various determinants of health;
- Collect data on the economic, social demographic, occupational and environmental

correlates of health;

- And to provide information on a selection of individuals who will be followed over time to reflect the dynamic process of health and illness;

The first cycle of data collection took place in 1994/95; and second, in 1996/97. The third cycle began in June 1998 and will continue through June 1999. This survey and its longitudinal components are expected to last 20 years.

This will be the data set used in this paper. The data set has many components. We used the 1994/95 health component of the cross sectional proportion of the data set. The health component of the survey contains detailed health questions about one randomly selected individual per household. The answers are provided by the selected individual.

In Shields' paper the data are from the household longitudinal component of the survey that includes both the 1994/95 and 1996/97 cycles. Her results are based on 3,380 adult workers aged 25 to 54 (2,181 men and 1,649 women) who worked 35 hours or more per week throughout the year before their 1994/95 interview.

In an attempt to replicate a similar data set, the same criteria were applied. The differences in the count of the actual observations stem from the differences in the components of the NPHS used. The 1994/95 cross sectional health component of the NPHS has 17,626 observations in total. The first step taken was to limit the observations to those who were between the ages of 25-54. The age of the individual was categorized into specific age cohorts. Those observations were only kept if they were in an age cohort between 25-54 (variable agegrp). This brought the count down to 9291. Next, only those who worked were kept as an observation. The variable used in this limitation was lfs_q1 which indicates an individual's labour force status. Only those who were working for pay/profit were kept in the data set. This lowered the count to 4610. The third step was to eliminate those who worked less than 35 hours per week. To do this, all who only had one part-time job, were eliminated (variable dvwh94). This brought us to the final number of observations used in our data set: 4,231 (2,807 males and 1424 females). This is as close as we come to Margot Shields' paper. The difficulty in replicating exactly the same data set apart from the obvious that two different data sets were being used, arise from the fact that there is no variable which specifies the number of hours per week each individuals worked. Therefore, only those who had one part-time job can be eliminated safely from the pool of survey respondents as working less than 35 hours. Part-time is defined as less than 30 hours. Potentially, there could be individuals who could hold multiple part-time jobs yet work less than 30 hours a week. This part-time workers remains in the data set due to the uncertainty. There are no questions which asks how many hours individuals are working. This serves as the best possible proxy for standard hour workers, those who worked 35-40 hours and long working hour workers who worked 41+ hours.

The remainder of the replication a similar data set required collapsing numerous levels in some variable to the specifications of the paper. The first handles the education levels of the individuals. The three resulting categories based on dwhhin94 are: somesec, high school graduation or less; somepost, some post-secondary education; and post-secondary graduation. The second transformation is of the level of the income variable. The classifications are according to the Shields' paper: lowest income; lower-mid income; middle income; upper-middle income; and highest income. After all these modifications, the observations are similar to those used in "Long working hours and health."

Results:

"There was, however, no relationship between working hours and the propensity to be a daily smoker in 1994/95." Though this is stated in the paper, there is no data provided to back up this claim. Thus we ran a regression to see if in fact there was a relationship between working hours and the propensity to be a daily smoker. We regress the dummy variable of long working hours on the dummy variable of the daily smoker. Our null hypothesis is that the propensity to be a daily smoker, i.e., the coefficient would be = 0. In running our regression and the corresponding F-Test that the null hypothesis is indeed true and thus we accept the null hypothesis and therefore the propensity to be a daily smoker is not associated with longer working hours using the 1994/95 data.

	Long Hours
Smoker	0.015 (0.018)

Standard error shown in parentheses

H_0 : longwh = 0.0

F(1, 4229) = 2.19

Prob > F = 0.1390

The second regression we run to test the association of numbers of visits to the general practitioner on long working hours controlling for income, gender, and education level. Table 2 shows the results. The null hypothesis that we test is that the coefficient for dummy variable standard working hours is = 0.0 The test accepts the null hypothesis thus the coefficient is insignificant. Again, either working standard hours or long hours does not affect the number of visits to the general practitioner.

Thus, we conclude that longer working hours have insignificant impact on propensity of being a smoker or the opportunity of visiting the general practitioner. The cross-sectional nature of the tests may explain these results as that it may signal that at the time of the survey, longer working hours have no significant impact. Accompanied with the longitudinal aspect of this survey, a better understanding of the effect of longer working hours can then be achieved.

Summary:

“Long working hours and health” was a very interesting read. It uses econometric methods to examine important issues like health. Work is an important aspect of life. However, like everything else, moderation is key. Overworking can lead to fatigue and higher stress levels. This could lead to complications and deteriorating health. By looking at the cross-sectional data it is possible to capture a glimpse of the current health status of individuals. By incorporating this with longitudinal statistics, it would be possible then to measure the dynamic influences on health.

The troubles we encountered were distinguishing the differences in the cross-sectional and longitudinal aspects of the survey. Further research was necessary to understand all the concepts and definitions of the survey. “The National Population Health Survey – its longitudinal nature” written by Larry Swain, Garry Catlin and Marie Beaudet proved to be helpful in this respect. Also, we were uncertain as to graphing of our data. Due to the usage of dummy variables, it seemed the odds ratios were what our results seem to indicate. Odds should be under one and over 0 which was not what our data indicated due to the linear nature of the equations. Thus we hope to be able to incorporate this into our next project.

The national population health survey serves a need for information regarding Canadian health and how different things are affecting it. Does high socio-economic status facilitate access to conditions that promote good health, or does good health enable an individual to achieve high socio-economic status? The cross-sectional component combined with the longitudinal statistics should be able to provide the answers.

Appendix:

Link to Stata Log file.

Table:

Table 1: The Numbers of Visits to General Practicioners

Variables	**Coeffecients and Error terms**
Number of visits to General practicioners	
_cons	3.06
	(0.40)
Standard	-0.24
	(0.16)
female	1.19
	(0.13)
Educational Attainment	
somepost	-0.03
	(0.14)
postsec	-0.14
	(0.18)
Income Levels	
lowmid	-0.41
	(0.44)
mid	-0.68
	(0.38)
upmid	-0.6
	(0.38)
highest	-0.64
	(0.39)

Number of Observations	4231

Standard errors shown in brackets below each coeffecient value.
Data from the National Population Health Survey - its longtitudinal nature
1994/95 Health Sample, dataset 337, nphs in the DLI

****

This is a Stata log file for a QED session

Course: Econ 452

Students: th

Date and time: Sun, 25 Mar 2001, 16:23:06

At the end of the QED session, this file will be copied to:

84_282_Sun_th.log

These files will also be uploaded to:

http://edith.econ.queensu.ca/statausr/logfiles/Econ452

Type help QEDstata for a list of QED commands

Student work begins below this line

****

pause: "Type BREAK to end session started at 25 Mar 2001 16:23:06"

-> . Qextract, ds(337)

getting information about file 337 ...

loading variables from 337 (nphs94h) only (no data yet)... done

-> . do "C:\WINDOWS\TEMP\STD0d0000.tmp"

. /*summarize the variables*/

. summarize

Variable		Obs	Mean	Std. Dev.	Min	Max
QEDid	|	17626	8813.5	5088.332	1	17626
agegrp	|	17626	7.247192	3.817422	1	15
sex	|	17626	1.542834	.498176	1	2
marstatg	|	17622	1.645557	.775166	1	3
hhsizeg	|	17626	2.643765	1.272117	1	5
numle5g	|	17626	1.851072	.3560273	1	2
num6t11g	|	17626	1.847044	.3599552	1	2
ut_q2a	|	17584	3.741413	5.193526	0	31
dvedc294	|	17601	5.882961	3.136378	1	12

lfs_q1	16985	3.461172	2.164576	1	8
dvwh94	10851	1.758824	1.284787	1	5
inc_q2g	17413	1.664504	1.098315	1	6
dvhhin94	16893	7.012964	2.455526	1	11
dvsmkt94	17618	3.974061	2.026519	1	6

```
. /* keep only those between the age of 25-54*/
. drop if agegrp <=3
(3242 observations deleted)
. drop if agegrp >=10
(5093 observations deleted)
. count
9291
. /* keep only those who are working*/
. drop if lfs_q1 ==1
(1497 observations deleted)
. drop if lfs_q1 >=3
(3184 observations deleted)
. /* drop those working only part time (less than 30 hours) and those who did not report their income */
. drop if dvwh94 ==2
(187 observations deleted)
. drop if dvwh94 >=6
(41 observations deleted)
. drop if dvhhin94 == .
(145 observations deleted)
. /* education is coded into 11 categories, will collapse to 3*/
. tab dvedc294
```

Highest level of education attained	Freq.	Percent	Cum.
Elementary school	110	2.60	2.60

Some secondary school	|	574	13.56	16.16
Secondary school graduation	|	668	15.78	31.95
Other beyond high school	|	19	0.45	32.40
Some trade school etc	|	404	9.55	41.94
Some community college	|	392	9.26	51.21
Some university	|	229	5.41	56.62
Trade school diploma/cert.	|	550	13.00	69.61
Community college diploma/cert	|	390	9.22	78.83
Bachelor degree (incl llb)	|	681	16.09	94.92
master/doctorate/medicine deg.	|	215	5.08	100.00
Total	|	4232	100.00	

. gen somesec = cond(dvedc294 <5,1,0)

. gen somepost = cond(dvedc294 >4& dvedc294 <11,1,0)

-> . gen postsec = cond(dvedc294 >10,1,0)

-> . gen school = somesec + 2*somepost + 3*postsec

. tab school

school	|	Freq.	Percent	Cum.
1	|	1352	31.91	31.91
2	|	1984	46.83	78.73
3	|	901	21.27	100.00
Total	|	4237	100.00	

. label define schools 1 "High school graduate or less" 2 "some post secondary edcation" 3 "post secondary graduate"

. label values school schools

. /* income is coded into 11 categories, will collapse to 5 based on household size*/

. tab dvhhin94 hhsizeg

household	|	Household size				|	
income	|	1 person	2 persons	3 persons	4 persons	|	Total

No income	6	10	4	4	26
Less than 5,000	4	8	4	1	17
5,000 - 9,999	44	11	8	3	68
10,000 - 14,999	75	39	20	10	148
15,000 - 19,999	80	51	25	27	201
20,000 - 29,999	223	146	68	71	537
30,000 - 39,999	226	198	115	97	699
40,000 - 49,999	169	194	133	134	688
50,000 - 59,999	99	203	110	125	594
60,000 - 79,999	60	229	140	176	667
80,000 or more	26	225	146	139	592
Total	1012	1314	773	787	4237

```
. gen lowest = cond(dvhhin94 <4 & hhsizeg <5 | dvhhin94 < 5 & hhsizeg ==5, 1, 0)
. gen lowmid = cond(dvhhin94 ==4 & hhsizeg <3 | dvhhin94 < 6 & dvhhin94 > 3 & hhsizeg < 5 & hhsizeg >2 | dvhhin94 >4 & dvhhin94 < 7 & hhsizeg ==5,1,0)
. gen mid = cond(dvhhin94 <7 & dvhhin94 >4 & hhsizeg <3 | dvhhin94 < 8 & dvhhin94 > 5 & hhsizeg < 5 & hhsizeg >2 | dvhhin94 >6 & dvhhin94 < 10 & hhsizeg ==5,1,0)
. gen upmid = cond(dvhhin94 <10 & dvhhin94 >6 & hhsizeg <3 | dvhhin94 < 11 & dvhhin94 > 7 & hhsizeg < 5 & hhsizeg >2 | dvhhin94 ==10 & hhsizeg ==5,1,0)
. gen highest = cond(dvhhin94 <=11 & dvhhin94 >9 & hhsizeg <3 | dvhhin94 ==11 & hhsizeg >2,1,0)
. gen inc = lowest + 2*lowmid + 3*mid + 4*upmid + 5*highest
. tab inc
```

inc	Freq.	Percent	Cum.
1	115	2.71	2.71
2	243	5.74	8.45
3	1029	24.29	32.74
4	1969	46.47	79.21
5	881	20.79	100.00
Total	4237	100.00	

```
. label define incs 1 "lowest" 2 "lower-middle" 3 "middle" 4 "upper-middle" 5 "highest"
. label values inc incs
```

. /* working hours is coded into 4 categories, will collapse to 2 : standard hours and long hours*/

gen standard = cond(dvwh94 ==1|dvwh94 ==4,1,0)

. gen longwh = cond(dvwh94 ==3|dvwh94 ==5,1,0)

. gen hours = standard + 2*longwh

. label define hour 1 "standard hours" 2 "long working hours"

. label values hours hour

. /* generate male and female dummy variables*/

. gen male = cond(sex ==1,1,0)

. gen female = cond(sex ==2,1,0)

. /*tabulations for smoking*/

. tab dvsmkt94 hours

	hours		
Type of smoker	standard	Long work	Total
Daily smoker	1053	241	1294
Occ smoker (former daily)	102	30	132
Always an occasional	64	18	82
Former daily smoker	792	157	949
Former occasional smo	221	68	289
Never smoked	1246	244	1490
Total	3478	758	4236

. gen smoker = cond(dvsmkt94 == 1,1,0)

. label define smokers 0 "non daily smoker" 1 "daily smoker

. label values smoker smokers

. tab smoker hours

	hours		
smoker	standard	longwork	Total

Non daily smoker	2426	517	2943
Daily smoker	1053	241	1294
Total	3479	758	4237

. /*regressions*/

. regress smoker longwh

Source	SS	df	MS			
						Numberofobs=4237
						F(1,4235)=0.68
Model	.14510197	1	.14510197			Prob>F=0.4083
Residual	898.661129	4235	.212198614			R-squared=0.0002
						AdjR-squared=-0.0001
Total	898.806231	4236	.212182774			RootMSE=.46065
smoker	Coef.	Std. Err.	t- statistic	P>|t|	[95%Conf.Interval]	
longwh	.0152688	.0184645	0.827	0.408	-.0209314	.051469
_cons	.3026732	.0078099	38.755	0.000	.2873617	.3179846

. test longwh

(1) longwh = 0.0

F(1, 4235) = 0.68

Prob > F = 0.4083

/* other tabulations */

. tab hours sex

	sex		
hours	male	female	Total
Standard hours	2303	1176	3479
Long working hours	507	251	758
Total	2810	1427	4237

. tab ut_q2a sex

Number of |

Visits to general practitioner Past year	sex male	female	Total
0	898	217	1115
1	736	331	1067
2	443	271	714
3	226	185	411
4	154	103	257
5	72	60	132
6	93	91	184
7	17	20	37
8	38	26	64
9	3	5	8
10	29	29	58
11	1	1	2
12	49	36	85
13	1	0	1
14	2	2	4
15	13	10	23
16	2	4	6
17	0	1	1
18	0	2	2
20	6	13	19
24	6	3	9
25	5	3	8
26	2	0	2

30	4	5	9
31	7	6	13
Total	2807	1424	4231

. tab ut_q2a hours

Number of Visits to general practitioner Past year	hours standard	Long work	Total
0	935	180	1115
1	878	189	1067
2	580	134	714
3	334	77	411
4	208	49	257
5	106	26	132
6	150	34	184
7	26	11	37
8	54	10	64
9	7	1	8
10	49	9	58
11	2	0	2
12	66	19	85
13	1	0	1
14	4	0	4
15	21	2	23
16	5	1	6
17	0	1	1

18	|	2	0	| 2
20	|	15	4	| 19
24	|	7	2	| 9
25	|	6	2	| 8
26	|	2	0	| 2
30	|	6	3	| 9
31	|	10	3	| 13
Total	|	3474	757	| 4231

end of do-file

-> . regress ut_q2a longwh

Source	SS	df	MS	
				Number of obs =
				F(1, 4229) = 2.19
Model	34.0889489	1	34.0889489	Prob > F = 0.1390
Residual	65834.5653	4229	15.5674073	R-squared = 0.0005
				Adj R-squared=0.0003
Total	65868.6542	4230	15.5717859	RootMSE=3.9456

ut_q2a	Coef.	Std.Err.	t	P>|t|	[95%Conf.Interval]	
longwh	.2341884	.1582584	1.480	0.139	-.0760811	.544458
_cons	2.574266	.0669412	38.456	0.000	2.443026	2.705506

-> . test longwh

(1) longwh = 0.0

F(1, 4229) = 2.19

Prob > F = 0.1390

> . summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
QEDid	4237	8925.158	5111.942	12	17625
agegrp	4237	6.261034	1.659756	4	9

sex	|	4237	1.336795	.4726699	1	2
marstatg	|	4237	1.502242	.7162667	1	3
hhsizeg	|	4237	2.563606	1.261969	1	5
numle5g	|	4237	1.839509	.3671041	1	2
num6t11g	|	4237	1.847062	.3599706	1	2
ut_q2a	|	4231	2.616166	3.94611	0	31
dvedc294	|	4232	7.237713	3.089956	2	12
lfs_q1	|	4237	2	0	2	2
dvwh94	|	4237	1.599481	1.290932	1	5
inc_q2g	|	4231	1.101867	.531534	1	6
dvhhin94	|	4237	8.026434	2.128821	1	11
dvsmkt94	|	4236	3.773607	2.078046	1	6
somesec	|	4237	.3190937	.4661804	0	1
somepost	|	4237	.4682558	.4990502	0	1
postsec	|	4237	.2126505	.4092307	0	1
school	|	4237	1.893557	.7214824	1	3
lowest	|	4237	.0271418	.1625158	0	1
lowmid	|	4237	.0573519	.2325412	0	1
mid	|	4237	.2428605	.4288621	0	1
upmid	|	4237	.4647156	.4988123	0	1
highest	|	4237	.2079301	.4058745	0	1
inc	|	4237	3.76894	.9334302	1	5
standard	|	4237	.8210998	.3833139	0	1
longwh	|	4237	.1789002	.3833139	0	1
hours	|	4237	1.1789	.3833139	1	2
male	|	4237	.6632051	.4726699	0	1
female	|	4237	.3367949	.4726699	0	1

smoker	|	4237	.3054048	.460633	0	1

-> . regress ut_q2a somesec somepost postsec female lowest lowmid mid upmid highest standard longwh

Source | SS df MS Numberofobs=4231

F(8, 4222) = 11.86

Model | 1447.24022 8 180.905028 Prob > F = 0.0000

Residual | 64421.414 4222 15.2585064 R-squared = 0.0220

AdjR-squared=0.0201

Total | 65868.6542 4230 15.5717859 Root MSE = 3.9062

ut_q2a	|	Coef.	Std.Err.	t	P>|t|	[95%Conf.Interval]	
somesec	|	(dropped)					
somepost	|	-.0279715	.1389673	-0.201	0.840	-.3004205	.2444774
postsec	|	-.1385647	.1757837	-0.788	0.431	-.4831933	.2060638
female	|	1.19053	.1279169	9.307	0.000	.9397451	1.441314
lowest	|	(dropped)					
lowmid	|	-.4089956	.4429538	-0.923	0.356	-1.277418	.4594268
mid	|	-.678305	.3845172	-1.764	0.078	-1.432161	.0755509
upmid	|	-.6003574	.3754849	-1.599	0.110	-1.336505	.1357905
highest	|	-.6369635	.39045	-1.631	0.103	-1.402451	.1285238
standard	|	-.2414124	.1578991	-1.529	0.126	-.5509778	.068153
longwh	|	(dropped)					
_cons	|	3.055879	.3984067	7.670	0.000	2.274793	3.836966

-> . exit

Session ended at 25 Mar 2001; 16:36:31

****

Project 1

Queen's University

Department of Economics

Economics 452*

Project II.1

Age-Wage Gap Between Younger & Older Workers: A Multiple Linear Regression Analysis on SCF(1995) data

MARCH 2001

I Introduction

Researches have shown that since the late 1970s, the real earnings among younger workers are declining. Longitudinal studies suggest that the decline of earnings among young workers is persistent. That is, younger workers will not be able to earn as much in the future, as their older cohorts are earning now (Morissette, 1997; Beaudry and Green, 1997). Hence, the gap in earnings between younger and older workers is growing.

Traditionally, younger workers enjoy an education premium over their older counterparts, while older workers benefited from greater experiences. Changes in *education* and *experience* of workers from different age groups would thus definitely affect the age-wage gap among younger and older workers. Recent observations have shown that the relative education premium enjoyed by the younger workers has largely vanished. Older male workers' education level has almost caught up with their younger counterparts during mid-90s. The relative decline in educational attainment among younger workers over the past two decades would be one of the major factors contributing to the widening of the age-wage gap.

Kapsalis, Morissette and Picot, in their paper "*The Returns to Education, and the Increasing Wage Gap between Younger and Older Workers*", attempted to show that changes in relative educational attainment between younger and older workers, could strongly affect the age-wage gap. They used a regression decomposition approach to study the changes in the age-wage gap over the 1981-1995 period, by employing data from various years of Surveys of Consumer Finance (SCF), Surveys of Work History (SWH) and the Labor Market Activity Surveys (LMAS). They found that during the 80s, the growth in the relative educational attainment of older workers has contributed to about one-quarter of the increase in the age-wage gap of both male and female workers, while the gap increased to a much lesser extent in the 1990s.

In addition, they also attempted to observe the trends of the expected real wages for younger workers over the studied period. Although the educational attainment of younger workers has been rising throughout the 1981-1995 period, their real hourly wages and annual earnings have been falling. This suggests that the real expected wages for younger workers with any level of education might have been falling. Using a wage equation that controls for changes in other characteristics such as industry of employment, full-time part-time status and region, they found that during the 1980s, the expected weekly wages associated with all levels of education fell for younger workers of both genders.

This project reviews the work done by Kapsalis, Morissette and Garnett (1999), and attempts to use a similar set of data to perform a multiple linear regression for the age-wage gap for both female and male workers. In Section II, we describe the data and variables employed in our regression, and compare that with data used in the paper. We briefly describe our model and methodology in Section III. In Section IV, we report and interpret our results and compare them to the paper. Section V is a concluding section where summary of our work and additional, final remarks will be addressed.

II The Data -- SCF (1995)

Kapsalis, Morissette and Picot (1999) used two sets of data in their research. They used a series of the Survey of Consumer Finance (SCF) to examine weekly earnings (wagsal) over the 1981-95 period. They also used a combination of resources obtained from the Survey of Work History (SWH) and the Labor Market Activity Survey (LMAS) to examine both weekly earnings and hourly wages over the 1981-1988 period. Kapsalis, Morissette and Picot used the SCF to carry out analysis for two separate periods: 1981-88 and 1989-1995. When they used the SCF data, they restricted their sample to workers with *positive* weekly earnings and no self-employment income.

Age is one of the major variables in the regression. The authors of the paper restrict their attention to two age groups: individuals aged 25-34 and 45-54, in order to keep the wage comparisons tractable. They excluded workers in the age group of 18-24 to avoid problems associated with shifting patterns over time in the rates of school attendance and part-time employment. Furthermore, at any point in time, a significant proportion of 18-24 years olds are still in school and therefore a very small portion of them are strongly attached to the labor market. Individuals aged 45-54 was selected because changing patterns of early retirements among the 55-64 year old population may also influence the results by changing in the composition of workers in the sample over time.

Since education is highly correlated to earnings, summary of education level (receduc) was included in the regression. Moreover, in order to obtain the expected real wages, the authors control for changes in characteristics like industry of employment (occ13), full-time part-time status (wrkft_pt) and region (prov) in their wage equation.

Since the main objective of our project is to pursue the work done in the paper, we attempted to follow the authors' sample selection criteria closely. Unfortunately, due to limited resources from the QED DLI data archive, we could only obtain SCF data for year 1995. The SWH-LMAS data was not available for this project. We carefully picked the variables used in the paper from the 1995 "SCF - economic families" dataset for our regression. Variables used in our regression are almost the same as those employed in the paper. A comparison of the variables used in this project and the paper was stated in

Table 1 of Appendix II. However, selections of dummy variables like occ13 and prov were not strictly followed, further details will be explained in the following section.

III The Multiple Linear Regression Model

The main objective of this project is to construct a multiple linear regression model that is similar to the regression performed by Kapsalis, Morissette and Picot (1999). In the paper, they tried to determine the extent to which the improvement in the relative educational attainment of older workers accounts for the growth in their relative wages. To do so, they used a regression decomposition technique. This allows them to decompose the change in the wage gap between young and older workers into two components: 1) changes in the characteristics of workers employed in the two age groups and, 2) changes in the expected returns to these characteristics.

In the paper, the authors setup log wage (*ln w*) equations for the younger and older workers (25-34 & 45-54) with education, province, occupation and full-/part-time work status as the control variables. The wage equation for age group j is in the form of:

$$Y_{it}^{j} = X_{it}^{j} b^{j} + X_{it}^{j} D_{95} d^{j} + u_{it}^{j} \qquad (1)$$

where Y_{it}^{j} is the log earnings of the i^{th} individual of age group j in year 1995, X_{it}^{j} are control variables, D_{95} is a dummy variable which equals one in 1995, zero otherwise, and u_{it}^{j} is a random term. Our controls consist of dummy variables for seven education levels, five regions, five industrial groups and full-time/part-time status. This is exactly what the authors have done in their paper. On top of that, they also used a method of decomposition suggested by Blinder (1973) and Oaxaca (1973), to obtain the difference in mean log earnings across for periods 1981-1988 and 1989-1995[1]. However, we cannot replicate their work, as there is only one survey dataset (SCF 1995) available in the QED DLI data archive.

Yet, their regression (Equation (1)) is inefficient in a way that they constructed 2 separate regressions the 2 age groups for each gender and all together 4 regressions for both sexes. We improved their work by pooling the data for the two age groups and obtain one wage model for each gender. The *pooled full-interaction regression* wage equation for age group *j* and *k* would be in the form of:

$$Y_i^j = X_i^j b^j + Z_i^j g^j + u_i^j \qquad (2)$$

where Y_i^j is the log earnings of the i^{th} individual of age group j in year 1995, b^j is the coefficient for the control variables X_i^j, and γ^j is the coefficient for the interaction terms Z_i^j. The difference between analyzing two age groups separately and making them into one regression is that we have now constrained the variance of u for both age groups to be equal. We used STATA, software for statistical analysis, to perform the regression and the results are reported and interpreted in the following section.

IV Results

We started our regression by generating dummy variables from the raw data (SCF - 1995). The variable sex was generated into 2 dummy variables, male and female. Out of these two gender groups, we created three dummy age groups (24-34, 45-54 and all others) for both sexes. Other required dummy variables (receduc, wrkft_pt, occ13, prov) are

[1] Sample formula for Difference in Mean Log Earnings between 1981-1988 for workers of age group j:
$\bar{Y}_{88}^{j} - \bar{Y}_{81}^{j} = (\bar{b}^{j} + \bar{d}^{j})(\bar{X}_{88}^{j} - \bar{X}_{81}^{j}) + \bar{X}_{81}^{j}\bar{d}^{j}$. Where $\bar{b}^{j} + \bar{d}^{j} = b_{88}^{j}$ and that $\beta^{j} = \beta_{81}^{j}$.

also generated from the raw data in the same way, and weekly earnings (wagsal) is the dependent variable with all the negative values removed from the data.

First of all, we perform the analysis for male workers only, by creating a dataset for log wages restricted to observations in the two age groups. The dataset consists of 4583 observations for age 25-34 and 4843 observations for age 45-54. We set age group 25-34 as the base group and run a pooled regression as illustrated in Eq. (2) by constraining the variances (u). Results are shown in Table 2 of Appendix II.

By simply looking at the pairwise coefficient differences (or, interaction terms) between the two age groups (γ), we discovered that there *is* an age-wage gap since these coefficient estimates do not equal to zero. Moreover, we can observe, to what extent, does each of the coefficients contribute to the gap. A negative estimated coefficient of an interaction term represents an age premium that the variable favors the younger age group over the old, and vice versa. For example, the coefficient for Old*no_school is negative for both sexes. This indicates that no school contributes more to the decrease in older male workers' earnings than younger workers.

We confirm our conjecture by performing a hypothesis test. We test for full coefficient equality by stating the hypothesis,

$$H_0: \gamma_j = \beta_j^{25\text{-}34} - \beta_j^{45\text{-}54} = 0 \qquad \forall\, j = 1, 2, 3, \ldots, k$$
$$H_A: \gamma_j = \beta_j^{25\text{-}34} - \beta_j^{45\text{-}54} \neq 0 \qquad j = 1, 2, 3, \ldots, k$$

where β is the coefficient estimate for the corresponding age group and k is the total number of regression coefficients in the unrestricted model, which equals to 17 for the male regression. We reject the null since the F-value was 23.90 with a p-value of 0. Hence, we conclude that there *is* an age-wage gap between younger and older male workers in year 1995, and decided to keep our model unrestricted, which was implied by the alternative hypothesis. Hence, we rejected the restricted model where all interactive coefficients equal 0. Our final unrestricted pooled full-interaction regression equation for 1995 male:

*ln w_i = 8.983557 - .3040725 no_school -.1233848 Gr.9-10 .0825255 Gr.11-13 + .2023544 Post-secondary diploma + .2856605 University + .1806115 Quebec + .3412613 Ontario + .1661092 Manitoba/Saskatchewan + .2561718 Alberta + .2168915 B.C. + .1700787 Occ-Manager/Admin. + .0022088 Occ-Sales - .0255549 Occ-Services - .08568 Occ-Construction + .0410144 Occ-Transport + 1.111473 Full Time + .5192982 Old*Young - .077381 Old*no_school - .0275861 Old*Gr.9-10 + .0294106 Old*Gr.11-13 - .0636513 Old*Post-Sec. diploma + .0985861 Old*University - .1158353 Old*Quebec - .0461316 Old*Ont - .1359847 Old*Manitoba/Saskatchewan -.2404863 Old*Alberta - .039621 Old*B.C. + .1863668 Old*Manager/Admin. - .01372 Old*Sales .0103647 Old*Services + .0957005 Old*Construction + .107952Old*Transport - .1575084Old*FT*

The regression model suggests that the young has an age premium over their older counterparts, since most of the values of the pairwise coefficient difference ($\hat{g}$) are negative. We also observed that education plays an important role in affecting the size of the gap since the coefficient estimate for the interaction term for university education suggests that older males enjoy higher rewards from university education than younger male workers. This matches with the paper's conclusion.

In order to test for the linear coefficient restrictions, we perform a general F-statistics for the significance of the restricted and unrestricted model:

$$F = \frac{(R_U^2 - R_R^2)/(k - k_0)}{(1 - R_U^2)/(N - k)} \qquad (3)$$

where:

R_U^2 = the R-squared for the *unrestricted* model = 0.2087;
R_R^2 = the R-squared for the *restricted* model = 0.0.1682;
k_0 = the number of free regression coefficients in the *restricted* model = 17;
k = the number of free regression coefficients in the *unrestricted* model = 34;
k - k_0 = the number of independent linear coefficient restrictions specified by the null hypothesis H_0 = 17;
N - k = the degrees of freedom for RSS_1, the *unrestricted* RSS = 9392.

The unrestricted R^2 is *greater* than the restricted R^2. The F-value calculated is 28.27[2] with a p-value of 0. The F-statistic in effect determines whether imposing the coefficient

[2] The corresponding F-value is calculated using Equation (3).

restrictions specified by the null hypothesis H_0 significantly reduces the coefficient of determination, R^2. Due to the p-value for the F-test is 0, we reject the null, which means the unrestricted model is more significant than the restricted model.

We regressed our model with the assumption that the variances (*u*) for both age groups are equal (constrained variances). If *u* is known to have the same variance in the two groups, the standard errors obtained from the pooled regression are better -- they are more efficient. However, if the variances are actually different, then the standard errors obtained from the pooled regression are wrong! Therefore, we pooled the data again without constraining the residual variance and observe the difference between the two models. Identical results were obtained for our regression, which indicates that the variances for both age groups are the same. The F-value for the unconstrained variance model is 23.99 (p-value = 0), which is slightly larger than that for the constrained variance model (23.90 with a p-value = 0). However, the values are close enough to conclude that there is an age-wage gap between young and old workers, no matter the variances are constrained or not.

Similar work was done to the data for female workers (see Table 3 in Appendix II for regression results). 1006 female workers at age 25-34 and 829 females at age 45-54 were being tested. Results show that there is an age-wage gap among female workers between the two age groups as well, since the null hypothesis for full coefficient equality was rejected (F-value is 3.22[3] and p-value = 0). Thus, implied by the alternative hypothesis, we use the unrestricted model for our female group. The final restricted regression model for women with a base age group of 25-34 in year 1995:

---

[3] The corresponding F-value is calculated using Equation (3). The required values for the calculations are stated in Table 3 of Appendix II.

ln w_i *= 8.428428 -.4165621No_schooling + .0569279Gr.9-10 + .2257894Gr.11-13 + .4069217Post-Secondary Diploma + .529994University + .2944867Quebec + .3906155Ontario .121286 Manitoba/Saskatchewan + .3609925Alberta + .4615904B.C. + .1613444Occ-Manager/Admin. -.1464164Occ-Sales - .5164822Occ-Services -.2654534Occ-Transport + .9360749Full-time + .3953911Old*Young + .055615Old*no_schooling - .4546646Old*Gr.9-10 - .2653172Old*Gr.11-13 - .3931499Old*Post-Sec. diploma - .2842871Old*University - .1587404Old*Quebec - .0977524 Old*Ont + .105598Manitoba/Saskatchewan - .1051346Old*B.C. + .1938023Old*Manager/Admin. - .0948978Old*Sales + .2757955Old*Services - .0531885Old*Transportation + .2836946 Old*Transport*

It was found that Education favors workers in the younger age group since the interaction terms for almost *all* education levels are negative. We also performed F-tests for the significance of the constrained and unconstrained variance models, and the F-values are close enough to conclude that it does not matter whether the variance are constrained or not (F-values: 4.89 vs 4.94).

It is difficult to compare our results with those generated by the authors owing to the limited resource of data. They obtained the percentage changes in the log wage gap by comparing the ln weekly wages of the two age groups between 2 time periods (1981-88 & 1989-95), this is what we cannot do since we only have one year of SCF data.

V Conclusion

In this project, we ran a multiple linear regression model by tightly following the sample selection criteria of the paper "The Returns to Education, and the Increasing Wage Gap Between Younger and Older Workers" by Kapsalis, Morissette and Picot (1999). The main purpose of the paper was to observe the change in the age-wage gap throughout two time periods: 1981-1988 and 1989-1995 using series of SCF data, and results were confirmed by performing a similar test using different set of data (SWH-LMAS). Due to limited sources of data, we only construct a wage model for male and female workers in year 1995 for this project. Special attention should be paid to the big differences in sample sizes between the two genders since sample size affects accuracy of estimation and the larger the sample size, the more closer the estimated results to the real value.

We improve the regression in the paper by setting up one single pooled full-interaction regression wage equation instead of 2 separate multiple linear regression equations for the two age groups. The interaction terms reveal whether an age-wage gap exists and the effect of each variable in the pooled regression on the gap.

We set up a hypothesis to test whether the gap exists between older and younger workers by assuming all interaction terms (γ) being zero. We rejected the null (p-value = 0) and concluded that there is a wage gap between the two age groups of both genders. We also found that younger workers enjoy an education premium over older workers. This is in particularly reflected in the data for female workers since the γ's for almost all education levels are negative. As mentioned in the introduction section, the paper declares that the education premium for younger male workers is disappearing. This matches with our regression results since the γ for male with university education is positive (.0985861) which means that university education does a positive effect to the wage of older male workers over their younger cohorts.

We found that the work done by Kapsalis, Morissette and Picot was imprudent, inefficient and unprofessional. Some of their simple additions and subtractions in the decomposition were miscalculated, which led to their misinterpretation of results for the expected earnings of young workers. Fortunately, we did not replicate this part of their work and thus our results do not contradict with the paper. However, this kind of careless mistake should not be found in a professional research paper. Moreover, their work was inefficient and confusing as they could have simplify their methodology by reducing the chunky tables of results into several more effective and representable charts. In addition, the variable for wage and salary in the 1995 SCF data was ambiguously defined. One of the observations in the wage/salary data has a value of negative billion, which seems impossible and hard to explain.

Bibliography

Abbott, M. **2000**. *Notes for Economics 351* - Introductory Econometrics*, Queen's University.

Gujarati, Damodar N. **1995**. *Basic Econometrics*, 3rd Edition. McGraw Hill.

Kapsalis, C, R. Morissette, Garnett Picot. **1999**. "*The Returns to Education, and the Increasing Wage Gap Between Young and Older Workers*", Analytical Studies Branch - Research Paper Series, Catalogue No. 11F0019MPE-131, Ottawa Statistical Canada.

Appendices

Appendix I - Log Files

```
do "C:\WINDOWS\TEMP\STD010000.tmp"

. /*project 1 Ferrall*/
. set more off

.
. use wagsal age sex receduc wrkft_pt occ13 prov using c:\ass1.dta
(363 : scfef95 : survey of consumer finance - economic families)

. gen lnw = ln(wagsal)
(10137 missing values generated)

. tab sex, gen(dsex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      25877       75.45       75.45
     female |       8419       24.55      100.00
------------+-----------------------------------
      Total |      34296      100.00

. tab receduc, gen(deduc)

             summary education level |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
    no schooling or grade 8 or lower |       5773       16.83       16.83
                          grade 9-10 |       4257       12.41       29.25
            grade 11-13 not graduate |       2004        5.84       35.09
               grade 11-13 garaduate |       5923       17.27       52.36
some post-secondary no dipl,deg,cert |       2421        7.06       59.42
         post-secondary cert or dipl |       9375       27.34       86.75
                   university degree |       4543       13.25      100.00
-------------------------------------+-----------------------------------
                               Total |      34296      100.00

. tab wrkft_pt, gen(dftpt)

    worked mostly full or part |
        time in reference year |      Freq.     Percent        Cum.
-------------------------------+-----------------------------------
                     full-time |      21053       61.39       61.39
                     part-time |       2506        7.31       68.69
did not work in reference year |      10732       31.29       99.99
                             7 |          1        0.00       99.99
                            20 |          1        0.00       99.99
                            27 |          1        0.00       99.99
                            40 |          1        0.00      100.00
                            41 |          1        0.00      100.00
-------------------------------+-----------------------------------
                         Total |      34296      100.00
```

```
. tab occ13, gen(docc)

  1980 occupational classification - 13 |
                                 groups |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
          managerial and administrative |       3353        9.78        9.78
                       natural sciences |       2986        8.71       18.48
                               teaching |       1010        2.94       21.43
                               clerical |       1868        5.45       26.87
                                  sales |       2210        6.44       33.32
                               services |       3269        9.53       42.85
farming,fishing,forestry and logging op |       1845        5.38       48.23
        mining,processing and machining |       1883        5.49       53.72
product fabricating,assembling and repa |       2547        7.43       61.15
                    construction trades |       2446        7.13       68.28
transport,material handling,other craft |       2669        7.78       76.06
                    never worked before |        947        2.76       78.82
      last worked more than 5 years ago |       7263       21.18      100.00
----------------------------------------+-----------------------------------
                                  Total |      34296      100.00

. tab prov, gen(dprov)

           province |      Freq.     Percent        Cum.
--------------------+-----------------------------------
 special family unit |          8        0.02        0.02
        newfoundland |       1213        3.55        3.57
prince edward island |        881        2.58        6.15
         nova scotia |       2254        6.60       12.75
       new brunswick |       2072        6.06       18.81
              quebec |       6868       20.10       38.92
             ontario |      10414       30.48       69.40
            manitoba |       2436        7.13       76.53
        saskatchewan |       2205        6.45       82.98
             alberta |       2649        7.75       90.74
    british columbia |       3165        9.26      100.00
--------------------+-----------------------------------
               Total |      34165      100.00

. gen dprov12 = 0

. replace dprov12 = 1 if  prov == 46|  prov==47
(4641 real changes made)

. label var dprov12 "Man/Sask"

. drop if wagsal < 0
(4 observations deleted)
.
. /*create dummy age group*/
. gen ageg = 0

. replace ageg =1 if age > 24 & age <35
(6314 real changes made)

. replace ageg = 2 if age >44 & age < 55
(6459 real changes made)
```

```
. tab ageg, gen(dageg)

       ageg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      21519       62.75       62.75
          1 |       6314       18.41       81.16
          2 |       6459       18.84      100.00
------------+-----------------------------------
      Total |      34292      100.00

.
. /*pool*/
. gen g2 = (ageg==2)

.
. /*for men*/
. drop if sex ==2
(8418 observations deleted)

.
.
. /*for education group 2*/
. gen g2deduc1 = g2 * deduc1

. gen g2deduc2 = g2 * deduc2

. gen g2deduc3 = g2 * deduc3

. gen g2deduc4 = g2 * deduc4

. gen g2deduc5 = g2 * deduc5

. gen g2deduc6 = g2 * deduc6

. gen g2deduc7 = g2 * deduc7

.
. /*for prov group 2*/
. gen g2dprov6 = g2 * dprov6
(101 missing values generated)

. gen g2dprov7 = g2 * dprov7
(101 missing values generated)

. gen g2dpro10 = g2 * dprov10
(101 missing values generated)

. gen g2dpro11 = g2 * dprov11
(101 missing values generated)

. gen g2dpro12 = g2 * dprov12

.
. /*occupation for group 2*/
. gen g2docc1 = g2 *docc1
. gen g2docc5 = g2 *docc5
. gen g2docc6 = g2 *docc6
. gen g2docc10 = g2 *docc10
. gen g2docc11 = g2 *docc11
.
. /*group 2 for full-time part-time*/
. gen g2ftpt = g2 * dftpt1

.
```

```
. /*TRY separately*/
. regress  lnw deduc1 deduc2 deduc4 deduc6 deduc7  dprov6 dprov7 dprov12 dprov1
> 0 dprov11 docc1 docc5  docc6 docc10 docc11 dftpt1 if ageg ==1

  Source |       SS       df       MS              Number of obs =    4583
---------+------------------------------           F( 16,  4566) =   57.71
   Model |  622.340189    16  38.8962618           Prob > F      =  0.0000
Residual |  3077.60238  4566  .674025927           R-squared     =  0.1682
---------+------------------------------           Adj R-squared =  0.1653
   Total |  3699.94257  4582  .807495105           Root MSE      =  .82099

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.3040725   .0795197     -3.824   0.000      -.4599695   -.1481754
  deduc2 |  -.1233848   .0534782     -2.307   0.021       -.228228   -.0185416
  deduc4 |   .0825255   .0413013      1.998   0.046        .001555    .1634959
  deduc6 |   .2023544   .0380835      5.313   0.000       .1276924    .2770164
  deduc7 |   .2856605   .0448405      6.371   0.000       .1977515    .3735694
  dprov6 |   .1806115   .0395862      4.562   0.000       .1030034    .2582196
  dprov7 |   .3412613   .0368503      9.261   0.000       .2690169    .4135057
 dprov12 |   .1661092   .0441241      3.765   0.000       .0796047    .2526137
 dprov10 |   .2561718   .0502085      5.102   0.000        .157739    .3546047
 dprov11 |   .2168915   .0496179      4.371   0.000       .1196165    .3141665
   docc1 |   .1700787   .0405803      4.191   0.000       .0905217    .2496357
   docc5 |   .0022088   .0468683      0.047   0.962      -.0896758    .0940934
   docc6 |  -.0255549   .0416638     -0.613   0.540      -.1072361    .0561262
  docc10 |    -.08568   .0409866     -2.090   0.037      -.1660335   -.0053264
  docc11 |   .0410144   .0399842      1.026   0.305      -.0373739    .1194028
  dftpt1 |   1.111473    .045351     24.508   0.000       1.022563    1.200383
   _cons |   8.983557   .0609666    147.352   0.000       8.864033    9.103081
------------------------------------------------------------------------------

. regress  lnw deduc1 deduc2 deduc4 deduc6 deduc7  dprov6 dprov7 dprov12 dprov1
> 0 dprov11 docc1 docc5  docc6 docc10 docc11 dftpt1 if ageg ==2

  Source |       SS       df       MS              Number of obs =    4843
---------+------------------------------           F( 16,  4826) =   79.61
   Model |  951.097295    16  59.4435809           Prob > F      =  0.0000
Residual |  3603.69954  4826  .746725972           R-squared     =  0.2088
---------+------------------------------           Adj R-squared =  0.2062
   Total |  4554.79684  4842  .940685014           Root MSE      =  .86413

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.3814534   .0525549     -7.258   0.000      -.4844849   -.2784219
  deduc2 |  -.1509709   .0519265     -2.907   0.004      -.2527704   -.0491714
  deduc4 |   .1119361   .0490953      2.280   0.023        .015687    .2081852
  deduc6 |   .1387031   .0436016      3.181   0.001       .0532241    .2241821
  deduc7 |   .3842466   .0484108      7.937   0.000       .2893394    .4791538
  dprov6 |   .0647762   .0390343      1.659   0.097      -.0117487    .1413012
  dprov7 |   .2951297   .0360458      8.188   0.000       .2244636    .3657959
 dprov12 |   .0301245   .0444138      0.678   0.498      -.0569468    .1171958
 dprov10 |   .0156856   .0530767      0.296   0.768      -.0883689    .1197401
 dprov11 |   .1772705   .0516807      3.430   0.001       .0759527    .2785883
   docc1 |   .3564455   .0369045      9.659   0.000       .2840959    .4287952
   docc5 |  -.0115112   .0489788     -0.235   0.814       -.107532    .0845097
   docc6 |  -.0151903   .0456884     -0.332   0.740      -.1047603    .0743798
  docc10 |   .0100205   .0433385      0.231   0.817      -.0749428    .0949837
  docc11 |   .1489664   .0422807      3.523   0.000       .0660769    .2318559
  dftpt1 |   .9539645   .0429423     22.215   0.000        .869778    1.038151
   _cons |   9.502856   .0594369    159.881   0.000       9.386332    9.619379
------------------------------------------------------------------------------
```

```
.
. /*pool*/
. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6 docc10 docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4
> g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc
> 5 g2docc6 g2docc10 g2docc11 g2ftpt if ageg ==1 |ageg==2

  Source |       SS       df       MS              Number of obs =    9426
---------+------------------------------           F( 33,  9392) =   75.08
   Model |  1762.54059    33  53.4103209           Prob > F      =  0.0000
Residual |  6681.30192  9392  .711382232           R-squared     =  0.2087
---------+------------------------------           Adj R-squared =  0.2060
   Total |  8443.84251  9425   .89589841           Root MSE      =  .84343

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 | -.3040725   .0816936     -3.722   0.000      -.4642096   -.1439353
  deduc2 | -.1233848   .0549402     -2.246   0.025      -.2310795   -.0156901
  deduc4 |  .0825255   .0424303      1.945   0.052      -.0006472    .1656982
  deduc6 |  .2023544   .0391246      5.172   0.000       .1256618    .2790471
  deduc7 |  .2856605   .0460663      6.201   0.000       .1953606    .3759604
  dprov6 |  .1806115   .0406684      4.441   0.000       .1008926    .2603304
  dprov7 |  .3412613   .0378577      9.014   0.000        .267052    .4154705
 dprov12 |  .1661092   .0453303      3.664   0.000        .077252    .2549664
 dprov10 |  .2561718   .0515811      4.966   0.000       .1550618    .3572819
 dprov11 |  .2168915   .0509743      4.255   0.000       .1169708    .3168121
   docc1 |  .1700787   .0416896      4.080   0.000        .088358    .2517994
   docc5 |  .0022088   .0481496      0.046   0.963      -.0921748    .0965925
   docc6 | -.0255549   .0428028     -0.597   0.550      -.1094576    .0583478
  docc10 |   -.08568   .0421071     -2.035   0.042       -.168219    -.003141
  docc11 |  .0410144   .0410773      0.998   0.318      -.0395059    .1215348
  dftpt1 |  1.111473   .0465908     23.856   0.000       1.020145    1.202801
      g2 |  .5192982   .0853725      6.083   0.000       .3519497    .6866467
g2deduc1 |  -.077381   .0964631     -0.802   0.422      -.2664695    .1117076
g2deduc2 | -.0275861   .0747473     -0.369   0.712       -.174107    .1189348
g2deduc4 |  .0294106   .0640046      0.460   0.646      -.0960524    .1548736
g2deduc6 | -.0636513   .0578087     -1.101   0.271      -.1769689    .0496664
g2deduc7 |  .0985861   .0659908      1.494   0.135      -.0307701    .2279423
g2dprov6 | -.1158353   .0557268     -2.079   0.038      -.2250719   -.0065987
g2dprov7 | -.0461316   .0516818     -0.893   0.372       -.147439    .0551759
g2dpro12 | -.1359847   .0627221     -2.168   0.030      -.2589335   -.0130359
g2dpro10 | -.2404863   .0731054     -3.290   0.001      -.3837887   -.0971839
g2dpro11 |  -.039621   .0717137     -0.552   0.581      -.1801955    .1009534
 g2docc1 |  .1863668   .0550954      3.383   0.001       .0783678    .2943658
 g2docc5 |   -.01372   .0678511     -0.202   0.840      -.1467228    .1192827
 g2docc6 |  .0103647   .0618119      0.168   0.867         -.1108    .1315293
g2docc10 |  .0957005   .0596853      1.603   0.109      -.0212956    .2126966
g2docc11 |   .107952    .058227      1.854   0.064      -.0061857    .2220896
  g2ftpt | -.1575084   .0626695     -2.513   0.012      -.2803541   -.0346626
   _cons |  8.983557   .0626332    143.431   0.000       8.860783    9.106332
------------------------------------------------------------------------------

.
```

```
. /*Hypothese Test*/
. /*test whether or not there is a wage-age gap*/
. test g2  g2deduc1 g2deduc2 g2deduc4 g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpr
> o12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2docc6 g2docc10 g2docc11 g2ftpt

 ( 1)  g2 = 0.0
 ( 2)  g2deduc1 = 0.0
 ( 3)  g2deduc2 = 0.0
 ( 4)  g2deduc4 = 0.0
 ( 5)  g2deduc6 = 0.0
 ( 6)  g2deduc7 = 0.0
 ( 7)  g2dprov6 = 0.0
 ( 8)  g2dprov7 = 0.0
 ( 9)  g2dpro12 = 0.0
 (10)  g2dpro10 = 0.0
 (11)  g2dpro11 = 0.0
 (12)  g2docc1 = 0.0
 (13)  g2docc5 = 0.0
 (14)  g2docc6 = 0.0
 (15)  g2docc10 = 0.0
 (16)  g2docc11 = 0.0
 (17)  g2ftpt = 0.0

       F( 17,  9392) =   23.90
            Prob > F =    0.0000


.
. /*for prediction which may not present in the paper*/
. /*contain the standard error of linear prediction X_j*b */
. predict stdp
(option xb assumed; fitted values)
(101 missing values generated)

. predict yhatmun, stdp
(101 missing values generated)

. sum yhatmun

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
 yhatmun |   25773    .0566266   .0156076   .0309614   .1049903

.
. /*Restricted model*/
.Hence the age group ==1

.
. /*Run the same process again for women*/
. use wagsal age sex receduc wrkft_pt occ13 prov using c:\ass1.dta
(363 : scfef95 : survey of consumer finance - economic families)

. gen lnw = ln(wagsal)
(10137 missing values generated)

. tab sex, gen(dsex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      25877       75.45       75.45
     female |       8419       24.55      100.00
------------+-----------------------------------
      Total |      34296      100.00
```

```
. tab receduc, gen(deduc)

            summary education level |      Freq.     Percent        Cum.
------------------------------------+-----------------------------------
    no schooling or grade 8 or lower |       5773       16.83       16.83
                         grade 9-10 |       4257       12.41       29.25
            grade 11-13 not graduate |       2004        5.84       35.09
               grade 11-13 garaduate |       5923       17.27       52.36
some post-secondary no dipl,deg,cert |       2421        7.06       59.42
          post-secondary cert or dipl |       9375       27.34       86.75
                   university degree |       4543       13.25      100.00
------------------------------------+-----------------------------------
                               Total |      34296      100.00

. tab wrkft_pt, gen(dftpt)

    worked mostly full or part |
        time in reference year |      Freq.     Percent        Cum.
------------------------------+-----------------------------------
                     full-time |      21053       61.39       61.39
                     part-time |       2506        7.31       68.69
did not work in reference year |      10732       31.29       99.99
                             7 |          1        0.00       99.99
                            20 |          1        0.00       99.99
                            27 |          1        0.00       99.99
                            40 |          1        0.00      100.00
                            41 |          1        0.00      100.00
------------------------------+-----------------------------------
                         Total |      34296      100.00

. tab occ13, gen(docc)

  1980 occupational classification - 13 |
                                 groups |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
          managerial and administrative |       3353        9.78        9.78
                       natural sciences |       2986        8.71       18.48
                               teaching |       1010        2.94       21.43
                               clerical |       1868        5.45       26.87
                                  sales |       2210        6.44       33.32
                               services |       3269        9.53       42.85
farming,fishing,forestry and logging op |       1845        5.38       48.23
        mining,processing and machining |       1883        5.49       53.72
product fabricating,assembling and repa |       2547        7.43       61.15
                    construction trades |       2446        7.13       68.28
transport,material handling,other craft |       2669        7.78       76.06
                     never worked before |        947        2.76       78.82
        last worked more than 5 years ago |       7263       21.18      100.00
---------------------------------------+-----------------------------------
                                  Total |      34296      100.00

. tab prov, gen(dprov)

            province |      Freq.     Percent        Cum.
---------------------+-----------------------------------
  special family unit |          8        0.02        0.02
         newfoundland |       1213        3.55        3.57
prince edward island |        881        2.58        6.15
          nova scotia |       2254        6.60       12.75
        new brunswick |       2072        6.06       18.81
               quebec |       6868       20.10       38.92
              ontario |      10414       30.48       69.40
             manitoba |       2436        7.13       76.53
```

```
      saskatchewan |       2205        6.45       82.98
           alberta |       2649        7.75       90.74
  british columbia |       3165        9.26      100.00
-------------------+-----------------------------------
             Total |      34165      100.00

. gen dprov12 = 0

. replace dprov12 = 1 if  prov == 46|  prov==47
(4641 real changes made)

. label var dprov12 "Man/Sask"

. drop if wagsal < 0
(4 observations deleted)

.
.
. /*create dummy age group*/
. gen ageg = 0

. replace ageg =1 if age > 24 & age <35
(6314 real changes made)

. replace ageg = 2 if age >44 & age < 55
(6459 real changes made)

. tab ageg, gen(dageg)

       ageg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      21519       62.75       62.75
          1 |       6314       18.41       81.16
          2 |       6459       18.84      100.00
------------+-----------------------------------
      Total |      34292      100.00

.
. /*pool*/
. gen g2 = (ageg==2)

.
. /*for women*/
. drop if sex ==1
(25874 observations deleted)

.
. /*for education group 2*/
. gen g2deduc1 = g2 * deduc1

. gen g2deduc2 = g2 * deduc2

. gen g2deduc3 = g2 * deduc3

. gen g2deduc4 = g2 * deduc4

. gen g2deduc5 = g2 * deduc5

. gen g2deduc6 = g2 * deduc6

. gen g2deduc7 = g2 * deduc7

.
. /*for prov group 2*/
```

```
. gen g2dprov6 = g2 * dprov6
(30 missing values generated)

. gen g2dprov7 = g2 * dprov7
(30 missing values generated)

. gen g2dpro10 = g2 * dprov10
(30 missing values generated)

. gen g2dpro11 = g2 * dprov11
(30 missing values generated)

. gen g2dpro12 = g2 * dprov12

.
. /*occupation for group 2*/
. gen g2docc1 = g2 *docc1

. gen g2docc5 = g2 *docc5

. gen g2docc6 = g2 *docc6

. gen g2docc10 = g2 *docc10

. gen g2docc11 = g2 *docc11

.
. /*group 2 for full-time part-time*/
. gen g2ftpt = g2 * dftpt1

.
.
. /*TRY separately*/
. regress  lnw deduc1 deduc2 deduc4 deduc6 deduc7  dprov6 dprov7 dprov12 dprov1
> 0 dprov11 docc1 docc5  docc6 docc11 dftpt1 if ageg ==1

  Source |       SS       df       MS              Number of obs =    1006
---------+------------------------------           F( 15,   990) =   28.02
   Model |  351.878833    15  23.4585888           Prob > F      =  0.0000
Residual |  828.730313   990  .837101327           R-squared     =  0.2980
---------+------------------------------           Adj R-squared =  0.2874
   Total |  1180.60915  1005  1.17473547           Root MSE      =  .91493

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.4165621   .2248779     -1.852   0.064      -.8578542    .0247299
  deduc2 |   .0569279   .1510604      0.377   0.706      -.2395075    .3533633
  deduc4 |   .2257894   .0991555      2.277   0.023       .0312104    .4203684
  deduc6 |   .4069217   .0917855      4.433   0.000       .2268052    .5870383
  deduc7 |    .529994   .0992403      5.341   0.000       .3352486    .7247394
  dprov6 |   .2944867   .1014777      2.902   0.004       .0953506    .4936228
  dprov7 |   .3906155   .0878886      4.444   0.000       .2181461    .5630849
 dprov12 |    .121286   .1049884      1.155   0.248      -.0847394    .3273114
 dprov10 |   .3609925   .1165852      3.096   0.002       .1322101    .5897749
 dprov11 |   .4615904   .1139551      4.051   0.000       .2379692    .6852116
   docc1 |   .1613444   .0852017      1.894   0.059      -.0058522    .3285411
   docc5 |  -.1464164   .1164095     -1.258   0.209       -.374854    .0820213
   docc6 |  -.5164822   .0789152     -6.545   0.000      -.6713425   -.3616218
  docc11 |  -.2654534   .1775983     -1.495   0.135      -.6139657    .0830589
  dftpt1 |   .9360749    .075359     12.422   0.000       .7881932    1.083957
   _cons |   8.428428   .1202296     70.103   0.000       8.192494    8.664362
------------------------------------------------------------------------------
```

```
. regress  lnw deduc1 deduc2 deduc4 deduc6 deduc7  dprov6 dprov7 dprov12 dprov1
> 0 dprov11 docc1 docc5  docc6 docc11 dftpt1 if ageg ==2

  Source |       SS       df       MS              Number of obs =     829
---------+------------------------------           F( 15,   813) =   24.85
   Model |  337.647643    15  22.5098429           Prob > F      =  0.0000
Residual |  736.309866   813  .905670192           R-squared     =  0.3144
---------+------------------------------           Adj R-squared =  0.3017
   Total |  1073.95751   828  1.29705013           Root MSE      =  .95167

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.3609471   .1638745     -2.203   0.028      -.6826141   -.0392801
  deduc2 |  -.3977368   .1490945     -2.668   0.008      -.6903923   -.1050812
  deduc4 |  -.0395279   .1226837     -0.322   0.747       -.280342    .2012862
  deduc6 |   .0137719   .1127509      0.122   0.903      -.2075454    .2350891
  deduc7 |    .245707   .1261139      1.948   0.052      -.0018403    .4932542
  dprov6 |   .1357463   .1138358      1.192   0.233      -.0877005     .359193
  dprov7 |   .2928631   .1097151      2.669   0.008       .0775049    .5082213
 dprov12 |    .226884   .1391614      1.630   0.103      -.0462739     .500042
 dprov10 |   .2558579   .1563971      1.636   0.102      -.0511318    .5628477
 dprov11 |   .2409352   .1363009      1.768   0.077       -.026608    .5084784
   docc1 |   .3551467   .0986264      3.601   0.000       .1615543    .5487392
   docc5 |  -.2413141   .1274525     -1.893   0.059      -.4914889    .0088607
   docc6 |  -.2406866   .1020735     -2.358   0.019      -.4410453    -.040328
  docc11 |  -.3186418   .2090391     -1.524   0.128      -.7289619    .0916782
  dftpt1 |   1.219769   .0825975     14.768   0.000        1.05764    1.381899
   _cons |   8.823819   .1466059     60.187   0.000       8.536048     9.11159
------------------------------------------------------------------------------

.
. /*pool*/
. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6  docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4 g2dedu
> c6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2do
> cc6  g2docc11 g2ftpt if ageg ==1 |ageg==2

  Source |       SS       df       MS              Number of obs =    1835
---------+------------------------------           F( 31,  1803) =   27.06
   Model |   728.07756    31  23.4863729           Prob > F      =  0.0000
Residual |  1565.04018  1803  .868020066           R-squared     =  0.3175
---------+------------------------------           Adj R-squared =  0.3058
   Total |  2293.11774  1834  1.25033683           Root MSE      =  .93168

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.4165621   .2289932     -1.819   0.069       -.865682    .0325578
  deduc2 |   .0569279   .1538249      0.370   0.711      -.2447659    .3586216
  deduc4 |   .2257894     .10097      2.236   0.025       .0277588    .4238199
  deduc6 |   .4069217   .0934652      4.354   0.000       .2236102    .5902333
  deduc7 |    .529994   .1010564      5.245   0.000       .3317941    .7281939
  dprov6 |   .2944867   .1033348      2.850   0.004       .0918182    .4971552
  dprov7 |   .3906155    .089497      4.365   0.000       .2150867    .5661442
 dprov12 |    .121286   .1069098      1.134   0.257       -.088394     .330966
 dprov10 |   .3609925   .1187187      3.041   0.002       .1281518    .5938332
 dprov11 |   .4615904   .1160405      3.978   0.000       .2340025    .6891783
   docc1 |   .1613444   .0867609      1.860   0.063      -.0088181    .3315069
   docc5 |  -.1464164   .1185398     -1.235   0.217      -.3789062    .0860735
   docc6 |  -.5164822   .0803594     -6.427   0.000      -.6740895   -.3588748
  docc11 |  -.2654534   .1808484     -1.468   0.142      -.6201478     .089241
  dftpt1 |   .9360749   .0767381     12.198   0.000         .78557     1.08658
      g2 |   .3953911   .1886501      2.096   0.036       .0253953    .7653869
```

```
g2deduc1 |    .055615   .2796003      0.199   0.842      -.4927597    .6039897
g2deduc2 |  -.4546646   .2120546     -2.144   0.032      -.8705632    -.038766
g2deduc4 |  -.2653172   .1569093     -1.691   0.091      -.5730604    .0424259
g2deduc6 |  -.3931499   .1446376     -2.718   0.007      -.6768248    -.109475
g2deduc7 |  -.2842871   .1595492     -1.782   0.075      -.5972077    .0286336
g2dprov6 |  -.1587404   .1519801     -1.044   0.296      -.4568161    .1393352
g2dprov7 |  -.0977524   .1398095     -0.699   0.485       -.371958    .1764533
g2dpro12 |    .105598   .1731777      0.610   0.542       -.234052    .4452481
g2dpro10 |  -.1051346   .1937456     -0.543   0.587       -.485124    .2748549
g2dpro11 |  -.2206552   .1768361     -1.248   0.212      -.5674805    .1261701
 g2docc1 |   .1938023   .1298085      1.493   0.136      -.0607886    .4483933
 g2docc5 |  -.0948978   .1721062     -0.551   0.581      -.4324463    .2426507
 g2docc6 |   .2757955   .1282322      2.151   0.032       .0242962    .5272948
g2docc11 |  -.0531885   .2731061     -0.195   0.846      -.5888261    .4824492
   g2ftpt |   .2836946   .1114785      2.545   0.011       .0650539    .5023352
    _cons |   8.428428   .1224299     68.843   0.000       8.188309    8.668547
------------------------------------------------------------------------------

.
. /*Hypothese Test*/
. /*test whether or not there is a wage-age gap*/
. test g2  g2deduc1 g2deduc2 g2deduc4 g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpr
> o12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2docc6  g2docc11 g2ftpt

 ( 1)  g2 = 0.0
 ( 2)  g2deduc1 = 0.0
 ( 3)  g2deduc2 = 0.0
 ( 4)  g2deduc4 = 0.0
 ( 5)  g2deduc6 = 0.0
 ( 6)  g2deduc7 = 0.0
 ( 7)  g2dprov6 = 0.0
 ( 8)  g2dprov7 = 0.0
 ( 9)  g2dpro12 = 0.0
 (10)  g2dpro10 = 0.0
 (11)  g2dpro11 = 0.0
 (12)  g2docc1 = 0.0
 (13)  g2docc5 = 0.0
 (14)  g2docc6 = 0.0
 (15)  g2docc11 = 0.0
 (16)  g2ftpt = 0.0

       F( 16,  1803) =    4.94
            Prob > F =    0.0000


.
. /*for prediction which may not present in the paper*/
. /*contain the standard error of linear prediction X_j*b */
. predict stdp
(option xb assumed; fitted values)
(30 missing values generated)

. predict yhatfun, stdp
(30 missing values generated)

. sum yhatfun

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
 yhatfun |    8388    .1444209   .0511473    .072206   .2845741

.
. /*Restricted model*/
Hence the age group == 1
```

```
.

. /*end project, but the following is for unconstraint variance model
> summary of constraint v.s unconstranint is discusssed at the end*/
.
. /*First begin with men*/
. use wagsal age sex receduc wrkft_pt occ13 prov using c:\ass1.dta
(363 : scfef95 : survey of consumer finance - economic families)

. gen lnw = ln(wagsal)
(10137 missing values generated)

. tab sex, gen(dsex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      25877       75.45       75.45
     female |       8419       24.55      100.00
------------+-----------------------------------
      Total |      34296      100.00

. tab receduc, gen(deduc)

             summary education level |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
    no schooling or grade 8 or lower |       5773       16.83       16.83
                          grade 9-10 |       4257       12.41       29.25
            grade 11-13 not graduate |       2004        5.84       35.09
               grade 11-13 garaduate |       5923       17.27       52.36
some post-secondary no dipl,deg,cert |       2421        7.06       59.42
         post-secondary cert or dipl |       9375       27.34       86.75
                   university degree |       4543       13.25      100.00
-------------------------------------+-----------------------------------
                               Total |      34296      100.00

. tab wrkft_pt, gen(dftpt)

    worked mostly full or part |
        time in reference year |      Freq.     Percent        Cum.
-------------------------------+-----------------------------------
                     full-time |      21053       61.39       61.39
                     part-time |       2506        7.31       68.69
did not work in reference year |      10732       31.29       99.99
                             7 |          1        0.00       99.99
                            20 |          1        0.00       99.99
                            27 |          1        0.00       99.99
                            40 |          1        0.00      100.00
                            41 |          1        0.00      100.00
-------------------------------+-----------------------------------
                         Total |      34296      100.00

. tab occ13, gen(docc)

  1980 occupational classification - 13 |
                                 groups |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
          managerial and administrative |       3353        9.78        9.78
                       natural sciences |       2986        8.71       18.48
                               teaching |       1010        2.94       21.43
                               clerical |       1868        5.45       26.87
                                  sales |       2210        6.44       33.32
                               services |       3269        9.53       42.85
```

```
farming,fishing,forestry and logging op |       1845        5.38       48.23
       mining,processing and machining |       1883        5.49       53.72
product fabricating,assembling and repa |       2547        7.43       61.15
                    construction trades |       2446        7.13       68.28
transport,material handling,other craft |       2669        7.78       76.06
                    never worked before |        947        2.76       78.82
      last worked more than 5 years ago |       7263       21.18      100.00
----------------------------------------+-----------------------------------
                                  Total |      34296      100.00

. tab prov, gen(dprov)

           province |      Freq.     Percent        Cum.
--------------------+-----------------------------------
 special family unit |          8        0.02        0.02
        newfoundland |       1213        3.55        3.57
prince edward island |        881        2.58        6.15
         nova scotia |       2254        6.60       12.75
       new brunswick |       2072        6.06       18.81
              quebec |       6868       20.10       38.92
             ontario |      10414       30.48       69.40
            manitoba |       2436        7.13       76.53
        saskatchewan |       2205        6.45       82.98
             alberta |       2649        7.75       90.74
    british columbia |       3165        9.26      100.00
--------------------+-----------------------------------
               Total |      34165      100.00

. gen dprov12 = 0

. replace dprov12 = 1 if  prov == 46|  prov==47
(4641 real changes made)

. label var dprov12 "Man/Sask"

. drop if wagsal < 0
(4 observations deleted)

.
.
. /*create dummy age group*/
. gen ageg = 0

. replace ageg =1 if age > 24 & age <35
(6314 real changes made)

. replace ageg = 2 if age >44 & age < 55
(6459 real changes made)

. tab ageg, gen(dageg)

       ageg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      21519       62.75       62.75
          1 |       6314       18.41       81.16
          2 |       6459       18.84      100.00
------------+-----------------------------------
      Total |      34292      100.00

.
. /*pool*/
. gen g2 = (ageg==2)

.
```

```
. /*for men*/
. drop if sex ==2
(8418 observations deleted)

.
.
. /*for education group 2*/
. gen g2deduc1 = g2 * deduc1

. gen g2deduc2 = g2 * deduc2

. gen g2deduc3 = g2 * deduc3

. gen g2deduc4 = g2 * deduc4

. gen g2deduc5 = g2 * deduc5

. gen g2deduc6 = g2 * deduc6

. gen g2deduc7 = g2 * deduc7

.
. /*for prov group 2*/
. gen g2dprov6 = g2 * dprov6
(101 missing values generated)

. gen g2dprov7 = g2 * dprov7
(101 missing values generated)

. gen g2dpro10 = g2 * dprov10
(101 missing values generated)

. gen g2dpro11 = g2 * dprov11
(101 missing values generated)

. gen g2dpro12 = g2 * dprov12

.
. /*occupation for group 2*/
. gen g2docc1 = g2 *docc1

. gen g2docc5 = g2 *docc5

. gen g2docc6 = g2 *docc6

. gen g2docc10 = g2 *docc10

. gen g2docc11 = g2 *docc11

.
. /*group 2 for full-time part-time*/
. gen g2ftpt = g2 * dftpt1

.
. /*pool*/
. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6 docc10 docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4
> g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc
> 5 g2docc6 g2docc10 g2docc11 g2ftpt if ageg ==1 |ageg==2

  Source |       SS       df       MS              Number of obs =    9426
---------+------------------------------           F( 33,  9392) =   75.08
   Model |  1762.54059    33  53.4103209           Prob > F      =  0.0000
Residual |  6681.30192  9392  .711382232           R-squared     =  0.2087
```

```
---------+------------------------------               Adj R-squared =  0.2060
   Total |  8443.84251  9425   .89589841               Root MSE      =  .84343

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.3040725   .0816936     -3.722   0.000      -.4642096   -.1439353
  deduc2 |  -.1233848   .0549402     -2.246   0.025      -.2310795   -.0156901
  deduc4 |   .0825255   .0424303      1.945   0.052      -.0006472    .1656982
  deduc6 |   .2023544   .0391246      5.172   0.000       .1256618    .2790471
  deduc7 |   .2856605   .0460663      6.201   0.000       .1953606    .3759604
  dprov6 |   .1806115   .0406684      4.441   0.000       .1008926    .2603304
  dprov7 |   .3412613   .0378577      9.014   0.000        .267052    .4154705
 dprov12 |   .1661092   .0453303      3.664   0.000        .077252    .2549664
 dprov10 |   .2561718   .0515811      4.966   0.000       .1550618    .3572819
 dprov11 |   .2168915   .0509743      4.255   0.000       .1169708    .3168121
   docc1 |   .1700787   .0416896      4.080   0.000        .088358    .2517994
   docc5 |   .0022088   .0481496      0.046   0.963      -.0921748    .0965925
   docc6 |  -.0255549   .0428028     -0.597   0.550      -.1094576    .0583478
  docc10 |    -.08568   .0421071     -2.035   0.042      -.168219    -.003141
  docc11 |   .0410144   .0410773      0.998   0.318      -.0395059    .1215348
  dftpt1 |   1.111473   .0465908     23.856   0.000       1.020145    1.202801
      g2 |   .5192982   .0853725      6.083   0.000       .3519497    .6866467
g2deduc1 |   -.077381   .0964631     -0.802   0.422      -.2664695    .1117076
g2deduc2 |  -.0275861   .0747473     -0.369   0.712       -.174107    .1189348
g2deduc4 |   .0294106   .0640046      0.460   0.646      -.0960524    .1548736
g2deduc6 |  -.0636513   .0578087     -1.101   0.271      -.1769689    .0496664
g2deduc7 |   .0985861   .0659908      1.494   0.135      -.0307701    .2279423
g2dprov6 |  -.1158353   .0557268     -2.079   0.038      -.2250719   -.0065987
g2dprov7 |  -.0461316   .0516818     -0.893   0.372       -.147439    .0551759
g2dpro12 |  -.1359847   .0627221     -2.168   0.030      -.2589335   -.0130359
g2dpro10 |  -.2404863   .0731054     -3.290   0.001      -.3837887   -.0971839
g2dpro11 |   -.039621   .0717137     -0.552   0.581      -.1801955    .1009534
 g2docc1 |   .1863668   .0550954      3.383   0.001       .0783678    .2943658
 g2docc5 |    -.01372   .0678511     -0.202   0.840      -.1467228    .1192827
 g2docc6 |   .0103647   .0618119      0.168   0.867        -.1108     .1315293
g2docc10 |   .0957005   .0596853      1.603   0.109      -.0212956    .2126966
g2docc11 |    .107952    .058227      1.854   0.064      -.0061857    .2220896
  g2ftpt |  -.1575084   .0626695     -2.513   0.012      -.2803541   -.0346626
   _cons |   8.983557   .0626332    143.431   0.000       8.860783    9.106332
------------------------------------------------------------------------------

.
. /*Notes: the number 17 stands for the number of coefficents to estimate for e
> ach group*/
. predict r, resid
(6028 missing values generated)

. sum r if ageg==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
       r |    4583    2.18e-10   .8195561  -9.945498   2.294081

. gen w = r(Var)*(r(N)-1)/(r(N)-17) if ageg==1
(20915 missing values generated)

. sum r if ageg==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
       r |    4843   -2.12e-11   .8627042  -10.32154   9.877984

. replace w = r(Var)*(r(N)-1)/(r(N)-17) if ageg==2
```

```
(5367 real changes made)

. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6 docc10 docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4
> g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc
> 5 g2docc6 g2docc10 g2docc11 g2ftpt if ageg ==1 |ageg==2 [aw=1/w]
(sum of wgt is  1.3285e+004)

  Source |       SS       df       MS                  Number of obs =    9426
---------+------------------------------               F( 33,  9392) =   74.65
   Model |  1747.95662    33  52.9683824               Prob > F      =  0.0000
Residual |  6663.78699  9392  .709517355               R-squared     =  0.2078
---------+------------------------------               Adj R-squared =  0.2050
   Total |  8411.74361  9425  .892492691               Root MSE      =  .84233

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.3040725   .0795197     -3.824   0.000      -.4599483   -.1481966
  deduc2 |  -.1233848   .0534782     -2.307   0.021      -.2282137   -.0185559
  deduc4 |   .0825255   .0413013      1.998   0.046       .0015661    .1634849
  deduc6 |   .2023544   .0380835      5.313   0.000       .1277026    .2770063
  deduc7 |   .2856605   .0448405      6.371   0.000       .1977635    .3735575
  dprov6 |   .1806115   .0395862      4.562   0.000        .103014     .258209
  dprov7 |   .3412613   .0368503      9.261   0.000       .2690268    .4134958
 dprov12 |   .1661092    .044124      3.765   0.000       .0796165    .2526019
 dprov10 |   .2561718   .0502085      5.102   0.000       .1577524    .3545913
 dprov11 |   .2168915   .0496179      4.371   0.000       .1196298    .3141532
   docc1 |   .1700787   .0405803      4.191   0.000       .0905326    .2496248
   docc5 |   .0022088   .0468683      0.047   0.962      -.0896632    .0940809
   docc6 |  -.0255549   .0416638     -0.613   0.540       -.107225    .0561151
  docc10 |    -.08568   .0409866     -2.090   0.037      -.1660226   -.0053374
  docc11 |   .0410144   .0399842      1.026   0.305      -.0373632    .1193921
  dftpt1 |   1.111473    .045351     24.508   0.000       1.022575    1.200371
      g2 |   .5192982    .085145      6.099   0.000       .3523957    .6862008
g2deduc1 |   -.077381   .0953174     -0.812   0.417      -.2642236    .1094617
g2deduc2 |  -.0275861   .0745404     -0.370   0.711      -.1737015    .1185293
g2deduc4 |   .0294106   .0641572      0.458   0.647      -.0963513    .1551725
g2deduc6 |  -.0636513   .0578917     -1.099   0.272      -.1771316     .049829
g2deduc7 |   .0985861   .0659869      1.494   0.135      -.0307625    .2279348
g2dprov6 |  -.1158353   .0555944     -2.084   0.037      -.2248124   -.0068582
g2dprov7 |  -.0461316   .0515484     -0.895   0.371      -.1471777    .0549145
g2dpro12 |  -.1359847    .062606     -2.172   0.030      -.2587061   -.0132633
g2dpro10 |  -.2404863   .0730618     -3.292   0.001      -.3837032   -.0972693
g2dpro11 |   -.039621   .0716438     -0.553   0.580      -.1800584    .1008163
 g2docc1 |   .1863668   .0548516      3.398   0.001       .0788457    .2938879
 g2docc5 |    -.01372   .0677906     -0.202   0.840      -.1466043    .1191642
 g2docc6 |   .0103647   .0618328      0.168   0.867      -.1108411    .1315704
g2docc10 |   .0957005   .0596501      1.604   0.109      -.0212266    .2126275
g2docc11 |    .107952   .0581927      1.855   0.064      -.0061184    .2220223
  g2ftpt |  -.1575084    .062456     -2.522   0.012      -.2799357    -.035081
   _cons |   8.983557   .0609666    147.352   0.000        8.86405    9.103065
------------------------------------------------------------------------------

.
. /*F-statistics for unconstraint variance model*/
. test g2 g2deduc1 g2deduc2 g2deduc4 g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpro
> 12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2docc6 g2docc10 g2docc11 g2ftpt

 ( 1)  g2 = 0.0
 ( 2)  g2deduc1 = 0.0
 ( 3)  g2deduc2 = 0.0
 ( 4)  g2deduc4 = 0.0
 ( 5)  g2deduc6 = 0.0
```

```
 ( 6)  g2deduc7 = 0.0
 ( 7)  g2dprov6 = 0.0
 ( 8)  g2dprov7 = 0.0
 ( 9)  g2dpro12 = 0.0
 (10)  g2dpro10 = 0.0
 (11)  g2dpro11 = 0.0
 (12)  g2docc1 = 0.0
 (13)  g2docc5 = 0.0
 (14)  g2docc6 = 0.0
 (15)  g2docc10 = 0.0
 (16)  g2docc11 = 0.0
 (17)  g2ftpt = 0.0

       F( 17,  9392) =   23.99
            Prob > F =    0.0000


. drop _all

.
. /*do the same for women*/
. use wagsal age sex receduc wrkft_pt occ13 prov using c:\ass1.dta
(363 : scfef95 : survey of consumer finance - economic families)

. gen lnw = ln(wagsal)
(10137 missing values generated)

. tab sex, gen(dsex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |      25877       75.45       75.45
     female |       8419       24.55      100.00
------------+-----------------------------------
      Total |      34296      100.00

. tab receduc, gen(deduc)

             summary education level |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
    no schooling or grade 8 or lower |       5773       16.83       16.83
                          grade 9-10 |       4257       12.41       29.25
           grade 11-13 not graduate |       2004        5.84       35.09
              grade 11-13 garaduate |       5923       17.27       52.36
some post-secondary no dipl,deg,cert |       2421        7.06       59.42
         post-secondary cert or dipl |       9375       27.34       86.75
                   university degree |       4543       13.25      100.00
-------------------------------------+-----------------------------------
                               Total |      34296      100.00

. tab wrkft_pt, gen(dftpt)

    worked mostly full or part |
        time in reference year |      Freq.     Percent        Cum.
-------------------------------+-----------------------------------
                     full-time |      21053       61.39       61.39
                     part-time |       2506        7.31       68.69
did not work in reference year |      10732       31.29       99.99
                             7 |          1        0.00       99.99
                            20 |          1        0.00       99.99
                            27 |          1        0.00       99.99
                            40 |          1        0.00      100.00
                            41 |          1        0.00      100.00
-------------------------------+-----------------------------------
```

```
                         Total |      34296      100.00

. tab occ13, gen(docc)

  1980 occupational classification - 13 |
                                  groups |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
         managerial and administrative |       3353        9.78        9.78
                      natural sciences |       2986        8.71       18.48
                              teaching |       1010        2.94       21.43
                              clerical |       1868        5.45       26.87
                                 sales |       2210        6.44       33.32
                              services |       3269        9.53       42.85
farming,fishing,forestry and logging op |       1845        5.38       48.23
       mining,processing and machining |       1883        5.49       53.72
product fabricating,assembling and repa |       2547        7.43       61.15
                   construction trades |       2446        7.13       68.28
transport,material handling,other craft |       2669        7.78       76.06
                   never worked before |        947        2.76       78.82
      last worked more than 5 years ago |       7263       21.18      100.00
----------------------------------------+-----------------------------------
                                  Total |      34296      100.00

. tab prov, gen(dprov)

            province |      Freq.     Percent        Cum.
---------------------+-----------------------------------
   special family unit |          8        0.02        0.02
         newfoundland |       1213        3.55        3.57
prince edward island |        881        2.58        6.15
          nova scotia |       2254        6.60       12.75
        new brunswick |       2072        6.06       18.81
               quebec |       6868       20.10       38.92
              ontario |      10414       30.48       69.40
             manitoba |       2436        7.13       76.53
         saskatchewan |       2205        6.45       82.98
              alberta |       2649        7.75       90.74
     british columbia |       3165        9.26      100.00
---------------------+-----------------------------------
                Total |      34165      100.00

. gen dprov12 = 0

. replace dprov12 = 1 if  prov == 46|  prov==47
(4641 real changes made)

. label var dprov12 "Man/Sask"

. drop if wagsal < 0
(4 observations deleted)

.
.
. /*create dummy age group*/
. gen ageg = 0

. replace ageg =1 if age > 24 & age <35
(6314 real changes made)

. replace ageg = 2 if age >44 & age < 55
(6459 real changes made)

. tab ageg, gen(dageg)
```

```
       ageg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      21519       62.75       62.75
          1 |       6314       18.41       81.16
          2 |       6459       18.84      100.00
------------+-----------------------------------
      Total |      34292      100.00

.
. /*pool*/
. gen g2 = (ageg==2)

.
. /*for men*/
. drop if sex ==1
(25874 observations deleted)

.
.
. /*for education group 2*/
. gen g2deduc1 = g2 * deduc1

. gen g2deduc2 = g2 * deduc2

. gen g2deduc3 = g2 * deduc3

. gen g2deduc4 = g2 * deduc4

. gen g2deduc5 = g2 * deduc5

. gen g2deduc6 = g2 * deduc6

. gen g2deduc7 = g2 * deduc7

.
. /*for prov group 2*/
. gen g2dprov6 = g2 * dprov6
(30 missing values generated)

. gen g2dprov7 = g2 * dprov7
(30 missing values generated)

. gen g2dpro10 = g2 * dprov10
(30 missing values generated)

. gen g2dpro11 = g2 * dprov11
(30 missing values generated)

. gen g2dpro12 = g2 * dprov12

.
. /*occupation for group 2*/
. gen g2docc1 = g2 *docc1

. gen g2docc5 = g2 *docc5

. gen g2docc6 = g2 *docc6

. gen g2docc10 = g2 *docc10

. gen g2docc11 = g2 *docc11

.
. /*group 2 for full-time part-time*/
```

```
. gen g2ftpt = g2 * dftpt1

.
. /*pool*/
. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6  docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4 g2dedu
> c6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2do
> cc6  g2docc11 g2ftpt if ageg ==1 |ageg==2

  Source |       SS       df       MS              Number of obs =    1835
---------+------------------------------           F( 31,  1803) =   27.06
   Model |  728.07756    31  23.4863729            Prob > F      =  0.0000
Residual |  1565.04018  1803  .868020066           R-squared     =  0.3175
---------+------------------------------           Adj R-squared =  0.3058
   Total |  2293.11774  1834  1.25033683           Root MSE      =  .93168

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.4165621   .2289932     -1.819   0.069       -.865682    .0325578
  deduc2 |   .0569279   .1538249      0.370   0.711      -.2447659    .3586216
  deduc4 |   .2257894     .10097      2.236   0.025       .0277588    .4238199
  deduc6 |   .4069217   .0934652      4.354   0.000       .2236102    .5902333
  deduc7 |    .529994   .1010564      5.245   0.000       .3317941    .7281939
  dprov6 |   .2944867   .1033348      2.850   0.004       .0918182    .4971552
  dprov7 |   .3906155    .089497      4.365   0.000       .2150867    .5661442
 dprov12 |    .121286   .1069098      1.134   0.257       -.088394     .330966
 dprov10 |   .3609925   .1187187      3.041   0.002       .1281518    .5938332
 dprov11 |   .4615904   .1160405      3.978   0.000       .2340025    .6891783
   docc1 |   .1613444   .0867609      1.860   0.063      -.0088181    .3315069
   docc5 |  -.1464164   .1185398     -1.235   0.217      -.3789062    .0860735
   docc6 |  -.5164822   .0803594     -6.427   0.000      -.6740895   -.3588748
  docc11 |  -.2654534   .1808484     -1.468   0.142      -.6201478     .089241
  dftpt1 |   .9360749   .0767381     12.198   0.000         .78557     1.08658
      g2 |   .3953911   .1886501      2.096   0.036       .0253953    .7653869
g2deduc1 |    .055615   .2796003      0.199   0.842      -.4927597    .6039897
g2deduc2 |  -.4546646   .2120546     -2.144   0.032      -.8705632    -.038766
g2deduc4 |  -.2653172   .1569093     -1.691   0.091      -.5730604    .0424259
g2deduc6 |  -.3931499   .1446376     -2.718   0.007      -.6768248    -.109475
g2deduc7 |  -.2842871   .1595492     -1.782   0.075      -.5972077    .0286336
g2dprov6 |  -.1587404   .1519801     -1.044   0.296      -.4568161    .1393352
g2dprov7 |  -.0977524   .1398095     -0.699   0.485       -.371958    .1764533
g2dpro12 |    .105598   .1731777      0.610   0.542       -.234052    .4452481
g2dpro10 |  -.1051346   .1937456     -0.543   0.587       -.485124    .2748549
g2dpro11 |  -.2206552   .1768361     -1.248   0.212      -.5674805    .1261701
 g2docc1 |   .1938023   .1298085      1.493   0.136      -.0607886    .4483933
 g2docc5 |  -.0948978   .1721062     -0.551   0.581      -.4324463    .2426507
 g2docc6 |   .2757955   .1282322      2.151   0.032       .0242962    .5272948
g2docc11 |  -.0531885   .2731061     -0.195   0.846      -.5888261    .4824492
  g2ftpt |   .2836946   .1114785      2.545   0.011       .0650539    .5023352
   _cons |   8.428428   .1224299     68.843   0.000       8.188309    8.668547
------------------------------------------------------------------------------

.
. /*Notes: the number 17 stands for the number of coefficents to estimate for e
> ach group*/
. predict r, resid
(4196 missing values generated)

. sum r if ageg==1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
       r |    1006   -1.74e-09   .9080789  -6.130455    2.58335
```

```
. gen w = r(Var)*(r(N)-1)/(r(N)-17) if ageg==1
(7063 missing values generated)

. sum r if ageg==2

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
       r |     829    8.14e-10   .9430075  -9.025227   2.477489

. replace w = r(Var)*(r(N)-1)/(r(N)-17) if ageg==2
(1092 real changes made)

. regress lnw deduc1 deduc2 deduc4 deduc6 deduc7 dprov6 dprov7 dprov12 dprov10
> dprov11 docc1 docc5 docc6  docc11 dftpt1 g2 g2deduc1 g2deduc2 g2deduc4 g2dedu
> c6 g2deduc7 g2dprov6 g2dprov7 g2dpro12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2do
> cc6  g2docc11 g2ftpt if ageg ==1 |ageg==2 [aw=1/w]
(sum of wgt is  2.1148e+003)

  Source |       SS       df       MS                Number of obs =    1835
---------+------------------------------             F( 31,  1803) =   27.01
   Model |  725.671413    31  23.4087553              Prob > F      =  0.0000
Residual |  1562.73927  1803  .866743909              R-squared     =  0.3171
---------+------------------------------             Adj R-squared =  0.3054
   Total |  2288.41068  1834  1.24777027              Root MSE      =  .93099

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  deduc1 |  -.4165621   .2248667     -1.852   0.064       -.8575888    .0244646
  deduc2 |   .0569279   .1510529      0.377   0.706       -.2393293     .353185
  deduc4 |   .2257894   .0991505      2.277   0.023        .0313273    .4202514
  deduc6 |   .4069217    .091781      4.434   0.000        .2269135      .58693
  deduc7 |    .529994   .0992353      5.341   0.000        .3353657    .7246223
  dprov6 |   .2944867   .1014727      2.902   0.004        .0954703    .4935031
  dprov7 |   .3906155   .0878843      4.445   0.000        .2182498    .5629812
 dprov12 |    .121286   .1049832      1.155   0.248       -.0846156    .3271875
 dprov10 |   .3609925   .1165794      3.097   0.002        .1323477    .5896374
 dprov11 |   .4615904   .1139494      4.051   0.000        .2381037    .6850772
   docc1 |   .1613444   .0851975      1.894   0.058       -.0057517    .3284405
   docc5 |  -.1464164   .1164037     -1.258   0.209       -.3747167    .0818839
   docc6 |  -.5164822   .0789113     -6.545   0.000       -.6712494   -.3617149
  docc11 |  -.2654534   .1775894     -1.495   0.135       -.6137561    .0828493
  dftpt1 |   .9360749   .0753552     12.422   0.000        .7882821    1.083868
      g2 |   .3953911   .1896039      2.085   0.037        .0235247    .7672575
g2deduc1 |    .055615   .2782501      0.200   0.842       -.4901116    .6013416
g2deduc2 |  -.4546646   .2122471     -2.142   0.032       -.8709408   -.0383885
g2deduc4 |  -.2653172   .1577464     -1.682   0.093       -.5747022    .0440678
g2deduc6 |  -.3931499   .1453893     -2.704   0.007       -.6782991   -.1080006
g2deduc7 |  -.2842871   .1604815     -1.771   0.077       -.5990362    .0304621
g2dprov6 |  -.1587404    .152502     -1.041   0.298       -.4578396    .1403588
g2dprov7 |  -.0977524   .1405791     -0.695   0.487       -.3734675    .1779628
g2dpro12 |    .105598   .1743264      0.606   0.545        -.236305    .4475011
g2dpro10 |  -.1051346   .1950738     -0.539   0.590       -.4877289    .2774598
g2dpro11 |  -.2206552   .1776645     -1.242   0.214        -.569105    .1277946
 g2docc1 |   .1938023    .130334      1.487   0.137       -.0618192    .4494239
 g2docc5 |  -.0948978    .172615     -0.550   0.583       -.4334442    .2436486
 g2docc6 |   .2757955   .1290242      2.138   0.033        .0227428    .5288482
g2docc11 |  -.0531885   .2743003     -0.194   0.846       -.5911683    .4847914
  g2ftpt |   .2836946   .1118105      2.537   0.011        .0644029    .5029863
   _cons |   8.428428   .1202237     70.106   0.000        8.192636     8.66422
------------------------------------------------------------------------------

.
```

```
. /*F-statistics for unconstraint variance model*/
. test g2 g2deduc1 g2deduc2 g2deduc4 g2deduc6 g2deduc7 g2dprov6 g2dprov7 g2dpro
> 12 g2dpro10 g2dpro11 g2docc1 g2docc5 g2docc6  g2docc11 g2ftpt

 ( 1)  g2 = 0.0
 ( 2)  g2deduc1 = 0.0
 ( 3)  g2deduc2 = 0.0
 ( 4)  g2deduc4 = 0.0
 ( 5)  g2deduc6 = 0.0
 ( 6)  g2deduc7 = 0.0
 ( 7)  g2dprov6 = 0.0
 ( 8)  g2dprov7 = 0.0
 ( 9)  g2dpro12 = 0.0
 (10)  g2dpro10 = 0.0
 (11)  g2dpro11 = 0.0
 (12)  g2docc1 = 0.0
 (13)  g2docc5 = 0.0
 (14)  g2docc6 = 0.0
 (15)  g2docc11 = 0.0
 (16)  g2ftpt = 0.0

       F( 16,  1803) =    4.89
            Prob > F =    0.0000


.
.
. /*Conclusion: it does not matter much
> If there were more groups, and the variance differences were great among
> the groups, this could become more important. */
.
. /*end project 3*/
.
end of do-file

.
```

Appendix II - Tables

Table 1. Comparison of Variables used by Kapsalis, Morissette, Picot (1999) with those used in this project.

	Variable	Kapsalis, Morissette, Picot (1999)	ECON452 Project II.1 (2001)
Summary Education Level (receduc)	No schooling or grade 8 or lower		*
	Grade 9-10	*	*
	Grade 11-13 not graduate	*	*
	Grade 11-13 graduate	*	*
	Some post-secondary not graduate	*	*
	Post-secondary cert/diploma	*	*
	University degree	*	*
Region (prov)	Quebec	*	*
	Ontario	*	*
	Manitoba/Saskatchewan	*	*
	Alberta	*	*
	British Columbia	*	*
Occupation (occ13)	Manufacturing durables	*	
	Manufacturing non-durables	*	
	Construction trades	*	*
	Transportation/Communication	*	*
	Sales	*	*
	F.I.R.E.	*	
	Services	*	*
	Managerial & Administrative	*	*
Work Status (wkft_pt)	Full-time	*	*
	Part-time		

SOURCE: Survey of Consumers Finance -- economic families (1995).

Table 2. Regression Results: For Males, base age group 25-34, 1995, Dependent Variable: ln Weekly Wages (SCF)

	Variable	Coefficient	(Std Error)
Base Age Group 25-34			
	Constant	8.983557	(.0626332)
Education Level	No school or < Gr. 8	-.3040725	(.0816936)
	Gr. 9-10	-.1233848	(.0549402)
	Gr. 11-13 Graduate	.0825255	(.0424303)
	Post-secondary diploma	.2023544	(.0391246)
	University Degree	.2856605	(.0460663)
Region	Quebec	.1806115	(.0406684)
	Ontario	.3412613	(.0378577)
	Manitoba/Saskatchewan	.1661092	(.0453303)
	Alberta	.2561718	(.0515811)
	British Columbia	.2168915	(.0509743)
Occupation	Managerial & Administrative	.1700787	(.0416896)
	Sales	.0022088	(.0481496)
	Services	-.0255549	(.0428028)
	Construction Trades	-.08568	(.0421071)
	Transportation/communication	.0410144	(.0410773)
Work Status	Full-time	1.111473	(.0465908)
Pairwise coefficient differences: ($g = b_{45-54} - b_{25-34}$)			
	Old*Constant	.5192982	(.0853725)
Education Level	Old*no_school (or < Gr. 8)	-.077381	(.0964631)
	Old*Gr. 9-10	-.0275861	(.0747473)
	Old*Gr. 11-13	.0294106	(.0640046)
	Old*Post-Sec. Diploma	-.0636513	(.578087)
	Old*University	.0985861	(.0659908)
Region	Old*Quebec	-.1158353	(.0557268)
	Old*Ontario	-.0461316	(.0516818)
	Old*Manitoba/Saskatchewan	-.1359847	(.0627221)
	Old*Alberta	-.2404863	(.0731054)
	Old*British Columbia	-.039621	(.0717137)
Occupation	Old*Managerial/Administrative	.1863668	(.0550954)
	Old*Sales	-.01372	(.0678511)
	Old*Services	.0103647	(.0618119)
	Old*Construction	.0957005	(.0596853)
	Old*Transportation	.107952	(.058227)
Work Status	Old*FT	-.1575084	(.0626695)

Number of Observations: Age 24-35 = 4583; Age 45-54 = 4843

$R_U^2 = 0.2087$; $R_R^2 = 0.1682$

SOURCE: Data from SCF(1995) and Statistical Analysis performed by STATA software..

Table 3. Regression Results: For Females, base age group 25-34, 1995, Dependent Variable: ln Weekly Wages (SCF)

Variable		Coefficient	(Std Error)
Base Age Group 25-34			
	Constant	8.428428	(.1224299)
Education Level	No school or < Gr. 8	-.4165621	(.2289932)
	Gr. 9-10	.0569279	(.1538249)
	Gr. 11-13 Graduate	.2257894	(.10097)
	Post-secondary diploma	.4069217	(.0934652)
	University Degree	.529994	(.1010564)
Region	Quebec	.2944867	(.1033348)
	Ontario	.3906155	(.089497)
	Manitoba/Saskatchewan	.121286	(.1069098)
	Alberta	.3609925	(.1187187)
	British Columbia	.4615904	(.1160405)
Occupation	Managerial & Administrative	.1613444	(.0867609)
	Sales	-.1464164	(.1185398)
	Services	-.5164822	(.0803594)
	Transportation/communication	-.2654534	(.1808484)
Work Status	Full-time	.9360749	(.0767381)
Pairwise coefficient differences: $(g = b_{45-54} - b_{25-34})$			
	Old*Constant	.3953911	(.1886501)
Education Level	Old*no_school (or < Gr. 8)	.055615	(.2796003)
	Old*Gr. 9-10	-.4546646	(.2120546)
	Old*Gr. 11-13	-.2653172	(.1569093)
	Old*Post-Sec. Diploma	-.3931499	(.1446376)
	Old*University	-.2842871	(.1595492)
Region	Old*Quebec	-.1587404	(.1519801)
	Old*Ontario	-.0977524	(.1398095)
	Old*Manitoba/Saskatchewan	.105598	(.1731777)
	Old*Alberta	-.1051346	(.1937456)
	Old*British Columbia	-.2206552	(.1768361)
Occupation	Old*Managerial/Administrative	.1938023	(.1298085)
	Old*Sales	-.0948978	(.1721062)
	Old*Services	.2757955	(.1282322)
	Old*Transportation	-.0531885	(.2731061)
Work Status	Old*FT	.2836946	(.1114785)

Number of Observations: Age 24-35 = 1006; Age 45-54 = 829

$R_U^2 = 0.3175$; $R_R^2 = 0.2980$

SOURCE: Data from SCF(1995) and Statistical Analysis performed by STATA software..

Economics 452, Part II, Project 1:

Modeling "*The Effect of Illicit Drug Use on the Labor Supply of Young Adults*" by Robert Kaestner

Authors:

Data Set: Canadian Alcohol and Drugs Survey, 1995

Data Set #6

Model Paper: Kaestner, Robert. "*The Effect of Illicit Drug Use on The Labor Supply of Young Adults*", The Journal of Human Resources, XXIX, 1, 1993.

-

Introduction

When discussing the negative repercussions of substance abuse, one of the most important considerations from an economic perspective is the effect on the labor supply. In terms of empirical research, the bulk of academic analysis has focussed on the relationship between alcohol abuse and the labor market, while the relative effects of illicit drug use have been virtually ignored. In his paper "The Effect of Illicit Drug Use on the Labor Supply of Young Adults," Robert Kaestner investigates whether the frequency and timing of marijuana and cocaine use are systematically related to labor supply decisions in young adults. Through this study, Kaestner becomes one of the first researchers to apply economic theory to the relationship between illicit drug use and labor market participation.

-

Economic Theory and Analytical Model

In order to construct a reflective economic model, Kaestner decides to treat illicit drugs as a consumption good and, using Becker and Murphy's (1988) formulation, derives an age–specific utility function;

(1) $U_t = u(L_t, D_t, S_t, X_t)$

Where L is the amount of leisure, D is the quantity of illicit drugs consumed, S ids the stock of drug consumption capital, X is a composite good representing all other consumption choices, and t = 1. . .T indexes age.

The corresponding utility-dependent cost function is defined as

$$(2)\ C_t = c(W_t, V_t, P_{xt}: U_t = u(L_t, D_t, S_t, X_t))$$

in which W represents the wage, V is the price of drug consumption (as a function of both the market price of drugs and the user costs of drug consumption capital) and P_x is the price of all other consumption. Equations isolating for the optimal supply of labour and other variables are arrived at by partially differentiating equation (2) with respect to the variable in question. Thus, the demand for leisure can be represented by:

$$(3)\ L_t = g(W_t, D_t, P_{xt}, U)$$

Using this framework, the choice amount of leisure is dependent on the amount of drug use. Assuming that leisure and drug use are complimentary goods, Kaestner hypothesizes that increasing levels of drug use also increase the demand for leisure, thus decreasing the quantity of labour supplied. As such, Kaestner uses a modified version of equation (3) as the basis for his OLS estimation, arriving at the model:

$$(3a)\ H_t = a + a_1 \ln OW_t + a_2 OW_t^{.5} SW_t^{-.5} + a_3 D_t + a_4 Z_t + a_5 \ln U + e_t$$

in which, $H_t = T - L_t$ is hours of work, OW_t is respondent's wage, SW_t is the spouse's wage, Z is a vector of other exogenous variables such as age and education, and *e* is the stochastic error term. Since drug use is determined endogenously, Kaestner utilizes an instrumental variables approach in his estimation. Wages are estimated through personal characteristics of the respondent, such as age, education and past labor force participation. Each respondent's level of drug use is estimated through measures of self-esteem, religious attendance, family characteristics, in addition to age and education.

Description of Data

Building on this theoretical framework, Kaestner moves into building appropriate regressions to estimate his hypothesis. The cross-sectional data set used for empirical analysis is the National Longitudinal Survey of Youth (NLSY), 1984 and 1988. This survey describes experiences of young persons including their labor market exposure, personal background and history of illicit drug use.

Kaestner's regressions focus on the effects of historical and current consumption of marijuana and cocaine on labor supply. The individuals chosen for the sample had to be 21 years old in 1984, either living independently or with their parents, could not be in school, could not have served in the military, and could not be in prison over the period between 1984 and 1988. All analyses of the relevant data were done separately for gender and marital status. In the case of gender, previous articles have shown a significant difference between the illicit drug use of male and that of females. With respect to marital status, the expected impact of other family members on the respondent's labor supply is analyzed.

Two issues of concern discussed in the article were the degree of underreporting that occurred between the 1984 and 1988 surveys the lack of a variable indicating quantity of illicit drug use, rather than just one indicating frequency of use. This first issue is particularly the case with cocaine use, and is more common among light users than heavy user of the drug. The second had to do with the fact that even though quantity used and frequency of use, are highly correlated, they are not the same. The level of drug use is a stronger indicator of heavy users relative to light users.

Concern for the inconsistency in the survey data due to underreporting led the authors to estimate their models twice, once using the original drug variables and once using "internally consistent" variables. Internally consistent variables refer to where an individual reports previous use of an illicit drug in 1984, but not in 1988, and the 1988 value is replaced with the 1984 value. These estimates were essentially identical, but the results discussed in the article are those pertaining to internally consistent variables.

The dependent variable used in the estimates was hours of work, referring to the number of hours worked per week multiplied by the number of weeks worked at a job, and represents the variable used to measure labour supply in the model. The one problem indicated with their variable is that it ignores the loss of work due to absenteeism. Labour Force participation depends on whether the individual worked at all over the past year.

When estimating drug use in the model, a number of dummy variables were used to measure the frequency of lifetime use of illicit drugs. For marijuana, the coding was categorized by no use, 1-39 times, and 40 times or greater. For cocaine, the categorization was no use, 1-9 times and 10 times or more of use. In addition, a variable was used to indicate use within the past year, where a dummy variable was coded as either use or no use.

Additional variables include experience and several personal and family background variables. Experience refers to the actual sum of weeks worked since 1975. The personal and family background variables were taken from scores respondents received in questionnaires relating to self-esteem, an individuals feeling of control over the world, frequency of religious attendance, and criminal record prior to 1980.

In the article Kaestner discusses the cross-sectional estimates obtained from his model. He finds that in 1984 married men who use marijuana 40 or more times over their lifetime work between 503 and 587 hours less than do those individuals who have not used any illicit substances. Estimates were similar for 1988, with a decrease of between 342 and 339 hours less per year for men who have used marijuana than for those who have not.

For cocaine, the impact does not appear to be significant for 1984, but in 1988, cocaine use is found to be associated with less hours worked per year for both married and single men. A married man who has used cocaine 20 times is expected to work 230 hours less than a comparable male non-user. For single men, the pattern is similar with an expected decline of 112 hours.

Among females, marijuana use is only found to be significant for single females in the 1988 survey. A single woman who uses marijuana 40 or more times in her life is expected to work between 518 and 587 hours less per year than female non-users. In the case of cocaine use over the past year, the results are only found to be significant for single females in 1988, and even then it is barely significant at 10 percent.

Our Data, Model and Results

Because the survey data Kaestner used was unavailable to us, we tried to replicate his findings using the 1994 Canadian Alcohol and Drugs survey. Considerable differences existed between these two sources of data. Enough variables could be extracted, however, to attempt a test of the hypothesized negative relationship between labor supply and drug use. A full list of these variables can be found in the appendix.

The dependent variable used in our estimation, e5, is the reported number of hours an individual works on average per week. Among the independent variables chosen were province (*prov*), marital status (*stat2*) and four variables regarding the extent of cocaine and marijuana consumption. Once these variables were extracted, every effort was made to duplicate Kaestner's sample characteristics as closely as possible. All unemployed individuals were deleted along with any individuals who did not provide full drug-use information. Other adjustments were made for missing observations and coding problems. Unfortunately, age variables in the survey were not helpful. For some reason, every individual had a recorded age of 0. As a result, we were unable to limit the sample to young adults.

Once the data had been edited to a suitable level, every independent variable chosen in our sample, with the exception of "number of children under 15 in household" needed to be transformed into a dummy variable due to the nature of the survey answers. Furthermore, transformations on some variables were required in order to achieve uniform spread in subcategories (i.e. 15-20, 20-25).

When working with the data, careful consideration was given to the problems of underreporting and misrepresentation. Dealing with a sensitive issue such as drug use can induce people to either provide no information or incorrect information regarding their habits. Consequently, any results calculated from this data must be inspected closely.

In order to mimic Kaestner's relation between labor supply and drug use for males and females, we used a pooled regression in the following form:

$$\mathbf{H = b_1 + b_{i2}Prov + b_{i3}m5a + b_{i4}m5b2 + b_{i5}m5dm + b_{i6}m5dn + b_{i7}b4 + b_{i8}hsdinc + b_9dvtotur + a_1 + a_{i2}fprov + a_{i3}fm5a + a_{i4}fm5b2 + a_{i5}fm5dm + a_{i6}fm5dn + a_{i7}fb4 + a_{i8}fhsdinc + a_9fhsdinc}$$

where *prov* is province, *m5a* is whether the subject has tried marijuana, *m5b2* is how often marijuana has been smoked in the last twelve months, *m5dm* is whether the subject has tried cocaine, m5dn is how often cocaine has been used in the last 12 months, *hsdinc* is household income and *dvtotur* is number of children under fifteen in household for men. The remaining variables are simply the female equivalents. This regression was run over 3771 observations and the outcome scrutinized.

Looking solely at the beta and alpha coefficients, it appears that the results obtained are inconsistent with Kaestner's. Historical marijuana and cocaine use (not in the last 12 months) do not seem to have a determinable effect on labor supply. Coefficients fluctuate positively and negatively with no apparent pattern. Analyzing the difference between men and women, it is difficult to identify any trend. Wild fluctuations in the coefficient terms made it nearly impossible to compare.

Due to the erratic nature of the coefficients that did not coincide with Kaestner's findings, we performed t-tests for individual coefficient restrictions to determine the statistical significance of the drug-use terms. Of the numerous variables in question, only one was significant at the 10% level. Incidentally, this coefficient was positive, suggesting a *positive* relationship between drug use and labor supply. In other words, our findings were almost entirely statistically insignificant, and conflicted with Kaestner's.

Conclusion

In an effort to mimic the statistical results from Robert Kaestner's paper "*The Effect of Illicit Drug Use on The Labor Supply of Young Adults,*" we extracted data from the Canadian Alcohol and Drugs survey of 1994. After running a multiple linear regression model similar to the one used by Kaestner, we were unable to achieve our goal. Our results conflicted greatly with his, producing no significant statistical relationship.

The inability to produce outcomes that resemble Kaestner's most likely results from the difference in the data sets utilized. It was our mistaken conclusion that these differences were not great enough to skew our results. The availability of relevant information was significantly limited by the Canadian survey, leaving us with data that most likely did not allow us to draw an accurate comparison.

References

Becker, Gary S., and Murphy, Kevin M. 1988. "A Theory of Rational Addition." *Journal of Political Economy* 96(4): 675-700.

Kaestner, Robert. 1993. "The Effect of Illicit Drug Use on The Labor Supply of Young Adults", *The Journal of Human Resources 29(1)*: 127-151.

Log File____

****

This is a Stata log file for a QED session

Course: Econ 452

Students: colin

Date and time: Sun, 25 Mar 2001, 12:20:17

At the end of the QED session, this file will be copied to:

84_213_Sun_colin.log

These files will also be uploaded to:

http://edith.econ.queensu.ca/statausr/logfiles/Econ452

Type help QEDstata for a list of QED commands

Student work begins below this line

****

pause: "Type BREAK to end session started at 25 Mar 2001 12:20:17"

-> . Qextract

getting information about file 6 ...

```
loading variables from 6 (cads94) only (no data yet)...  done

-> . drop if m5a>5

(359 observations deleted)

-> . drop if m5dm>5

(22 observations deleted)

-> . mvencode e5, mv(1000)

e5:      5111 missing values

-> . drop if e5>995

(5177 observations deleted)

-> . mvencode m5b2, mv(-1)

m5b2:    6037 missing values

-> . drop if m5b2>6

(1 observation deleted)

-> . drop if b4>95

(0 observations deleted)

-> . drop if hsdinc>95

(1537 observations deleted)

-> . mvencode m5dn, mv(0)

m5dn:    4754 missing values

-> . mvencode stat2, mv(0)

stat2:   1281 missing values

-> . drop if stat2<1

(1281 observations deleted)

-> . drop if stat2>6

(5 observations deleted)

-> . browse

-> . browse

-> . browse

-> . drop if e5<2

(2 observations deleted)
```

```
-> . browse

-> . xi: regress e5 i.prov i.m5a i.m5b2 i.m5dm i.m5dn i.b4 i.hsdinc dvtotur female2 i.fm5a i.fm5b2
i.fm5dm i.fm5dn i.fb4 i.fhsdinc fdvtotur i.fprov

i.prov            Iprov_10-59  (naturally coded; Iprov_10 omitted)

i.m5a             Im5a_0-2     (naturally coded; Im5a_0 omitted)

i.m5b2            Im5b2_1-6    (Im5b2_1 for m5b2==-1 omitted)

i.m5dm            Im5dm_1-2    (naturally coded; Im5dm_1 omitted)

i.m5dn            Im5dn_0-2    (naturally coded; Im5dn_0 omitted)

i.b4              Ib4_1-8      (naturally coded; Ib4_1 omitted)

i.hsdinc          Ihsdin_0-9   (naturally coded; Ihsdin_0 omitted)

i.fm5a            Ifm5a_0-2    (naturally coded; Ifm5a_0 omitted)

i.fm5b2           Ifm5b2_1-6   (Ifm5b2_1 for fm5b2==-1 omitted)

i.fm5dm           Ifm5dm_0-2   (naturally coded; Ifm5dm_0 omitted)

i.fm5dn           Ifm5dn_0-2   (naturally coded; Ifm5dn_0 omitted)

i.fb4             Ifb4_0-2     (naturally coded; Ifb4_0 omitted)

i.fhsdinc         Ifhsdi_0-9   (naturally coded; Ifhsdi_0 omitted)

i.fprov           Ifpro_0-59   (naturally coded; Ifpro_0 omitted)


  Source |       SS       df       MS              Number of obs =    3771
---------+------------------------------           F( 60,  3710) =   11.09
   Model |  101262.849    60  1687.71416           Prob > F      =  0.0000
Residual |  564770.334  3710  152.229201           R-squared     =  0.1520
---------+------------------------------           Adj R-squared =  0.1383
   Total |  666033.184  3770  176.666627           Root MSE      =  12.338


------------------------------------------------------------------------------
      e5 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Iprov_11 |  -.6672941    2.02012     -0.330   0.741      -4.627948     3.29336
Iprov_12 |  -2.739607   1.420942     -1.928   0.054       -5.52551    .0462973
```

Iprov_13	-3.976192	1.550634	-2.564	0.010	-7.01637	-.936014
Iprov_24	-4.323061	1.259247	-3.433	0.001	-6.791944	-1.854178
Iprov_35	-3.832577	1.337729	-2.865	0.004	-6.455333	-1.20982
Iprov_46	-.8311719	1.45564	-0.571	0.568	-3.685105	2.022761
Iprov_47	-.9695665	1.480573	-0.655	0.513	-3.872384	1.933251
Iprov_48	.0297464	1.296037	0.023	0.982	-2.511268	2.570761
Iprov_59	-3.764268	1.374835	-2.738	0.006	-6.459774	-1.068762
Im5a_1	-.2275208	1.210597	-0.188	0.851	-2.601021	2.14598
Im5a_2	.3668031	1.157169	0.317	0.751	-1.901946	2.635552
Im5b2_2	-.1133867	3.244568	-0.035	0.972	-6.474698	6.247925
Im5b2_3	-.4408187	2.333637	-0.189	0.850	-5.016156	4.134518
Im5b2_4	.1161507	2.782453	0.042	0.967	-5.339136	5.571438
Im5b2_5	-2.490759	2.283012	-1.091	0.275	-6.966841	1.985322
Im5b2_6	-1.67477	1.516463	-1.104	0.269	-4.647952	1.298412
Im5dm_2	.6394686	1.324878	0.483	0.629	-1.958093	3.23703
Im5dn_1	-2.090115	2.966548	-0.705	0.481	-7.906339	3.726108
Ib4_2	.1722756	.8566791	0.201	0.841	-1.507332	1.851884
Ib4_8	1.618297	12.36781	0.131	0.896	-22.63007	25.86667
Ihsdin_1	-7.473059	5.879016	-1.271	0.204	-18.99948	4.05336
Ihsdin_2	.9921548	4.993618	0.199	0.843	-8.79835	10.78266
Ihsdin_3	-.6169043	4.61269	-0.134	0.894	-9.66056	8.426752
Ihsdin_4	-.9429658	4.548781	-0.207	0.836	-9.861322	7.97539
Ihsdin_5	.4493167	4.454385	0.101	0.920	-8.283966	9.1826
Ihsdin_6	1.874545	4.429219	0.423	0.672	-6.809398	10.55849
Ihsdin_7	2.168682	4.410919	0.492	0.623	-6.47938	10.81674
Ihsdin_8	3.645411	4.430105	0.823	0.411	-5.040267	12.33109
Ihsdin_9	4.116367	4.448535	0.925	0.355	-4.605445	12.83818
dvtotur	.8590618	.2648241	3.244	0.001	.3398468	1.378277
female2	3.592138	8.912541	0.403	0.687	-13.88182	21.0661
Ifm5a_1	1.086544	1.817354	0.598	0.550	-2.476567	4.649656

lfm5a_2	-.9928594	1.714376	-0.579	0.563	-4.354071	2.368352
lfm5b2_2	1.418523	5.50726	0.258	0.797	-9.37903	12.21608
lfm5b2_3	-2.787442	4.582989	-0.608	0.543	-11.77287	6.197983
lfm5b2_4	-8.770969	5.239445	-1.674	0.094	-19.04344	1.501506
lfm5b2_5	5.304625	4.833837	1.097	0.273	-4.172612	14.78186
lfm5b2_6	-.7778761	2.588612	-0.300	0.764	-5.853118	4.297366
lfm5dm_1	.2780151	6.041076	0.046	0.963	-11.56614	12.12217
lfm5dm_2	-1.233912	2.260466	-0.546	0.585	-5.665788	3.197965
lfb4_1	-1.14704	1.329778	-0.863	0.388	-3.754208	1.460128
lfhsdi_1	.7833322	7.69216	0.102	0.919	-14.29794	15.86461
lfhsdi_2	-14.98264	6.893923	-2.173	0.030	-28.49889	-1.466391
lfhsdi_3	-4.565209	6.528082	-0.699	0.484	-17.36419	8.233771
lfhsdi_4	-3.255471	6.444997	-0.505	0.614	-15.89155	9.380611
lfhsdi_5	-5.704255	6.31327	-0.904	0.366	-18.08208	6.673565
lfhsdi_6	-6.079304	6.28835	-0.967	0.334	-18.40827	6.249658
lfhsdi_7	-6.067593	6.256519	-0.970	0.332	-18.33415	6.198961
lfhsdi_8	-7.622131	6.288864	-1.212	0.226	-19.9521	4.707839
lfhsdi_9	-5.406892	6.314176	-0.856	0.392	-17.78649	6.972704
fdvtotur	-2.411882	.4076475	-5.917	0.000	-3.211117	-1.612647
lfpro_10	-1.447279	3.079106	-0.470	0.638	-7.484184	4.589627
lfpro_12	-.3351006	3.001655	-0.112	0.911	-6.220156	5.549955
lfpro_13	.871759	3.059343	0.285	0.776	-5.1264	6.869918
lfpro_24	-.690217	2.797671	-0.247	0.805	-6.17534	4.794906
lfpro_35	-1.458304	2.84973	-0.512	0.609	-7.045494	4.128887
lfpro_46	-3.535761	2.983931	-1.185	0.236	-9.386067	2.314544
lfpro_47	-2.122482	3.026039	-0.701	0.483	-8.055345	3.810381
lfpro_48	-1.092135	2.851329	-0.383	0.702	-6.682461	4.498191
lfpro_59	-1.161995	2.902064	-0.400	0.689	-6.851791	4.527802
_cons	42.89942	7.278301	5.894	0.000	28.62955	57.16928

----

r(699);

-> . log close

THE EFFECTS OF ALCOHOL CONSUMPTION ON EARNINGS

The data set used for this paper is from the General Social Survey, 1985, found in the Queen's Economics Data Archive, (file number 28)

The reference paper is:
Hamilton, Vivian and Barton Hamilton (1997) “Alcohol and earnings: does drinking yield a wage premium?” *Canadian Journal of Economics*, Vol. 30, No.1, pp.135-151

I INTRODUCTION

The effects of alcohol consumption on wage earnings have been analyzed for many years. Berger and Leigh (1988) determined that those who have the minimum of one drink per week would earn a higher wage than non-drinkers. Kaestner (1991), Gill and Michaels (1992) and Register and Williams (1992), find that heavy substance among young adults may lead to higher earnings compared to those of non-users. Contrary to these previous findings, Mullahy and Sindelar (1991) established that alcoholism has a negative effect on earnings.

In Vivian and Barton Hamilton's paper, entitled "Alcohol and earnings: does drinking yield a wage premium?" the relationship between alcohol consumption and earnings for prime age males is examined. Hamilton and Hamilton also examined how earnings differed for different drinking types across age groups. They identify the positive effects of income on alcohol consumption, alcohol use effecting earnings and earnings effecting alcohol use. The testing conducted on these two relationships identifies drinking as an endogenous variable in the determination of earnings. Hamilton and Hamilton define heavy drinker according to medical literature and tests to see whether there is a threshold point at which heavy drinking has a negative impact on wages. For the purpose of this paper, an individual's drinker status will be defined by one of three types: non-, moderate or heavy drinker. The model used in this article is a "polychotomous choice model to estimate the wage differentials between drinker types, while accounting for the possible correlation between unobserved factors affecting both alcohol use and earnings".

The data used in this article was from the 1985 General Social Survey (GSS), which consists of a sample of Canadians, randomly collected during the period of September 25 to October 18, 1985. The sample was restricted to males between the ages of twenty-five and fifty-nine years and has reported positive earnings. The sample was also restricted to those who have worked at some employment in the past week and that the main activity in that week was work. The data set used was not optimal in capturing the possibility that alcohol abuse will have an additional negative impact on earnings through its effect on employment status as there is only data regarding heavy drinking, which is a less severe problem then alcoholism or alcohol dependence. Hamilton and Hamilton defined non-drinkers as those who never drank over the course of the past year or less than once a month; moderate drinkers as those who drink once a month, once a week or every day, yet never consumed eight or more drinks on any given day; and, heavy drinkers as those who drank at least once a week in the previous twelve months and drank eight or more drinks on one or more days in the previous week. The chosen cut-off between moderate and heavy drinkers is based on findings by Knupfer (1984) who observes that those individuals who consume at least eight drinks a day one or more times a week, face a risk of social disapproval or personal concern of their drinking habits.

Hamilton and Hamilton estimated an OLS wage regression, keeping moderate drinkers as a base and using dummy variables for non-drinker and heavy drinker, They determined that non-drinkers earn 7.4 percent less than moderate drinkers and heavy drinkers earn 6.6 percent more than moderate drinkers. These results are significantly different than previous tests that show no drop off in earnings for heavy versus moderate drinkers. Finally, the results show that heavy drinkers earn 14 per cent more than non-drinkers, the difference being significant at the 0.01 confidence level. The following attempts to duplicate the findings of Hamilton and Hamilton, using a similar data set. This paper will have the following format: Section II will be an overview of the data that is used similar to that utilized in the article, Section III includes our results, and the summary and conclusions stated in Section IV.

II. DATA

In accordance with Hamilton and Hamilton, the data used is from the GSS 1985. The survey questions Canadians about the frequency of their alcohol consumption over the past year as well as the quantity of drinks consumed in the past week. One drink is considered one pint of beer, one glass of wine or 1.5 ounces of spirits. Of the 2,648 males aged twenty-five to fifty-nine in the GSS, 495 were dropped because of missing data, 124 were dropped because salary equalled 0, 282 were excluded because they didn't work at a job last week that was their main activity, and six were excluded because (age-school-6) was negative. The final survey sample used consisted of 1741 males. The data set is comprised of variables to test against annual earnings as income, before taxes, from wages accumulated in the year 1984. The variables used to affect these earnings included age, education, martial status and regional dummy variables, as well as a dummy variable showing whether the individual was born outside Canada. Health status was measured by an individual's ability to perform activities such as walking, climbing stairs, standing for extended periods, etc; and, the number of chronic diseases an individual has reported.

We placed the same restrictions on the survey sample as Hamilton and Hamilton. We limited our sample to males between the ages of twenty-five and fifty-nine, who did earn over three hundred dollars in the previous year, and whose main activity that week was working at their job. We were unable to replicate the exact sample number used in the Hamilton and Hamilton article because we had difficulties identifying all of the 495 individuals with missing data and the 6 people whose (age-school-6) was negative. Our sample consisted of 1,823 males, 17.7 per cent were non-drinkers, 72.2 per cent were moderate-drinkers and 10.1 per cent were heavy-drinkers.

We generated a dummy variable for moderate drinkers and heavy drinkers using non-drinkers as the base case. We also generated dummy variables for marital status, six different age groups, for four regions of Canada (omitting Ontario, Yukon and the N.W.T), three education variables (high-school dropouts, some college excluding a B.A, and university degree or higher). We generated a dummy variable for whether the individual was born outside of Canada; those having difficulties with daily activities; and a variable for chronic diseases (any of high-blood pressure, heart problems, arthritis, rheumatism, bursitis and respiratory problems). Table 1 displays the dummy variables used for our regression, their definitions and their corresponding means and standard deviations.

III RESULTS

When conducting the regression, we regressed log wage against all the dummy variables interacted with both the moderate drinker and heavy drinker variable. Table 2 displays definitions of the interaction terms used in the regression. Table 3 shows the regression results of the dummy variables against log wage and Table 4 displays the results of running three separate regressions of each drinker type. In accordance with the article, we regressed each drinker type separately and achieved numbers very close to that of the article. This is seen in Table 4. The results provided show that for both non- and moderate drinkers, we find that older individuals, as well as those who possess a college diploma, and are married tend to earn more. The results for heavy drinkers however are somewhat different, there is a small increase associated with earnings as one's age increases. Also, there is very little indication of a valuable pay-off for having a college degree or being married as is for the other types of drinkers. Lower earnings for heavy drinkers are attributable to both lower mean characteristics and lower returns to these characteristics relative to moderate drinkers.

Although using somewhat of a larger sample size than that of Hamilton and Hamilton, our regression yielded similar results when testing for three separate drinker types. However, when regressing with only one equation we found numerous difficulties. Our first attempt using interaction variables proved to be a failure, no numbers of use were generated. In the second attempt we took a different approach and set non-drinkers as a base case and regressed moderate and heavy drinkers against this base. This provided us with numbers, however to not achieve perfect collinearity, STATA set all variables for heavy drinkers equal to zero thus cancelling them out. Finally, we were left with one regression yielding different results than the article as they compared moderate drinkers to heavy drinkers and disregarded the non-drinker status. These results were only for moderate drinkers and thus proved to be of little use once again. Therefore, we are forced to use the numbers provided from the three separate regressions of each drinker type and the numbers generated for moderate drinkers.

After conducting a regression for all drinker types, we were able to reject the null hypothesis. An F-test generated to determine whether the coefficients on our explanatory variables in the regression yielded 8.04. This shows that the coefficients for the variables between moderate and heavy drinkers were statistically different and thus implies lower relative returns for heavy versus moderate drinkers. This conclusion was consistent with that of Hamilton and Hamilton, in that they rejected that there is not a difference in earnings pending on drinking types.

IV CONCLUSION

After analyzing the regression results, we find that moderate alcohol consumption leads to increased earnings relative to heavy drinking. Heavy drinking also generates negative returns. These results are not surprising, in that an individual consuming at least eight drinks on any given day would, for example be less productive and thus yield lower earnings. This outcome is in contrast to previous research on the relationship between alcohol and earnings where no negative correlation is found, but is juxtaposed with the chosen article. Berger and Leigh (1988) have previously researched the positive correlation between moderate drinkers and earnings to conclude the beneficial effects of moderate consumption on health and in turn an increase in labor productivity and thus income. In conclusion, we find that heavy drinking leads to reduced earnings relative that of moderate and non-drinking.

APPENDIX A

TABLE 1:
VARIABLE DEFINITIONS, MEANS AND STANDARD DEVIATIONS

Variable Name	Description	Mean	Std. Deviation
lwage	Log earnings		
AGE Dummy Variables (excluded category age 25)			
age 30	age 30 = 1 if age>=30 & age <=35; 0 otherwise	.22092	.41498
age 35	age 35 = 1 if age>=35 & age <=39; 0 otherwise	.19056	.39828
age 40	age 40 = 1 if age>=40 & age <=45; 0 otherwise	.12931	.33561
age 45	age 45 = 1 if age>=45 & age <=49; 0 otherwise	.09074	.28731
age 50	age 50 = 1 if age>= 50 & age <=55; 0 otherwise	.08384	.27667
age 55	age 55 = 1 if age>=55 & age <=59; 0 otherwise	.06397	.24476
EDUCATION Dummy Variables			
Hsdrop	Hsdrop=1 if never graduated from high school; 0 if otherwise	.28176	.44999
Coll Inc	CollInc=1 if some college experience but no B.A. degree; 0 otherwise	.09754	.29677
Cgrad	Cgrad=1 if Bachelor’s degree or higher; 0 otherwise	.14927	.35643
REGIONAL Dummy Variables (excluded categories- Ontario)			
Atlantic	Atlantic=1 if lives in specified provinces; 0 otherwise	.21007	.40745
Quebec	Quebec=1 if lives in specified province; 0 otherwise	.15335	.36041
Prairie	Prairie=1 if lives in specified provinces; 0 otherwise	.52359	.49955
BritCol	BritCol=1 if lives in specified province; 0 otherwise	.11297	.31663
married	married = 1 if married; 0 otherwise	.73956	.43897
Forborn	Forborn=1 if born outside Canada; 0 otherwise	.16197	.36851
Hasadl	Hasadl=1 if has problem with activity of daily living; 0 otherwise	.18239	.38625
Numchron	Numchron=1 if has any of the following chronic diseases: high blood pressure, heart trouble, arthritis, rheumatism, bursitis, respiratory diseases; 0 otherwise	.30989	.46255

TABLE 2:
VARIABLE INTERACTION TERM DEFINITIONS

Interaction Term	Description
mod30	age30 interacted with moderate drinker
heavy30	age30 interacted with heavy drinker
mod35	age35 interacted with moderate drinker
heavy35	age35 interacted with heavy drinker
mod40	age40 interacted with moderate drinker
heavy40	age40 interacted with heavy drinker
mod45	age45 interacted with moderate drinker
heavy45	age45 interacted with heavy drinker
mod50	age50 interacted with moderate drinker
heavy50	age50 interacted with heavy drinker
mod55	age55 interacted with moderate drinker
heavy55	age55 interacted with heavy drinker
modhsdp	Hsdrop interacted with moderate drinker
hvhsdp	Hsdrop interacted with heavy drinker
modclinc	Collinc interacted with moderate drinker
hvcolinc	Collinc interacted with heavy drinker
modcgrad	Cgrad interacted with moderate drinker
hvcgrad	Cgrad interacted with heavy drinker
modatl	Atlantic interacted with moderate drinker
heavyatl	Atlantic interacted with heavy drinker
modque	Quebec interacted with moderate drinker
heavyque	Quebec interacted with heavy drinker
modprai	Prairie interacted with moderate drinker
hvprai	Prairie interacted with heavy drinker
modbrit	BritCol interacted with moderate drinker
hvbrit	BritCol interacted with heavy drinker
modmarry	married interacted with moderate drinker
hvmarry	married interacted with heavy drinker
moddforb	Forborn interacted with moderate drinker
hvdforb	Forborn interacted with heavy drinker
moddadl	Hasadl interacted with moderate drinker
hvdadl	Hasadl interacted with heavy drinker
moddnum	Numchron interacted with moderate drinker
hvdnum	Numchron interacted with heavy drinker

TABLE 3
WAGE REGRESSIONS

Variable	Coefficient	t-statistic
age30	-.017	-0.161
age35	.187	1.825
age40	.149	1.281
age45	.214	1.696
age50	.165	1.254
age55	.032	0.249
hsdrop	-.247	-3.345
collinc	-.287	-2.692
cgrad	.226	2.263
Atlantic	-.153	-1.227
Quebec	.001	0.015
Prairie	.115	1.029
BritCol		
married	.158	2.072
Forborn	-.192	-2.347
Hasadl	-.097	-1.253
Numchron	-.048	-0.704
mod30	.146	1.259
heavy30	dropped	
mod35	.064	0.578
heavy35	dropped	
mod40	.175	1.377
heavy40	dropped	
mod45	.076	0.547
heavy45	dropped	
mod50	.104	0.718
heavy50	dropped	
mod55	.108	0.726
heavy55	dropped	
modhsdp	.002	0.036
hvhsdp	dropped	
modclinc	.292	2.476
hvclinic	dropped	
modcgrad	-.125	-1.153
hvcgrad	dropped	
modmarry	.025	0.309
hvmarry	dropped	

modatl	-.065	-0.512
heavyatl	dropped	
modque	-.117	-0.909
heavyque	dropped	
modprai	-.112	-0.984
hvprai	dropped	
modbrit	.017	0.121
hvbrit	dropped	
moddforb	.128	1.398
hvdforb	dropped	
moddadl	.024	0.283
hvdadl	dropped	
moddnum	.071	0.928
hvdnum	dropped	
_cons	9.925	73.752

TABLE 4
WAGE REGRESSIONS WITH SELECTIVITY CORRECTIONS

Variable	Non-Drinkers		Moderate Drinkers		Heavy Drinkers	
	Coefficient	**t-statistic**	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**
cons	9.926	81.872	9.645	168.496	9.652	72.339
age30	-.017	-0.159	.140	3.017	.146	1.521
age35	. 187	1.801	.266	5.456	.231	1.975
age40	.149	1.264	.356	6.381	.207	1.278
age45	.214	1.673	.335	5.169	.284	0.592
age50	.165	1.238	.288	4.467	. 464	1.700
age55	.032	0.245	. 172	2.288	.006	0.027
Hsdrop	-.247	-3.300	-.251	-6.331	-.251	-2.794
Collinc	-.287	-2.656	-.009	-0.177	.109	0.960
Cgrad	.226	2.233	.108	2.433	.101	0.693
married	.158	2.045	.203	5.525	.134	1.686
Atlantic	-.155	-1.557	-.226	-4.370	.257	1.865
Quebec	-.108	-2.080	.155	2.915	-.111	-2.132
Prairie	.114	1.316	.245	5.624	.360	3.018
BritCol	-.001	-0.014	.240	3.899	.545	3.241
ForBorn	.192	-2.315	-.081	-1.896	.185	1.232
Hasadl	-.097	-1.236	-.082	-1.973	-.028	-0.295
NumChron	-.048	-0.695	.038	1.097	-.077	-0.887
R^2	0.1624		0.1253		0.1612	
No. of Obs	335		1289		199	

APPENDIX B

LOG FILE

```
. drop if var383>1
(6105 observations deleted)

. /* sex variable. restricted to males.*/
.
. drop if var389<3
(763 observations deleted)

. /* drop if age less than 25 years.*/
.
. drop if var389>9
(1684 observations deleted)

. /* drop if age is greater than 60 years*/
.
. drop if var377<300
(129 observations deleted)

. /* drop if income is less than 300*/
.
. drop if var362<2
(101 observations deleted)

. /* drop if have a job but didn't work in reference week.*/
.
. drop if var361>1
(214 observations deleted)

. /* drop if didn't work in  reference week.*/
.
. gen nond=0

. replace nond=1 if var87==4 | var86==2
(437 real changes made)

. /*generated the non-drinking dummy variable*/
.
. gen heavyd=0


.
. /*generated the heavy drinking dummy variable*/
.
. gen modd=0
```

. replace modd=1 if nond==0 & heavyd==0
(1767 real changes made)

. /*generated the moderate drinking dummy variable*/
.
. gen Atlantic=0

. replace Atlantic=1 if var382<4
(463 real changes made)

. /*generated Atlantic dummy variable*/
.
. gen Quebec=0

. replace Quebec=1 if var382==4
(338 real changes made)

. /*generated Quebec dummy variable*/
.
. gen Prairie=0

. replace Prairie=1 if var382>4 & var382<9
(1154 real changes made)

. /*generated Prairie dummy variable*/
.
. gen BritCol=0

. replace BritCol=1 if var382==9
(249 real changes made)

. /*generated British Columbia (BritCol) dummy variable*/
.
. gen Forborn=0

. replace Forborn=1 if var343==13
(357 real changes made)

. /*generated foreign born (Forborn) dummy variable*/
.
. gen Hasadl=0

. replace Hasadl=1 if var29==1 | var30==1 | var31==1 | var32==1 | var33==1 | var34==1 | var35==1 | var36==1 | var37==1 | var38==1 | var39==1 | var40==1 | var41==1 | var42==1 | var43==1 | var44==1 | var17==1| var18==1

```
(402 real changes made)

. /*generated having problems doing daily activities (Hasadl) dummy variable*/
.
. gen Numchron=0

. replace Numchron=1 if var4==1 | var6==1 | var8==1 | var10==1 | var11==1
(683 real changes made)

. /*generated number of chronic diseases (Numchron) dummy variable*/
.
. gen age30=0

. replace age30=1 if var389==4
(487 real changes made)

. /* generate age dummy variable*/
.
. gen age35=0

. replace age35=1 if var389==5
(420 real changes made)

.
. gen age40=0

. replace age40=1 if var389==6
(285 real changes made)

.
. gen age45=0

. replace age45=1 if var389==7
(200 real changes made)

. gen age50=0

. replace age50=1 if var389==8
(184 real changes made)

. gen age55=0

. replace age55=1 if var389==9
(141 real changes made)

. gen married=0
```

```
. replace married=1 if var390==1
(1630 real changes made)

.
. gen hsdrop=0

. replace hsdrop=1 if var419==1
(621 real changes made)

. /*generated high school drop out dummy variable*/
.
. gen collinc=0

. replace collinc=1 if var337==1
(215 real changes made)

. /*generated some college but no BA degree dummy variable*/
.
. gen cgrad=0

. replace cgrad=1 if var340==1
(329 real changes made)

. /*generated B.A. degree dummy variable*/
.
. rename var87 DrinkingFrequency

. rename var343 ForeignBorn

. rename var29 TroubleWalking

. rename var30 UnableWalk

. rename var31 StairTrouble

. rename var32 UnclimbStair

. rename var88 AgeBeganDrinking

. rename var4 HighBlood

. rename var6 HeartTrouble

. rename var8 Diabetes
```

```
. rename var11 Arthritis

. rename var17 Slowdown

. rename var18 CutdownMain

. rename var20 NumDays

. rename var361 WorkRefWeek

. rename var33 Trouble5kg

. rename var34 No5kg

. rename var35 TroubleStandingLong

. rename var36 NoStandLong

. rename var37 TroublePickup

. rename var38 NoPickup

. rename var39 TroubleToenails

. rename var382 Prov

. rename var383 Sex

. rename var389 AgeGroup

. rename var390 MartialStat

. rename var419 Education

. rename var40 Anya

. /* Rename variables*/
.
. gen drink=modd + 2*heavy

-> . rename var377 Wage
-> . edit
- preserve
-> . gen lwage=log(Wage)
(381 missing values generated)
```

```
-> . regress lwage nond age30 age35 age40 age45 age50 age55 hsdrop collinc cgrad
married Atlantic Quebec Prairie BritCol Forborn Hasadl Numchron

  Source |       SS       df       MS              Number of obs =     335
---------+------------------------------           F( 16,   318) =    3.85
   Model |  19.5642713    16  1.22276696           Prob > F      =  0.0000
Residual |  100.907139   318  .317318048           R-squared     =  0.1624
---------+------------------------------           Adj R-squared =  0.1203
   Total |   120.47141   334  .360692846           Root MSE      =  .56331

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    nond |  (dropped)
   age30 |  -.0175859   .1104111     -0.159   0.874      -.2348144    .1996427
   age35 |   .1879179   .1043599      1.801   0.073      -.0174051    .3932409
   age40 |    .149245   .1180513      1.264   0.207      -.0830153    .3815053
   age45 |   .2147938   .1283762      1.673   0.095      -.0377802    .4673677
   age50 |   .1652286   .1334971      1.238   0.217      -.0974205    .4278778
   age55 |   .0325002    .132536      0.245   0.806       -.228258    .2932585
  hsdrop |  -.2478545    .075099     -3.300   0.001      -.3956081   -.1001009
 collinc |  -.2875194   .1082574     -2.656   0.008      -.5005106   -.0745281
   cgrad |   .2262632   .1013468      2.233   0.026       .0268681    .4256582
 married |   .1587477   .0776353      2.045   0.042       .0060039    .3114915
Atlantic |  -.1553606   .0997827     -1.557   0.120      -.3516782     .040957
  Quebec |  (dropped)
 Prairie |    .114088   .0867231      1.316   0.189      -.0565356    .2847116
 BritCol |  -.0018317   .1278331     -0.014   0.989      -.2533373    .2496738
 Forborn |  -.1928535   .0832882     -2.315   0.021      -.3567191    -.028988
  Hasadl |  -.0978583   .0791416     -1.236   0.217      -.2535657     .057849
Numchron |  -.0486508   .0700369     -0.695   0.488       -.186445    .0891435
   _cons |   9.926882   .1212495     81.872   0.000       9.688329    10.16543
------------------------------------------------------------------------------
-> . regress lwage modd age30 age35 age40 age45 age50 age55 hsdrop collinc cgrad
married Atlantic Quebec Prairie BritCol Forborn Hasadl Numchron

  Source |       SS       df       MS              Number of obs =    1289
---------+------------------------------           F( 16,  1272) =   11.39
   Model |  56.7029535    16  3.54393459           Prob > F      =  0.0000
Residual |   395.74211  1272  .311118011           R-squared     =  0.1253
---------+------------------------------           Adj R-squared =  0.1143
   Total |  452.445063  1288  .351277223           Root MSE      =  .55778

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
```

```
    modd |  (dropped)
   age30 |   .1400242   .0464169     3.017   0.003       .048962    .2310863
   age35 |   .2668131   .0489026     5.456   0.000      .1708745    .3627517
   age40 |   .3562437   .0558328     6.381   0.000      .2467092    .4657783
   age45 |   .3358475   .0649772     5.169   0.000      .2083731    .4633218
   age50 |    .288303   .0645424     4.467   0.000      .1616817    .4149244
   age55 |   .1725517   .0754315     2.288   0.022      .0245679    .3205354
  hsdrop |  -.2517987   .0397692    -6.331   0.000     -.3298191   -.1737782
 collinc |  -.0098637   .0557576    -0.177   0.860     -.1192507    .0995233
   cgrad |   .1080163   .0443915     2.433   0.015      .0209276     .195105
 married |   .2037237   .0368753     5.525   0.000      .1313806    .2760668
Atlantic |  (dropped)
  Quebec |   .1557527   .0534276     2.915   0.004      .0509367    .2605687
 Prairie |   .2451456   .0435921     5.624   0.000      .1596253     .330666
 BritCol |   .2409912   .0618148     3.899   0.000      .1197211    .3622613
 Forborn |  -.0810919   .0427653    -1.896   0.058     -.1649901    .0028063
  Hasadl |  -.0823904   .0417563    -1.973   0.049     -.1643092   -.0004715
Numchron |   .0389979   .0355423     1.097   0.273     -.0307301    .1087258
   _cons |   9.645908    .057247   168.496   0.000      9.533599    9.758217
------------------------------------------------------------------------------
-> . regress lwage heavyd age30 age35 age40 age45 age50 age55 hsdrop collinc cgrad
married Atlantic Quebec Prairie BritCol Forborn Hasadl Numchron

  Source |       SS       df       MS              Number of obs =     199
---------+------------------------------           F( 16,   182) =    2.19
   Model |  9.47484344    16  .592177715           Prob > F      =  0.0069
Residual |  49.2869723   182   .27080754           R-squared     =  0.1612
---------+------------------------------           Adj R-squared =  0.0875
   Total |  58.7618157   198  .296776847           Root MSE      =  .52039

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  heavyd |  (dropped)
   age30 |   .1463849   .0962543     1.521   0.130     -.0435331    .3363028
   age35 |    .231766   .1173413     1.975   0.050      .0002417    .4632903
   age40 |   .2076696   .1624902     1.278   0.203     -.1129373    .5282765
   age45 |    .099064   .1673178     0.592   0.555     -.2310681    .4291962
   age50 |   .4643928   .2732189     1.700   0.091      -.074691    1.003477
   age55 |    .006355   .2336667     0.027   0.978     -.4546891     .467399
  hsdrop |  -.2515439   .0900409    -2.794   0.006     -.4292021   -.0738857
 collinc |   .1093947   .1139983     0.960   0.339     -.1155337     .334323
   cgrad |   .1016991    .146855     0.693   0.489     -.1880581    .3914563
 married |    .134393   .0797034     1.686   0.093     -.0228685    .2916544
Atlantic |   .2575978   .1381544     1.865   0.064     -.0149924     .530188
  Quebec |  (dropped)
```

```
 Prairie |   .3607328   .1195374      3.018   0.003       .1248755      .59659
 BritCol |   .5458372   .1684046      3.241   0.001       .2135608    .8781135
 Forborn |   .1851681   .1502817      1.232   0.219      -.1113504    .4816866
  Hasadl |  -.0284356   .0965523     -0.295   0.769      -.2189413    .1620702
Numchron |  -.0770195   .0868196     -0.887   0.376      -.2483217    .0942828
   _cons |   9.652593   .1334348     72.339   0.000       9.389315    9.915871
------------------------------------------------------------------------------
-> . gen mod30=age30*modd
-> . gen heavy30=age30*heavyd
-> . gen mod35=age35*modd
-> . gen heavy35=age35*heavyd
-> . gen mod40=age40*modd
-> . gen heavy40=age40*heavyd
-> . gen mod45=age45*modd
-> . gen heavy45=age45*heavyd
-> . gen mod50=age50*modd
-> . gen heavy50=age50*heavyd
-> . gen mod55=age55*modd
-> . gen heavy55=age55*heavyd
-> . gen modhsdp=hsdrop*modd
-> . gen hvhsdp=hsdrop*heavyd
-> . gen modclinc=collinc*modd
-> . gen hvcolinc=collinc*heavyd
-> . gen modcgrad=cgrad*modd
-> . gen hvcgrad=cgrad*heavyd
-> . gen modmarry=married*modd
-> . gen hvmarry=married*heavyd
-> . gen modatl=Atlantic*modd
-> . gen heavyatl=Atlantic*heavyd
-> . gen modque=Quebec*modd
-> . gen heavyque=Quebec*heavyd
-> . gen modprai=Prairie*modd
-> . gen hvprai=Prairie*heavyd
-> . gen modbrit=BritCol*modd
-> . gen hvbrit=BritCol*heavyd
-> . gen moddforb=Forborn*modd
-> . gen hvdforb=Forborn*heavyd
-> . gen moddadl=Hasadl*modd
-> . gen hvdadl=Hasadl*heavyd
-> . gen moddnum=Numchron*modd
-> . gen hvdnum=Numchron*heavyd
-> . set matsize 150
-> . regress lwage age30 age35 age40 age45 age50 age55 hsdrop collinc cgrad married
Atlantic Quebec Prairie BritCol Forborn Hasadl Numchron mod30 heavy30 mod35
heavy35 mod40 heavy40 mod45 heavy45 mod50 heavy50 mod55 heavy55 modhsdp
hvhsdp modclinc modcgrad hvcgrad modmarry hvmarry modatl heavyatl modque
```

heavyque modprai hvprai modbrit hvbrit moddforb hvdforb moddadl hvdadl moddnum hvdnum

Source	SS	df	MS
Model	81.951781	33	2.4833873
Residual	552.709242	1789	.30894871
Total	634.661023	1822	.348332065

Number of obs = 1823
F(33, 1789) = 8.04
Prob > F = 0.0000
R-squared = 0.1291
Adj R-squared = 0.1131
Root MSE = .55583

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age30	-.0175859	.1089453	-0.161	0.872	-.2312594	.1960876
age35	.1879179	.1029744	1.825	0.068	-.0140448	.3898807
age40	.149245	.1164841	1.281	0.200	-.0792143	.3777042
age45	.2147938	.1266719	1.696	0.090	-.0336467	.4632342
age50	.1652286	.1317248	1.254	0.210	-.0931221	.4235793
age55	.0325002	.1307765	0.249	0.804	-.2239905	.288991
hsdrop	-.2478545	.074102	-3.345	0.001	-.39319	-.102519
collinc	-.2875194	.1068202	-2.692	0.007	-.4970249	-.0780139
cgrad	.2262632	.1000014	2.263	0.024	.0301314	.422395
married	.1587477	.0766047	2.072	0.038	.0085037	.3089917
Atlantic	-.1535289	.1251704	-1.227	0.220	-.3990244	.0919666
Quebec	.0018317	.126136	0.015	0.988	-.2455578	.2492212
Prairie	.1159197	.1126149	1.029	0.303	-.1049509	.3367904
BritCol	(dropped)					
Forborn	-.1928535	.0821825	-2.347	0.019	-.3540373	-.0316697
Hasadl	-.0978583	.078091	-1.253	0.210	-.2510174	.0553007
Numchron	-.0486508	.0691071	-0.704	0.482	-.1841899	.0868884
mod30	.1469242	.1167169	1.259	0.208	-.0819915	.3758399
heavy30	(dropped)					
mod35	.0649295	.1123094	0.578	0.563	-.1553419	.2852009
heavy35	(dropped)					
mod40	.1757567	.1276507	1.377	0.169	-.0746035	.4261168
heavy40	(dropped)					
mod45	.0767092	.1402949	0.547	0.585	-.19845	.3518684
heavy45	(dropped)					
mod50	.1043079	.1453213	0.718	0.473	-.1807094	.3893252
heavy50	(dropped)					
mod55	.1081055	.1489344	0.726	0.468	-.1839982	.4002093
heavy55	(dropped)					
modhsdp	.0029969	.082519	0.036	0.971	-.158847	.1648407
hvhsdp	(dropped)					
modclinc	.2921562	.1180147	2.476	0.013	.060695	.5236173
modcgrad	-.1252628	.1085992	-1.153	0.249	-.3382573	.0877317

```
 hvcgrad |  (dropped)
modmarry |   .0257815   .0835148     0.309   0.758     -.1380152    .1895783
 hvmarry |  (dropped)
  modatl |  -.0650721   .1270059    -0.512   0.608     -.3141676    .1840233
heavyatl |  (dropped)
  modque |  -.1178476   .1297156    -0.909   0.364     -.3722576    .1365624
heavyque |  (dropped)
 modprai |  -.1126564   .1145084    -0.984   0.325     -.3372408    .1119279
  hvprai |  (dropped)
 modbrit |   .0176178   .1462023     0.121   0.904     -.2691274    .3043629
  hvbrit |  (dropped)
moddforb |   .1283598   .0918084     1.398   0.162     -.0517032    .3084229
 hvdforb |  (dropped)
 moddadl |   .0246497    .086988     0.283   0.777      -.145959    .1952585
  hvdadl |  (dropped)
 moddnum |   .0710525   .0765409     0.928   0.353     -.0790665    .2211715
  hvdnum |  (dropped)
   _cons |    9.92505   .1345733    73.752   0.000      9.661113    10.18899
------------------------------------------------------------------------------
```

-> . test age30 age35 age40 age45 age50 age55 hsdrop collinc cgrad married Atlantic Quebec Prairie BritCol Forborn Hasadl Numchron mod30 heavy30 mod35 heavy35 mod40 heavy40 mod45 heavy45 mod50 heavy50 mod55 heavy55 modhsdp hvhsdp modclinc modcgrad hvcgrad modmarry hvmarry modatl heavyatl modque heavyque modprai hvprai modbrit hvbrit moddforb hvdforb moddadl hvdadl moddnum hvdnum

```
 ( 1)  age30 = 0.0
 ( 2)  age35 = 0.0
 ( 3)  age40 = 0.0
 ( 4)  age45 = 0.0
 ( 5)  age50 = 0.0
 ( 6)  age55 = 0.0
 ( 7)  hsdrop = 0.0
 ( 8)  collinc = 0.0
 ( 9)  cgrad = 0.0
 (10)  married = 0.0
 (11)  Atlantic = 0.0
 (12)  Quebec = 0.0
 (13)  Prairie = 0.0
 (14)  BritCol = 0.0
 (15)  Forborn = 0.0
 (16)  Hasadl = 0.0
 (17)  Numchron = 0.0
 (18)  mod30 = 0.0
 (19)  heavy30 = 0.0
 (20)  mod35 = 0.0
 (21)  heavy35 = 0.0
```

(22) mod40 = 0.0
(23) heavy40 = 0.0
(24) mod45 = 0.0
(25) heavy45 = 0.0
(26) mod50 = 0.0
(27) heavy50 = 0.0
(28) mod55 = 0.0
(29) heavy55 = 0.0
(30) modhsdp = 0.0
(31) hvhsdp = 0.0
(32) modclinc = 0.0
(33) modcgrad = 0.0
(34) hvcgrad = 0.0
(35) modmarry = 0.0
(36) hvmarry = 0.0
(37) modatl = 0.0
(38) heavyatl = 0.0
(39) modque = 0.0
(40) heavyque = 0.0
(41) modprai = 0.0
(42) hvprai = 0.0
(43) modbrit = 0.0
(44) hvbrit = 0.0
(45) moddforb = 0.0
(46) hvdforb = 0.0
(47) moddadl = 0.0
(48) hvdadl = 0.0
(49) moddnum = 0.0
(50) hvdnum = 0.0

Constraint 14 dropped
Constraint 19 dropped
Constraint 21 dropped
Constraint 23 dropped
Constraint 25 dropped
Constraint 27 dropped
Constraint 29 dropped
Constraint 31 dropped
Constraint 34 dropped
Constraint 36 dropped
Constraint 38 dropped
Constraint 40 dropped
Constraint 42 dropped
Constraint 44 dropped
Constraint 46 dropped
Constraint 48 dropped
Constraint 50 dropped

F(33, 1789) =   8.04
Prob > F =   0.0000

Volunteer Works

in

Estimating

Males and Females

Earnings Gap

Survey of Volunteer Activity (VAS) – Volunteer File

Model Paper: Can volunteer work help explain the male-female earnings gap?
By Kathleen M. Day and Rose Anne Devlin.
Applied Economics, 1997, 29, 707-721

Introduction

For many years, researchers have tried to identify the causes of male-female wage differential, yet the goal of researchers is not achieved. A significant portion of the gap still remains unexplained. One difficulty that researchers might encounter while conducting a research on male-female wage differential is unavailability of sufficient information of individual's experience. Thus, instead of using inadequate information of individual's experience, Day and Devlin use another component of experience, volunteer work, to test if it is the cause the wage differential gap. Before doing that, three possible ways through which volunteering may affect earnings are introduced: first, volunteer work may provide individuals with an alternative means of acquiring skills and experience that make them more productive (the human capital hypothesis). Secondly, volunteering may influence earnings by providing a signal to employers of otherwise unobservable ability (the screening hypothesis). Finally, volunteering may provide access to informal networks of contacts (the networking hypothesis). Combining all three hypotheses, Day and Devlin wish to find some evidence that can explain male-female wage differential gap. In their paper, they examine the differential returns to volunteer work in the paid labor market and use the finding to explain whether the volunteer work affects the male-female earnings gap or not. As the result, one third of the male-female earnings gap may be caused by the different rewards to male-female volunteers.

Day and Devlin followed two approaches. The first was to add dummy variable representing volunteer experience to the male and female earnings equations; the second was to estimate separate earnings equations for male volunteers and non-volunteers, and female volunteers and non-volunteers. To explore the earnings gap, Day and Devlin use the earnings equation as follows:

$$\overline{\ln W^j} - \overline{\ln W^k} = \hat{\beta}^j(\overline{X}^j - \overline{X}^k) + (\hat{\beta}^j - \hat{\beta}^k)\overline{X}^k$$

where $\hat{\beta}^j$ is a row vector containing the estimated coefficients of the earnings equation for group j and $\overline{X}^j$ is a column vector containing the sample means of the explanatory variables for group j. The first term in the equation represents that portion of the differential in earnings between the groups j and k arising from differences in their stock of characteristics, while the second term of the differential is attributable to differences in the returns to those characteristics.

The data are drawn from the 1987 Survey of Volunteer Activity (VAS). Day and Devlin draw two subsamples of the VAS data set. The first sample was taken from the VAS Screening File. The subsample contains 5057 individuals, of whom 3374 were volunteers, 1683 were non-volunteers, 1956 were females and 3101 were males. The second sample was drawn from the VAS Volunteer File. The subsample consisted of 3687 individuals, of whom 2004 were volunteers, 1683 were non-volunteers, 1397 were females, and 2290 were males.

In their wage differential analysis, they first test married, education, family size, children, experience, city size, occupation, province, and constant as for human capital hypothesis. Two results are found. The first result is that the differential in log earnings between male volunteers and non-volunteers is 0.1950 or 21.53%. In other words, male volunteers earn 21.53% higher incomes than non-volunteers. The second result, a positive value of return effect (0.0976), which is derived from the calculation as well, indicates that male volunteers earn higher returns than male non-volunteers. On the other hand, the earnings differential between female volunteers and non-volunteers is smaller compared to the range between male volunteer and non-volunteer. The finding is only 0.1063 or 11.22%. This implies that female volunteers earn 11.22% higher incomes than non-volunteers. Nevertheless, a negative value of return effect (-0.0054) means that female rewarded is offset by a large and negative effect associated with the family size variable. This effect is that females receive almost no return to volunteering on the labour market. The findings for both male and female suggest that volunteering enhances males employment earnings but not to females.

Next, they use 4 variables, recreation, economy, religious, and multi-domain to test their screening hypothesis. By using screening hypothesis, four of the organizations have statistically significant coefficients in one or both equations. The result reports that women's returns to participating in recreational organizations are only slightly lower than those of men (0.087308 vs. 0.089966), their participation in recreational organization is significantly lower for women (16.43%) than for men (28.57%). These results suggest that the treatment of women who volunteer for recreational organizations is the same as that of their male counterparts on the labour market.

Furthermore, the statistics values of participation in a multi-domain type of organization show that more men are involved than women (15.13% compared to 9.55%). This appears to be some discrimination against women who participate in certain types of organizations and activities. When taking the type of volunteer organization into account, men receive a return to volunteering of about half of the total mean differential between volunteers and non-volunteers (0.0528 + 0.0472), while the return to women is once again negative (-0.0033 + 0.00004). This means that women' participation in recreational and economy related organizations are cancelled out by the negative effects of religious volunteering. When taking activities into account, about 60% of the mean earnings differential between volunteers and no-volunteers constitutes a return to volunteering for men, as compared to only 4% for women.

Lastly, using the standard decomposition of the male-female earnings gap, results indicate that the mean earnings differential (in logs) between males and females is 0.2613 or 29.90%. Of this differential, 68% is attributable to labour market discrimination ('the return' effect), which means that the discrimination is captured by the constant term.

When adding a dummy variable to the male and female earnings equations, the share of the earnings differential due to discrimination remains unchanged. The differential returns to volunteering (0.0651) account for 36% of discrimination against women. In other words, if the labour market returns to volunteer work were the same for men and women, the earnings gap would be reduced from 29.90% to 19.14%.

We choose Day and Devlin's paper for our model paper, because we want to know whether volunteer experience affects earnings or not. Wage differentials between males and females exist in Canada. Can volunteer experience reduce or narrow the wage gap? We will find the answer through our estimation in this paper.

This paper contains five sections which are introduction, data, model, result and conclusion. The data section of the paper describes data set that we use in this paper. The data set is drawn from the one that Day and Devlin use in their paper. However, we only use the VAS Volunteer File. The sample size in our paper is different from Day and Devlin's. As Day and Devlin, we focus on the sole wage earner in households and volunteers who volunteered during the year of survey and had volunteered in the past.

The model section will present the model or earnings equation we use in this paper. Our earnings equation is similar to Day and Devlin's. However, we just run one pooled regression equation to get the result. The equation contains four groups of individuals: male non-volunteers, male volunteers, female volunteers, and female non-volunteers. The result is shown in table B in the appendix section.

The fourth section will present the result of our estimation. The results of hypothesis tests shows that wage differentials between these four groups exist. Male volunteers benefit more with their volunteer experience than female volunteers. Due to different sample and earnings equation, our result is slightly different from Day and Devlin's paper.

The last part of the paper concludes our result and our difficulties. We learn that it is hard to get the similar result to the model paper. The data we drawn, the equation we use, and the result we get are different from the model paper.

Data

The date are drawn from the 1987 Survey of Volunteer Activity (VAS) conducted by Statistics Canada in conjunction with its November 1987 Labour Force Survey (LFS). The VAS contains data on the labour force characteristics as well as volunteer activities of individuals. In this paper, we choose to include volunteers who volunteered during the year of the survey and had volunteered at some time in the past. Although the VAS provides some information on both the type of volunteer organization and the type of volunteering activities which Day and Devlin use, we are not going to use those variables in order to simplify our estimation in our paper.

In their paper, Day and Devlin derive results from two subsamples of the VAS

Screening File and VAS Volunteer File. In this paper, we only use the sample that was drawn from the VAS Volunteer File. The VAS Volunteer File contains the responses of all non-volunteers to the VAS screening questionnaire, as well as the responses of those volunteers who returned the more detailed follow-up questionnaire. This file also provides variables, which state whether individuals volunteered during the year of survey and had volunteered in the past, so we choose this file to be used in our paper.

After excluding individuals who were not the sole wage earner in their households, we were left with a sample of 6878 individuals, of whom 1258 were male volunteers, 1816 were female volunteers, 1884 were male non-volunteers, and 1920 were female non-volunteers.

Model

The regression equation is similar to the one that Day and Devlin use in their paper, but we just generate one pooled regression equation to estimate wage differentials between four groups of individuals. They are male non-volunteers, male volunteers, female non-volunteers, and female volunteers. We use male non-volunteers as a base group in the pooled regression equation. The earnings equation is in log shown as follows:

$$
\begin{aligned}
\text{lnwage} = {} & \beta_1 + \beta_2\text{EXP} + \beta_3\text{EXP}^2 + \beta_4\text{EDUCATION} + \beta_5\text{LANGUAGE} + \beta_6\text{FAMSIZE} + \beta_7\text{KIDSOWN} \\
& + \beta_8\text{KIDSADSCH} + \beta_9\text{PROVINCES} + \beta_{10}\text{MV} + \beta_{11}\text{MVEXP} + \beta_{12}\text{MVEXP}^2 \\
& + \beta_{13}\text{MVEDUCATION} + \beta_{14}\text{MVLANGUAGE} + \beta_{15}\text{MVFAMSIZE} + \beta_{16}\text{MVKIDSOWN} \\
& + \beta_{17}\text{MVKIDSADSCH} + \beta_{86}\text{MVPROVINCES} + \beta_{19}\text{FV} + \beta_{20}\text{FVEXP} + \beta_{21}\text{FVEXP}^2 \\
& + \beta_{22}\text{FVEDUCATION} + \beta_{23}\text{FVLANGUAGE} + \beta_{24}\text{FAMSIZE} + \beta_{25}\text{FVKIDSOWN} \\
& + \beta_{26}\text{FVKIDSADSCH} + \beta_{27}\text{FVPROVINCES} + \beta_{28}\text{FNV} + \beta_{29}\text{FNVEXP} + \beta_{30}\text{FNVEXP}^2 \\
& + \beta_{31}\text{FNVEDUCATION} + \beta_{32}\text{FNVLANGUAGE} + \beta_{33}\text{FAMSIZE} + \beta_{34}\text{FNVKIDSOWN} \\
& + \beta_{35}\text{FNVKIDSADSCH} + \beta_{36}\text{FNVPROVINCES} + \mu_I
\end{aligned}
$$

where education, language, kidsown, and provinces are dummy variables. Moreover, these dummy variables contain four levels of education, three types of language, the number of children in four different age groups, and province of residence. Experience is calculated by 'age – schooling – 6'. The pooled equation includes three interaction terms, which are male volunteers, female volunteers, and female non-volunteers. Those interaction terms are to show the wage differentials between male non-volunteers and volunteers, male non-volunteers and female volunteers, and male non-volunteers and female non-volunteers. Table A shows all variables that include in the regression equation.

Result

To see whether volunteer experience affects the earnings differential between male and female, we use a pooled regression equation which contains four groups of individuals to calculate the earnings gap. The result is reported in table B, Appendix.

First, we look at the wage gap between male volunteers and non-volunteer. The initial earnings gap between these two groups is given by the MV coefficient and the MV-interaction terms. The result shows that male volunteers generally earn more than male non-volunteers. However, according to our result, male volunteers who are English speakers and have kids aged 0 to 5 years earn less than male non-volunteers. English speaking volunteers earn 4% less than English speaking non-volunteers, and volunteers with kids aged 0 to 5 years earn 15 to 17% less than non-volunteers. Volunteer experience seems to have less impact on English speakers than non-English speakers. On the other hand, the wage differential between male volunteers and non-volunteers who are other language speakers is 0.2777. In other words, if individuals are other language speakers, their earnings will increase by 27.77% with volunteer experience. Volunteers who have kids aged from 6 to 24 years earn 2 to 5% more than non-volunteers. This situation happens because if volunteers who have kids under aged 5, they have to arrange time among working, volunteering, and caring for their infants. They do not have extra time to devote themselves on working in order to earn more income. As their children get older and more independent, individuals will have extra time to work and volunteer; thus, volunteers will earn more. Although volunteer experience has positive impact on individuals, those individuals who live in far-east provinces, such as Newfoundland and PEI, earn less than those individuals who live in far-west provinces, such as Alberta and B.C. In larger provinces, due to a strong competition between individuals, volunteer experience is important to individuals when they seek for jobs. More volunteer experience means more working experience, so that volunteers tend to acquire higher wages than non-volunteers. However, in small provinces, less competition between individuals leads to less advantages of volunteer experience. Thus, volunteers who live in large provinces will earn much more than those who live in small provinces. In general, male volunteers tend to earn more than male non-volunteers, because they are better educated compared to non-volunteers. Hence, there is a positive wage gap between male volunteers and male non-volunteers.

We, then look at the wage differential among male non-volunteers, female non-volunteers, and female volunteers. The initial gap is indicated by the FV and FNV coefficients and the FV and FNV interaction terms. We have found that volunteer work do not benefit females. Females with volunteer works still earn less than male non-volunteers. Female volunteers with high education still obtain fewer wages than male non-volunteers and female non-volunteers. The possible reason is that female volunteers are less educated than female non-volunteers and male non-volunteers in our sample, so the wage differential between male non-volunteers and female volunteers is larger than the wage gap between male non-volunteers and female non-volunteers.

Females who speak English, French, or other language earn more than male non-volunteers regardless volunteer experience. In some occupations, females with or without volunteer experience will still earn 1 to 20% more than male non-volunteers, such as salesperson and receptionists. For females who have kids aged 0 to 24 years, they still earn less than male non-volunteers, regardless of their volunteer experience. In tradition, females have to stay at home and are responsible for caring for their kids and their husbands. Even though their kids are getting independent, females still have to take care of their daily needs, such as meals and housework. Hence, females do not have extra time to devote on working or volunteering; their earnings are 2 to 8% lower than male non-volunteers' earnings. The results are different from male volunteers. When children are getting older, males can devote more time on working and volunteering than females. Although female non-volunteers and volunteers generally earn less than males non-volunteers, volunteer experience has positive impacts on the provinces of residence of females. Living in far-east provinces has larger positive effects on earnings than living in far-west provinces.

Although the experience and languages of female volunteers is rewarded more highly than that of non-volunteers, they are offset by a large and negative effect associated with the number of kids and education variables. The net effect is that females receive almost no return to volunteering on the labour market. Hence, volunteer experience has similar negative effects on females, which means that female volunteers and female non-volunteers earn less than male volunteers.

We have done the hypothesis tests to test that whether volunteer works can affect the male and female wage gap.

H_0 : the coefficients of the interaction terms are all equal to zero
H_1 : at least one of the coefficients of the interaction terms is not equal zero

The result shows that the coefficients of the interaction terms are statistically significant from zero (p-value is 0.000). That means that there are wage differentials between male non-volunteers, male volunteers, female non-volunteers and female volunteers. Volunteer experience does not benefit female in the labour market. Also, we have tested separately whether there are wage differential between male non-volunteers and volunteers, between male non-volunteers and female volunteers, and male non-volunteers and female non-volunteers. We have found that three null hypothesis are rejected. The coefficients of the MV interaction terms, the FV interaction terms, and the FNV interaction terms are all statistically different from zero. Hence, volunteer experience does not change the wage gaps between males and females.

Conclusion

In our paper, we find that volunteer experience cannot narrow wage differential between males and females. Our results are consistent with Day and Devlin's. Volunteer works has a positive return on the labour market for men but not for women. Males will earn more if they have participated in volunteer works, compared with male non-volunteers; yet, volunteer works have no return on females. Hence, the wage gap between males and female get larger in account for volunteer works. Also, we have found that the effects on female non-volunteers and volunteer are similar. The possible reason is that female volunteers posses the same or even worse quality characteristics than female non-volunteers, so we have found that volunteer works contribute the same effect on female non-volunteers and volunteers.

We have learned that it was hard to find the same number of observation as the model paper, because Day and Devlin did not state clearly where those variables came from. We took similar variables, but we got different sample size. The result we found was slightly different from the model paper. Although we got different result from the model paper, we are interested in trying to find similar data and result to the model paper.

In their paper, Day and Devlin selected three estimations. However, in our paper, we just focus on estimating whether volunteer works narrow the wage gap between males and females. We did not include types of voluntary organization and types of volunteer activity. If we include those variables, we will make the earning equation more complicate. Hence, we only simply run a pooled regression equation and see whether volunteer works affect males and females wage gaps.

Appendix

Table A. Variable names and definations

Variables	Description	Sample Mean
Wage	Midpoint of range of seven reported income classes	28616.2
Female	Dummy variable: 1 if female, 0 otherwise	0.5431812
MV	Dummy variable: 1 if male volunteer, 0 otherwise	0.182902
MNV	Dummy variable: 1 if male non-volunteer, 0 otherwise	0.2739168
FV	Dummy variable: 1 if female volunteer, 0 otherwise	0.2640302
FNV	Dummy variable: 1 if female non-volunteer, 0 otherwise	0.2791509
Exp	Experience	25.88674
Expsq	Experience square	876.2526
Education:		
HIGHSH	Dummy variable: 1 if high school (some or complete), 0 otherwise	0.6304158
POSTSE	Dummy variable: 1 if some post-secondary education, 0 otherwise	0.0933411
DIPLO	Dummy variable: 1 if post-secondary diploma, 0 otherwise	0.1443734
UNIVER	Dummy variable: 1 if university degree, 0 otherwise	0.1318697
Language:		
ENGLIS	Dummy variable: 1 if language spoken at home is English, 0 otherwise	1.216924
FRENCH	Dummy variable: 1 if language spoken at home is Frence, 0 otherwise	1.805757
OTHLAN	Dummy variable: 1 if neither English nor French is spoken at home, 0 otherwise	1.96467
FAMSIZ	Number of individuals residing in the household	2.69686
Kids Own:		
KIDS1	Number of own children aged 0-2 years	0.1788311
KIDS2	Number of own children aged 3-5 years	0.1738878
KIDS3	Number of own children aged 6-15 years	0.5231172
KIDSAT	Number of children aged 16-24 attending school	0.1420471
Provinces:		
PROV1	Dummy variable: 1 if the province is Newfoundland, 0 otherwise	0.074731
PROV2	Dummy variable: 1 if the province is P.E.I., 0 otherwise	0.0191916
PROV3	Dummy variable: 1 if the province is Nova Scotia, 0 otherwise	0.0681884
PROV4	Dummy variable: 1 if the province is New Brunswick, 0 otherwise	0.0734225
PROV5	Dummy variable: 1 if the province is Quebec, 0 otherwise	0.1703984
PROV6	Dummy variable: 1 if the province is Ontario, 0 otherwise	0.1766502
PROV7	Dummy variable: 1 if the province is Manitoba, 0 otherwise	0.0738587
PROV8	Dummy variable: 1 if the province is Saskatchewan, 0 otherwise	0.0905787
PROV9	Dummy variable: 1 if the province is Alberta, 0 otherwise	0.1388485
PROV10	Dummy variable: 1 if the province is B.C., 0 otherwise	0.114132

Table B. Pooled Regression Equations for Earnings Equation for Male and Female Volunteers and Non-volunteers in log

Variables		Coefficient	Standard Error
Male non-volunteer	Experience	0.0172714*	0.0036864
	Experience square	-0.0002314*	0.0000602
	with high schools education only	-0.7930283	1.028375
	with some post-secondary education	-0.6085904	1.029336
	with post-secondary diploma	-0.5724835	1.029609
	with university degree	-0.3827609	1.029367
	English speaker	0.201768**	0.1063885
	French speaker	0.1830338	0.1165685
	Other language speaker	0.3829137*	0.0996634
	Size of family	0.0906728*	0.0170922
	Number of own children age 0-2	-0.0420437	0.0361049
	Number of own children age 3-5	-0.0296613	0.376273
	Number of own children age 6-15	-0.0015329	0.0208106
	Number of children 16-24 attending school	0.0069882	0.347504
	In province of Newfoundland	-0.0733663	0.967958
	In province of P.E.I.	**Dropped**	**Dropped**
	In province of Nova Scotia	0.1294816	0.0993639
	In province of New Brunswick	0.0218783	0.0989384
	In province of Quebec	0.2073272**	0.1063007
	In province of Ontario	0.3572267*	0.0916138
	In province of Manitoba	0.2103777*	0.0989447
	In province of Saskatchewan	0.2490585*	0.0978077
	In province of Alberta	0.2100216*	0.0941095
	In province of British Columbia	0.3197017*	0.0938421
Male volunteer		**Dropped**	**Dropped**
	Experience	0.0196792*	0.0051142
	Experience square	-0.0002673*	0.0000834
	with high schools education only	0.0705353	0.052935
	with some post-secondary education	**Dropped**	**Dropped**
	with post-secondary diploma	0.1688454*	0.0612862
	with university degree	0.3627873*	0.0566985
	English speaker	-0.0453145	0.1882461
	French speaker	0.0115812	0.1923823
	Other language speaker	0.2776568	0.1809811
	Size of family	0.607735*	0.239789
	Number of own children age 0-2	-0172733	0.0429104
	Number of own children age 3-5	-0.146204	0.0403437
	Number of own children age 6-15	0.0456517**	0.0246749
	Number of children 16-24 attending school	0.0209774	0.045667
	In province of Newfoundland	0.0998154	0.1506925
	In province of P.E.I.	**Dropped**	**Dropped**
	In province of Nova Scotia	0.2515977**	0.1488522
	In province of New Brunswick	0.3173755*	0.1513647
	In province of Quebec	0.3388386*	0.160607
	In province of Ontario	0.468882*	0.1420137
	In province of Manitoba	0.38559*	0.146004
	In province of Saskatchewan	0.2389253**	0.1439889
	In province of Alberta	0.3273247*	0.1412649
	In province of British Columbia	0.4435922*	0.1427984
Female volunteer		-0.0380826	1.0884
	Experience	0.0313845*	0.0045586
	Experience square	-0.0004589*	0.0000765
	with high schools education only	-0.3343738*	0.380135
	with some post-secondary education	-0.2598206*	0.0504793
	with post-secondary diploma	-0.1906531*	0.043263

	with university degree	**Dropped**	**Dropped**
	English speaker	0.1337197	0.1356852
	French speaker	0.1126686	0.1350572
	Other language speaker	0.1936609	0.1374526
	Size of family	0.1810094*	0.020187
	Number of own children age 0-2	-0.0212367	0.0368781
	Number of own children age 3-5	-0.0528429	0.0327745
	Number of own children age 6-15	-0.040641**	0.208445
	Number of children 16-24 attending school	-0.0764488*	0.035714
	In province of Newfoundland	-0.1168016	0.1015597
	In province of P.E.I.	**Dropped**	**Dropped**
	In province of Nova Scotia	0.0732168	0.099495
	In province of New Brunswick	0.0154526	0.1030586
	In province of Quebec	0.0572507	0.113699
	In province of Ontario	0.2140672*	0.0941572
	In province of Manitoba	0.0416957	0.0994003
	In province of Saskatchewan	0.0069937	0.0967014
	In province of Alberta	0.0770842	0.0939189
	In province of British Columbia	0.069854	0.0960784
Female non-volunteer		-0.1263828	1.035858
	Experience	0.0172368*	0.0036432
	Experience square	-0.0002349*	0.0000593
	with high schools education only	-0.0965939*	0.0466376
	with some post-secondary education	**Dropped**	**Dropped**
	with post-secondary diploma	0.1145373*	0.055643
	with university degree	0.2681002*	0.064838
	English speaker	0.0196446	0.1105807
	French speaker	0.1593518	0.1152824
	Other language speaker	0.1144466	0.1059502
	Size of family	0.1763353*	0.0185183
	Number of own children age 0-2	-0.0548205	0.0346519
	Number of own children age 3-5	-0.0783378*	0.0350914
	Number of own children age 6-15	-0.0803781*	0.0209758
	Number of children 16-24 attending school	-0.0118494	0.0331356
	In province of Newfoundland	0.0223841	0.0991462
	In province of P.E.I.	**Dropped**	**Dropped**
	In province of Nova Scotia	0.0968844	0.1019728
	In province of New Brunswick	0.0226228	0.1008827
	In province of Quebec	0.3140238*	0.1047993
	In province of Ontario	0.3275003*	0.0932092
	In province of Manitoba	0.0499093	0.1019962
	In province of Saskatchewan	0.1247453	0.1000367
	In province of Alberta	0.199237*	0.0965336
	In province of British Columbia	0.286456*	0.0976632
Constant		8.785488	0.8912568

* Significant at the 5% level.
** Significant at the 10% level.

Log File

```
. Qextract
getting information about file 378 ...
loading variables from 378 (vas87vol) only (no data yet)...  done

. drop if empfam~=1
(18073 observations deleted)

. drop if f06_q30a>2
(168 observations deleted)

. drop if f06_q35c == .
(1638 observations deleted)

. rename f03q34 sex

. rename f03q33 age

. rename f03q38 edu

. rename f06_q30a english

. rename f06_q30b french

. rename f06_q30c othlang

. rename f06_q35c wage

. recode wage 1=2499.5 2=7499.5 3=12499.5 4=17499.5 5=24999.5 6=34999.5
7=50000 8=50000
(6878 changes made)

. recode age 1=15.5 2=18 3=22 4=29.5 5=39.5 6=49.5 7=59.5 8=67 9=70
(6878 changes made)

. recode edu 1=6 2=6 3=7 4=8 5=9
```

```
(6878 changes made)

. gen exp = age - edu - 6

. gen expsq = exp*exp

. gen lnwage = ln(wage)

.
. gen voluntee = 0

. replace voluntee=1 if  f08_q23a ==1
(1956 real changes made)

. replace voluntee=1 if  f08_q23b ==1
(167 real changes made)

. replace voluntee=1 if  f08_q23c ==1
(110 real changes made)

. replace voluntee=1 if  f08_q23d ==1
(96 real changes made)

. replace voluntee=1 if  f08_q23e ==1
(103 real changes made)

. replace voluntee=1 if  f08_q23f ==1
(120 real changes made)

. replace voluntee=1 if  f08_q23g ==1
(115 real changes made)

. replace voluntee=1 if  f08_q23h ==1
(84 real changes made)

. replace voluntee=1 if  f08_q23i ==1
(57 real changes made)

. replace voluntee=1 if  f08_q23j ==1
(57 real changes made)

. replace voluntee=1 if  f08_q23k ==1
(150 real changes made)

. replace voluntee=1 if  f08_q23l ==1
(59 real changes made)

.
. /*create dummy variables for age , education and provinces*/
. tab age, gen(da)

        age group |      Freq.     Percent        Cum.
------------------+-----------------------------------
             15.5 |        196        2.85        2.85
               18 |        206        3.00        5.84
               22 |        650        9.45       15.30
             29.5 |       2234       32.48       47.78
```

```
           39.5 |       1518       22.07       69.85
           49.5 |        815       11.85       81.70
           59.5 |        843       12.26       93.95
             67 |        199        2.89       96.85
             70 |        217        3.15      100.00
----------------+-----------------------------------
          Total |       6878      100.00

. /*drop if age == 1
> drop if age == 2
> drop if age == 8
> drop if age == 9
> drop da3 da4 da8 da9
> gen young = 0
> replace young = 1 if age == 3
> replace young = 1 if age == 4
> rename da5 mid1
> rename da6 mid2
> rename da7 mid3
> */
. tab province, gen(dp)

 region and province |      Freq.     Percent        Cum.
---------------------+-----------------------------------
        newfoundland |        514        7.47        7.47
prince edward island |        132        1.92        9.39
         nova scotia |        469        6.82       16.21
       new brunswick |        505        7.34       23.55
              quebec |       1172       17.04       40.59
             ontario |       1215       17.67       58.26
            manitoba |        508        7.39       65.64
        saskatchewan |        623        9.06       74.70
             alberta |        955       13.88       88.59
    british columbia |        785       11.41      100.00
---------------------+-----------------------------------
               Total |       6878      100.00

. tab edu, gen(de)

                           education |      Freq.     Percent        Cum.
-------------------------------------+-----------------------------------
                                   6 |       4336       63.04       63.04
                                   7 |        642        9.33       72.38
                                   8 |        993       14.44       86.81
                                   9 |        907       13.19      100.00
-------------------------------------+-----------------------------------
                               Total |       6878      100.00

. gen nonvol = 0

. replace nonvol = 1 if volunt == 0
(3804 real changes made)

.
. /*create interaction terms for male and female*/
. gen male = 0
```

```
. replace male=1 if sex ==1
(3142 real changes made)

. gen mv=male*voluntee

. gen mvexp=male*exp*voluntee

. gen mvexpsq=male*expsq*voluntee

. gen mvhighsh=male*de1*voluntee

. gen mvpostse=male*de2*voluntee

. gen mvdiplo=male*de3*voluntee

. gen mvuniver=male*de4*voluntee

. gen mvfamsiz=male*famsize*voluntee

. gen mvkids1=male* ownkids1*voluntee

. gen mvkids2=male* ownkids2*voluntee

. gen mvkids3=male* ownkids3*voluntee

. gen mvkidsat=male* kidsatsh*voluntee

. gen mvenglis=male*english*voluntee

. gen mvfrench=male*french*voluntee

. gen mvothlan=male*othlang*voluntee

. gen mvprov1=male*dp1*voluntee

. gen mvprov2=male*dp2*voluntee

. gen mvprov3=male*dp3*voluntee

. gen mvprov4=male*dp4*voluntee

. gen mvprov5=male*dp5*voluntee

. gen mvprov6=male*dp6*voluntee

. gen mvprov7=male*dp7*voluntee

. gen mvprov8=male*dp8*voluntee

. gen mvprov9=male*dp9*voluntee

. gen mvprov10=male*dp10*voluntee

.
. gen mnv=male*nonvol

. gen mnvexp=male*exp*nonvol
```

```
. gen mnvexpsq=male*expsq*nonvol

. gen mnvhighs=male*de1*nonvol

. gen mnvposts=male*de2*nonvol

. gen mnvdiplo=male*de3*nonvol

. gen mnvuniv=male*de4*nonvol

. gen mnvfamsi=male*famsize*nonvol

. gen mnvkids1=male* ownkids1*nonvol

. gen mnvkids2=male* ownkids2*nonvol

. gen mnvkids3=male* ownkids3*nonvol

. gen mnvkiats=male* kidsatsh*nonvol

. gen mnvengli=male*english*nonvol

. gen mnvfrenc=male*french*nonvol

. gen mnvothla=male*othlang*nonvol

. gen mnvprov1=male*dp1*nonvol

. gen mnvprov2=male*dp2*nonvol

. gen mnvprov3=male*dp3*nonvol

. gen mnvprov4=male*dp4*nonvol

. gen mnvprov5=male*dp5*nonvol

. gen mnvprov6=male*dp6*nonvol

. gen mnvprov7=male*dp7*nonvol

. gen mnvprov8=male*dp8*nonvol

. gen mnvprov9=male*dp9*nonvol

. gen mnvprov0=male*dp10*nonvol

.
. gen female = 0

. replace female = 1 if sex == 2
(3736 real changes made)

. gen fv = female*voluntee

. gen fvexp = female*exp*voluntee
```

```
. gen fvexpsq = female*expsq*voluntee

. gen fvhighsh = female*de1*voluntee

. gen fvpostse = female*de2*voluntee

. gen fvdiplo = female*de3*voluntee

. gen fvuniver = female*de4*voluntee

. gen fvfamsiz=female*famsize*voluntee

. gen fvkids1 = female* ownkids1*voluntee

. gen fvkids2 = female* ownkids2*voluntee

. gen fvkids3 = female* ownkids3*voluntee

. gen fvkidsat = female* kidsatsh*voluntee

. gen fvenglis=female*english*voluntee

. gen fvfrench=female*french*voluntee

. gen fvothlan=female*othlang*voluntee

. gen fvprov1=female*dp1*voluntee

. gen fvprov2=female*dp2*voluntee

. gen fvprov3=female*dp3*voluntee

. gen fvprov4=female*dp4*voluntee

. gen fvprov5=female*dp5*voluntee

. gen fvprov6=female*dp6*voluntee

. gen fvprov7=female*dp7*voluntee

. gen fvprov8=female*dp8*voluntee

. gen fvprov9=female*dp9*voluntee

. gen fvprov10=female*dp10*voluntee

.
. gen fnv=female *nonvol

. gen fnvexp=female*exp*nonvol

. gen fnvexpsq=female*expsq*nonvol

. gen fnvhighs=female*de1*nonvol

. gen fnvposts=female*de2*nonvol
```

```
. gen fnvdiplo=female*de3*nonvol

. gen fnvuniv=female*de4*nonvol

. gen fnvfamsi=female*famsize*nonvol

. gen fnvkids1=female*ownkids1*nonvol

. gen fnvkids2=female*ownkids2*nonvol

. gen fnvkids3=female*ownkids3*nonvol

. gen fnvkiats=female*kidsatsh*nonvol

. gen fnvengli=female*english*nonvol

. gen fnvfrenc=female*french*nonvol

. gen fnvothla=female*othlang*nonvol

. gen fnvprov1=female*dp1*nonvol

. gen fnvprov2=female*dp2*nonvol

. gen fnvprov3=female*dp3*nonvol

. gen fnvprov4=female*dp4*nonvol

. gen fnvprov5=female*dp5*nonvol

. gen fnvprov6=female*dp6*nonvol

. gen fnvprov7=female*dp7*nonvol

. gen fnvprov8=female*dp8*nonvol

. gen fnvprov9=female*dp9*nonvol

. gen fnvprov0=female*dp10*nonvol

.
. set matsiz 120

.
. /*regression*/
. /*pooled regression for 4 groups*/
. regress lnwage mnvexp mnvexpsq mnvhighs mnvposts mnvdiplo mnvuniv
mnvengli mnvfrenc mnvothla mnvfamsi mnvkids1 mnvkids2 mnvkids3 mnvkiats
mnvprov1 mnvprov2 mnvprov3 mnvprov4 mnvprov5 mnvprov6 mnvprov7 mnvprov8
mnvprov9 mnvprov0 mv mvexp mvexpsq mvhighsh mvpostse mvdiplo mvuniver
mvenglis mvfrench mvothlan mvfamsiz mvkids1 mvkids2 mvkids3 mvkidsat
mvprov1 mvprov2 mvprov3 mvprov4 mvprov5 mvprov6 mvprov7 mvprov8 mvprov9
mvprov10 fv fvexp fvexpsq fvhighsh fvpostse fvdiplo fvuniver fvenglis
fvfrench fvothlan fvfamsiz fvkids1 fvkids2 fvkids3 fvkidsat fvprov1
fvprov2 fvprov3 fvprov4 fvprov5 fvprov6 fvprov7 fvprov8 fvprov9
fvprov10 fnv fnvexp fnvexpsq fnvhighs fnvposts fnvdiplo fnvuniv
fnvengli fnvfrenc fnvothla fnvfamsi fnvkids1 fnvkids2 fnvkids3 fnvkiats
```

```
fnvprov1 fnvprov2 fnvprov3 fnvprov4 fnvprov5 fnvprov6 fnvprov7 fnvprov8
fnvprov9 fnvprov0

  Source |       SS       df       MS              Number of obs =    6878
---------+------------------------------           F( 91,  6786) =   14.48
   Model |  389.999929    91  4.28571351           Prob > F      =  0.0000
Residual |  2008.88559  6786  .296033834           R-squared     =  0.1626
---------+------------------------------           Adj R-squared =  0.1513
   Total |  2398.88552  6877  .348827326           Root MSE      =  .54409

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  mnvexp |   .0172714   .0036864      4.685   0.000        .010045    .0244978
mnvexpsq |  -.0002314   .0000602     -3.842   0.000      -.0003494   -.0001133
mnvhighs |  -.7930283   1.028375     -0.771   0.441      -2.808966    1.222909
mnvposts |  -.6085904   1.029336     -0.591   0.554      -2.626411     1.40923
mnvdiplo |  -.5724835   1.029609     -0.556   0.578       -2.59084    1.445873
 mnvuniv |  -.3827609   1.029367     -0.372   0.710      -2.400643    1.635121
mnvengli |    .201768   .1063885      1.897   0.058      -.0067868    .4103228
mnvfrenc |   .1830338   .1165685      1.570   0.116      -.0454771    .4115447
mnvothla |   .3829137   .0996634      3.842   0.000       .1875422    .5782852
mnvfamsi |   .0906728   .0170922      5.305   0.000       .0571666    .1241789
mnvkids1 |  -.0420437   .0361049     -1.164   0.244      -.1128208    .0287333
mnvkids2 |  -.0296613   .0376273     -0.788   0.431      -.1034225    .0440999
mnvkids3 |  -.0015329   .0208106     -0.074   0.941      -.0423283    .0392625
mnvkiats |   .0069882   .0347504      0.201   0.841      -.0611336      .07511
mnvprov1 |  -.0733663   .0967958     -0.758   0.449      -.2631163    .1163838
mnvprov2 |  (dropped)
mnvprov3 |   .1294816   .0993639      1.303   0.193      -.0653028     .324266
mnvprov4 |   .0218783   .0989384      0.221   0.825      -.1720719    .2158286
mnvprov5 |   .2073272   .1063007      1.950   0.051      -.0010556      .41571
mnvprov6 |   .3572267   .0916138      3.899   0.000        .177635    .5368184
mnvprov7 |   .2103777   .0989447      2.126   0.034        .016415    .4043404
mnvprov8 |   .2490585   .0978077      2.546   0.011       .0573247    .4407923
mnvprov9 |   .2100216   .0941095      2.232   0.026       .0255376    .3945057
mnvprov0 |   .3197017   .0938421      3.407   0.001       .1357419    .5036616
      mv |  (dropped)
   mvexp |   .0196792   .0051142      3.848   0.000       .0096539    .0297046
 mvexpsq |  -.0002673   .0000834     -3.205   0.001      -.0004308   -.0001038
mvhighsh |   .0705353    .052935      1.332   0.183      -.0332338    .1743045
mvpostse |  (dropped)
 mvdiplo |   .1688454   .0612862      2.755   0.006       .0487052    .2889856
 mvuniver |   .3627873   .0566985      6.399   0.000       .2516405     .473934
mvenglis |  -.0453145   .1882461     -0.241   0.810      -.4143358    .3237068
mvfrench |   .0115812   .1923823      0.060   0.952      -.3655484    .3887109
mvothlan |   .2776568   .1809811      1.534   0.125      -.0771229    .6324365
mvfamsiz |   .0607735   .0239789      2.534   0.011       .0137673    .1077796
 mvkids1 |  -.0172733   .0429104     -0.403   0.687      -.1013912    .0668446
 mvkids2 |  -.0146204   .0403437     -0.362   0.717      -.0937066    .0644658
 mvkids3 |   .0456517   .0246749      1.850   0.064      -.0027189    .0940223
mvkidsat |   .0209774    .045667      0.459   0.646      -.0685443    .1104991
 mvprov1 |   .0998154   .1506925      0.662   0.508      -.1955893      .39522
 mvprov2 |  (dropped)
 mvprov3 |   .2515977   .1488522      1.690   0.091      -.0401993    .5433948
 mvprov4 |   .3173755   .1513647      2.097   0.036       .0206533    .6140977
 mvprov5 |   .3388386    .160607      2.110   0.035       .0239985    .6536786
 mvprov6 |    .468882   .1420137      3.302   0.001       .1904906    .7472735
 mvprov7 |     .38559    .146004      2.641   0.008       .0993764    .6718036
 mvprov8 |   .2389253   .1439889      1.659   0.097      -.0433381    .5211887
 mvprov9 |   .3273247   .1412649      2.317   0.021       .0504013    .6042481
mvprov10 |   .4435922   .1427984      3.106   0.002       .1636626    .7235219
      fv |  -.0380826     1.0884     -0.035   0.972      -2.171689    2.095524
```

```
   fvexp |   .0313845   .0045586      6.885   0.000       .0224483    .0403208
 fvexpsq |  -.0004589   .0000765     -5.999   0.000      -.0006089    -.000309
fvhighsh |  -.3343738   .0380135     -8.796   0.000      -.4088923   -.2598553
fvpostse |  -.2598206   .0504793     -5.147   0.000      -.3587759   -.1608653
 fvdiplo |  -.1906531    .043263     -4.407   0.000      -.2754623    -.105844
fvuniver |  (dropped)
fvenglis |   .1337197   .1356852      0.986   0.324      -.1322657    .3997051
fvfrench |   .1126686   .1350572      0.834   0.404      -.1520858    .3774231
fvothlan |   .1936609   .1374526      1.409   0.159      -.0757892    .4631111
fvfamsiz |   .1810094    .020187      8.967   0.000       .1414365    .2205824
 fvkids1 |  -.0212367   .0368781     -0.576   0.565      -.0935293    .0510558
 fvkids2 |  -.0528429   .0327745     -1.612   0.107      -.1170912    .0114055
 fvkids3 |   -.040641   .0208445     -1.950   0.051      -.0815027    .0002208
fvkidsat |  -.0764488    .035714     -2.141   0.032      -.1464595   -.0064382
 fvprov1 |  -.1168016   .1015597     -1.150   0.250      -.3158905    .0822873
 fvprov2 |  (dropped)
 fvprov3 |   .0732168    .099495      0.736   0.462      -.1218246    .2682582
 fvprov4 |   .0154526   .1030586      0.150   0.881      -.1865746    .2174798
 fvprov5 |   .0572507    .113699      0.504   0.615      -.1656349    .2801363
 fvprov6 |   .2140672   .0941572      2.274   0.023       .0294896    .3986448
 fvprov7 |   .0416957   .0994003      0.419   0.675      -.1531601    .2365515
 fvprov8 |   .0069937   .0967014      0.072   0.942      -.1825713    .1965587
 fvprov9 |   .0770842   .0939189      0.821   0.412      -.1070262    .2611946
fvprov10 |    .069854   .0960784      0.727   0.467      -.1184898    .2581979
     fnv |  -.1263828   1.035858     -0.122   0.903      -2.156989    1.904223
  fnvexp |   .0172368   .0036432      4.731   0.000        .010095    .0243787
fnvexpsq |  -.0002349   .0000593     -3.959   0.000      -.0003512   -.0001186
fnvhighs |  -.0965939   .0466376     -2.071   0.038      -.1880181   -.0051697
fnvposts |  (dropped)
fnvdiplo |   .1145373    .055643      2.058   0.040       .0054595    .2236151
 fnvuniv |   .2681002    .064838      4.135   0.000       .1409975     .395203
fnvengli |   .0196446   .1105807      0.178   0.859      -.1971283    .2364176
fnvfrenc |   .1593518   .1152824      1.382   0.167      -.0666379    .3853414
fnvothla |   .1144466   .1059502      1.080   0.280       -.093249    .3221422
fnvfamsi |   .1763353   .0185183      9.522   0.000       .1400337     .212637
fnvkids1 |  -.0548205   .0346519     -1.582   0.114      -.1227491    .0131081
fnvkids2 |  -.0783378   .0350914     -2.232   0.026      -.1471279   -.0095478
fnvkids3 |  -.0803781   .0209758     -3.832   0.000      -.1214973   -.0392589
fnvkiats |  -.0118494   .0331356     -0.358   0.721      -.0768055    .0531068
fnvprov1 |   .0223841   .0991462      0.226   0.821      -.1719736    .2167417
fnvprov2 |  (dropped)
fnvprov3 |   .0968844   .1019728      0.950   0.342      -.1030144    .2967831
fnvprov4 |   .0226228   .1008827      0.224   0.823      -.1751391    .2203846
fnvprov5 |   .3140238   .1047993      2.996   0.003       .1085844    .5194632
fnvprov6 |   .3275003   .0932092      3.514   0.000       .1447811    .5102195
fnvprov7 |   .0499093   .1019962      0.489   0.625      -.1500353    .2498538
fnvprov8 |   .1247453   .1000367      1.247   0.212       -.071358    .3208486
fnvprov9 |    .199237   .0965336      2.064   0.039       .0100008    .3884732
fnvprov0 |    .286458   .0976632      2.933   0.003       .0950075    .4779085
   _cons |   8.785488   .8912568      9.857   0.000       7.038346    10.53263
----------------------------------------------------------------------------

.
. /*hypothesis tests*/


. test mv mvexp mvexpsq mvhighsh mvpostse mvdiplo mvuniver mvenglis
mvfrench mvothlan mvfamsiz mvkids1 mvkids2 mvkids3 mvkidsat mvprov1
mvprov2 mvprov3 mvprov4 mvprov5 mvprov6 mvprov7 mvprov8 mvprov9
mvprov10 fv fvexp fvexpsq fvhighsh fvpostse fvdiplo fvuniver fvenglis
fvfrench fvothlan fvfamsiz fvkids1 fvkids2 fvkids3 fvkidsat fvprov1
fvprov2 fvprov3 fvprov4 fvprov5 fvprov6 fvprov7 fvprov8 fvprov9
```

```
fvprov10 fnv fnvexp fnvhighs fnvposts fnvdiplo fnvuniv fnvengli
fnvfrenc fnvothla fnvfamsi fnvkids1 fnvkids2 fnvkids3 fnvkiats fnvprov1
fnvprov2 fnvprov3 fnvprov4 fnvprov5 fnvprov6 fnvprov7 fnvprov8 fnvprov9
fnvprov0

 ( 1)  mv = 0.0
 ( 2)  mvexp = 0.0
 ( 3)  mvexpsq = 0.0
 ( 4)  mvhighsh = 0.0
 ( 5)  mvpostse = 0.0
 ( 6)  mvdiplo = 0.0
 ( 7)  mvuniver = 0.0
 ( 8)  mvenglis = 0.0
 ( 9)  mvfrench = 0.0
 (10)  mvothlan = 0.0
 (11)  mvfamsiz = 0.0
 (12)  mvkids1 = 0.0
 (13)  mvkids2 = 0.0
 (14)  mvkids3 = 0.0
 (15)  mvkidsat = 0.0
 (16)  mvprov1 = 0.0
 (17)  mvprov2 = 0.0
 (18)  mvprov3 = 0.0
 (19)  mvprov4 = 0.0
 (20)  mvprov5 = 0.0
 (21)  mvprov6 = 0.0
 (22)  mvprov7 = 0.0
 (23)  mvprov8 = 0.0
 (24)  mvprov9 = 0.0
 (25)  mvprov10 = 0.0
 (26)  fv = 0.0
 (27)  fvexp = 0.0
 (28)  fvexpsq = 0.0
 (29)  fvhighsh = 0.0
 (30)  fvpostse = 0.0
 (31)  fvdiplo = 0.0
 (32)  fvuniver = 0.0
 (33)  fvenglis = 0.0
 (34)  fvfrench = 0.0
 (35)  fvothlan = 0.0
 (36)  fvfamsiz = 0.0
 (37)  fvkids1 = 0.0
 (38)  fvkids2 = 0.0
 (39)  fvkids3 = 0.0
 (40)  fvkidsat = 0.0
 (41)  fvprov1 = 0.0
 (42)  fvprov2 = 0.0
 (43)  fvprov3 = 0.0
 (44)  fvprov4 = 0.0
 (45)  fvprov5 = 0.0
 (46)  fvprov6 = 0.0
 (47)  fvprov7 = 0.0
 (48)  fvprov8 = 0.0
 (49)  fvprov9 = 0.0
 (50)  fvprov10 = 0.0
 (51)  fnv = 0.0
 (52)  fnvexp = 0.0
```

```
 (53)  fnvhighs = 0.0
 (54)  fnvposts = 0.0
 (55)  fnvdiplo = 0.0
 (56)  fnvuniv = 0.0
 (57)  fnvengli = 0.0
 (58)  fnvfrenc = 0.0
 (59)  fnvothla = 0.0
 (60)  fnvfamsi = 0.0
 (61)  fnvkids1 = 0.0
 (62)  fnvkids2 = 0.0
 (63)  fnvkids3 = 0.0
 (64)  fnvkiats = 0.0
 (65)  fnvprov1 = 0.0
 (66)  fnvprov2 = 0.0
 (67)  fnvprov3 = 0.0
 (68)  fnvprov4 = 0.0
 (69)  fnvprov5 = 0.0
 (70)  fnvprov6 = 0.0
 (71)  fnvprov7 = 0.0
 (72)  fnvprov8 = 0.0
 (73)  fnvprov9 = 0.0
 (74)  fnvprov0 = 0.0
       Constraint 1 dropped
       Constraint 5 dropped
       Constraint 17 dropped
       Constraint 32 dropped
       Constraint 42 dropped
       Constraint 54 dropped
       Constraint 66 dropped

       F( 67,  6786) =   14.13
            Prob > F =    0.0000



. test mv mvexp mvexpsq mvhighsh mvpostse mvdiplo mvuniver mvenglis
mvfrench mvothlan mvfamsiz mvkids1 mvkids2 mvkids3 mvkidsat mvprov1
mvprov2 mvprov3 mvprov4 mvprov5 mvprov6 mvprov7 mvprov8 mvprov9
mvprov10

 ( 1)  mv = 0.0
 ( 2)  mvexp = 0.0
 ( 3)  mvexpsq = 0.0
 ( 4)  mvhighsh = 0.0
 ( 5)  mvpostse = 0.0
 ( 6)  mvdiplo = 0.0
 ( 7)  mvuniver = 0.0
 ( 8)  mvenglis = 0.0
 ( 9)  mvfrench = 0.0
 (10)  mvothlan = 0.0
 (11)  mvfamsiz = 0.0
 (12)  mvkids1 = 0.0
 (13)  mvkids2 = 0.0
 (14)  mvkids3 = 0.0
 (15)  mvkidsat = 0.0
 (16)  mvprov1 = 0.0
 (17)  mvprov2 = 0.0
```

```
 (18)  mvprov3 = 0.0
 (19)  mvprov4 = 0.0
 (20)  mvprov5 = 0.0
 (21)  mvprov6 = 0.0
 (22)  mvprov7 = 0.0
 (23)  mvprov8 = 0.0
 (24)  mvprov9 = 0.0
 (25)  mvprov10 = 0.0
       Constraint 1 dropped
       Constraint 5 dropped
       Constraint 17 dropped

       F( 22,  6786) =    8.68
            Prob > F =    0.0000



. test fv fvexp fvexpsq fvhighsh fvpostse fvdiplo fvuniver fvenglis
fvfrench fvothlan fvfamsiz fvkids1 fvkids2 fvkids3 fvkidsat fvprov1
fvprov2 fvprov3 fvprov4 fvprov5 fvprov6 fvprov7 fvprov8 fvprov9
fvprov10

 ( 1)  fv = 0.0
 ( 2)  fvexp = 0.0
 ( 3)  fvexpsq = 0.0
 ( 4)  fvhighsh = 0.0
 ( 5)  fvpostse = 0.0
 ( 6)  fvdiplo = 0.0
 ( 7)  fvuniver = 0.0
 ( 8)  fvenglis = 0.0
 ( 9)  fvfrench = 0.0
 (10)  fvothlan = 0.0
 (11)  fvfamsiz = 0.0
 (12)  fvkids1 = 0.0
 (13)  fvkids2 = 0.0
 (14)  fvkids3 = 0.0
 (15)  fvkidsat = 0.0
 (16)  fvprov1 = 0.0
 (17)  fvprov2 = 0.0
 (18)  fvprov3 = 0.0
 (19)  fvprov4 = 0.0
 (20)  fvprov5 = 0.0
 (21)  fvprov6 = 0.0
 (22)  fvprov7 = 0.0
 (23)  fvprov8 = 0.0
 (24)  fvprov9 = 0.0
 (25)  fvprov10 = 0.0
       Constraint 7 dropped
       Constraint 17 dropped

       F( 23,  6786) =   14.27
            Prob > F =    0.0000
```

```
. test fnv fnvexp fnvhighs fnvposts fnvdiplo fnvuniv fnvengli fnvfrenc
fnvothla fnvfamsi fnvkids1 fnvkids2 fnvkids3 fnvkiats fnvprov1 fnvprov2
fnvprov3 fnvprov4 fnvprov5 fnvprov6 fnvprov7 fnvprov8 fnvprov9 fnvprov0

 ( 1)  fnv = 0.0
 ( 2)  fnvexp = 0.0
 ( 3)  fnvhighs = 0.0
 ( 4)  fnvposts = 0.0
 ( 5)  fnvdiplo = 0.0
 ( 6)  fnvuniv = 0.0
 ( 7)  fnvengli = 0.0
 ( 8)  fnvfrenc = 0.0
 ( 9)  fnvothla = 0.0
 (10)  fnvfamsi = 0.0
 (11)  fnvkids1 = 0.0
 (12)  fnvkids2 = 0.0
 (13)  fnvkids3 = 0.0
 (14)  fnvkiats = 0.0
 (15)  fnvprov1 = 0.0
 (16)  fnvprov2 = 0.0
 (17)  fnvprov3 = 0.0
 (18)  fnvprov4 = 0.0
 (19)  fnvprov5 = 0.0
 (20)  fnvprov6 = 0.0
 (21)  fnvprov7 = 0.0
 (22)  fnvprov8 = 0.0
 (23)  fnvprov9 = 0.0
 (24)  fnvprov0 = 0.0
       Constraint 4 dropped
       Constraint 16 dropped

       F( 22,  6786) =   12.15
            Prob > F =    0.0000


.
. /*prediction*/
.
. predict yhat
(option xb assumed; fitted values)

. summ lnwage yhat

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |    6878    10.11653   .5906161   7.823846   10.81978
    yhat |    6878    10.11653   .2381402   9.283969   10.98948

.
. predict uhat
(option xb assumed; fitted values)

. summ lnwage yhat uhat

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
  lnwage |    6878    10.11653   .5906161   7.823846   10.81978
```

```
    yhat |      6878    10.11653    .2381402   9.283969   10.98948
    uhat |      6878    10.11653    .2381402   9.283969   10.98948

.
. summ wage province sex female mv mnv fv fnv age de1 de2 de3 de4
famsize ownkids1 ownkids2 ownkids3 kidsatsh english french othlang exp
expsq dp1 dp2 dp3 dp4 dp5 dp6 dp7 dp8 dp9 dp10

Variable |       Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    wage |      6878     28616.2    13680.14     2499.5      50000
province |      6878     34.0567    16.13448         10         59
     sex |      6878    1.543181    .4981681          1          2
  female |      6878    .5431812    .4981681          0          1
      mv |      6878     .182902    .3866143          0          1
     mnv |      6878    .2739168    .4459992          0          1
      fv |      6878    .2640302    .4408475          0          1
     fnv |      6878    .2791509    .4486145          0          1
     age |      6878    38.66444    14.25192       15.5         70
     de1 |      6878    .6304158    .4827272          0          1
     de2 |      6878    .0933411     .290931          0          1
     de3 |      6878    .1443734    .3514935          0          1
     de4 |      6878    .1318697    .3383737          0          1
 famsize |      6878     2.69686    1.162583          1          4
ownkids1 |      6878    .1788311    .4420855          0          3
ownkids2 |      6878    .1738878    .4331122          0          3
ownkids3 |      6878    .5231172    .8818385          0          5
kidsatsh |      6878    .1420471    .4219201          0          3
 english |      6878    1.216924    .4121801          1          2
  french |      6878    1.805757    .3956452          1          2
 othlang |      6878     1.96467    .1846261          1          2
     exp |      6878    25.88674    14.35825        3.5         58
   expsq |      6878    876.2526    887.6157      12.25       3364
     dp1 |      6878     .074731     .262976          0          1
     dp2 |      6878    .0191916     .137208          0          1
     dp3 |      6878    .0681884    .2520873          0          1
     dp4 |      6878    .0734225    .2608477          0          1
     dp5 |      6878    .1703984    .3760097          0          1
     dp6 |      6878    .1766502    .3814001          0          1
     dp7 |      6878    .0738587    .2615598          0          1
     dp8 |      6878    .0905787    .2870299          0          1
     dp9 |      6878    .1388485    .3458135          0          1
    dp10 |      6878     .114132    .3179947          0          1


. count if sex == 1
 3142

. count if sex == 2
 3736

. count if voluntee == 1
 3074

. count if voluntee == 0
 3804
```

```
. count if sex == 1 & voluntee == 1
 1258

. count if sex == 2 & voluntee == 1
 1816

. count if sex == 1 & voluntee == 0
 1884

. count if sex == 2 & voluntee == 0
 1920

. count if de1 ==1
 4336

. count if de2 ==1
  642

. count if de3 ==1
  993

. count if de4 ==1
  907

. count if de1==1 & voluntee ==1
 1572

. count if de2==1 & voluntee ==1
  330

. count if de3==1 & voluntee ==1
  547

. count if de4==1 & voluntee ==1
  625

. count if de1==1 & voluntee ==0
 2764

. count if de2==1 & voluntee ==0
  312

. count if de3==1 & voluntee ==0
  446

. count if de4==1 & voluntee ==0
  282
```