ECON 21110 - Applied Microeconometrics - Assignment 1

Jack Ogle

Problem 1

(a)

We then estimate the multiple regression for the following Models. See Table 1 for Regression Results.

$$(M1) \qquad lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + U_i$$

$$(M2) \qquad lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + U_i$$

Explanation of (M1) regression results

When $educ_i$ and KWW_i are constant the mean hourly wage of men ages 14-24 is 535.1 cents. Holding KWW_i constant every year increase of $educ_i$ there is a 2.1% increase in hourly wage in cents on average. Holding $educ_i$ constant every unit increase of KWW_i there is a 1.9% increase in hourly wage in cents on average.

Explanation of (M2) regression results

Table 1: Regression Results

	Dependent variable:	
	lwage	
	M1	M2
	(1)	(2)
educ	0.021***	
	(0.003)	
HS		0.125***
		(0.022)
College		0.195***
		(0.026)
KWW	0.019***	0.019***
	(0.001)	(0.001)
Constant	5.351***	5.514***
	(0.039)	(0.031)
Observations	2,963	2,963
R^2	0.196	0.199
Adjusted R^2	0.195	0.199
Residual Std. Error	0.396 (df = 2960)	0.395 (df = 2959)
F Statistic	360.893*** (df = 2; 2960)	245.547*** (df = 3; 2959)
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses	

When receiving both a high school or college education and KWW are constant the mean hourly wage of men ages 14-24 is 553.3 cents. Holding KWW_i constant obtaining a high school education results in a 3.6% increase in hourly wage in cents on average. Holding KWW_i constant obtaining a college education results in a 11.6% increase in hourly wage in cents on average. Holding both high school and college education constant every unit increase of KWW_i there is a 2.1% increase in hourly wage in cents on average.

(b)

The OLS estimator $\hat{\beta}_1$ is biased because it fails to satisfy one of four necessary assumptions. The zero conditional mean assumption does not hold.

$$\text{Zero Conditional Mean Assumption}$$
$$E[U_i|educ_i, KWW_i] = 0$$

There are variables in the error term that are correlated with independent variables. For example, if a parent of a child is educated, then they would be more likely to have their child educated. If they are educated then they would be more likely to have the required tuition funds because they have a higher paying job. Additionally they would be more likely to see the benefit of education as they benefited from the connections and a higher paying job. Highly educated parents are likely to produce highly educated offspring because they have the means and wherewithal. This means that parents education a variable in U_i is correlated with our independent variable $educ_i$.

(c)

In order for the Zero Conditional Mean Assumption to hold the variables in U_i cannot be correlated with the independent variables $educ_i$ and KWW_i. To minimize the correlation between KWW_i and any variables in the error term we would want to measure KWW_i as early in the subjects as possible. This would minimize the correlation because the child would not be exposed to varying levels of education that might give some students and advantage over their peers.

The study measured the participants KWW_i as soon as the study begins in 1966. This would mean that the youngest men would have been age 14 and the oldest would have been age 24.

(d)

$$(M3) \qquad lwage_i = \beta_0 + \beta_1 educ_i + \beta_2 HS_i + \beta_3 College_i + \beta_4 KWW_i + \beta_5 HS_i * educ_i + \beta_6 College_i * educ_i + U_i$$

(M2) is preferred to (M1) according to the beta coefficient values. If a young man has received a College education then when education is increased by one year there is a compensatory 4.6299% $[100(\beta_1 + \beta_6)]$ increase in hourly wage. Compared to if a young man has received a High School Education then when education is increased by one year there is a 1.27% $[100(\beta_1 + \beta_5)]$ decrease in the hourly wage. There is a big difference between receiving a college education and receiving only a High School education. This is because $educ_i$ is a misleading explanatory value. It is misleading because their is such a stark difference between the effects of college and high school on wages. So we should exclude the variable $educ_i$ and use (M2). See Table 2 for regression results.

(e)

For (M1) we see that the effect of education on wage per hour is smaller at age 24 compared to the effect when considering all ages $hat\beta_1$. However when the men reach ages 28 and 32 we see that beta estimator for those ages are larger than the beta estimator for All ages. This is likely because the effect of education is lagging and it takes a few years for the value of education to pay off and increase wages. See table 3 for the regression results.

We see a similar trend in (M2) the effect of education on wage per hour is less at age 24 compared to the effect when considering all ages. Then men at age 28 and 32 see an increase in wage dependent on income. Both β_1 and β_2 increase at these ages, so both High School and College education increases income for these ages. See table 4 for the regression results.

(f)

When experience is held constant we see that $hat\beta_1$ is less than β_1. This can be explained by understanding what it means to control for experience. Controlling for experience means that the differences of experiences of individual subjects has no effect on the wage of the subjects. Thus, the effect of getting an education has an larger effect on wage. Keeping experience constant, more educated young men have higher returns to education compared to less educated young men. See table 5 for regression results.

In (M2) we see that the High School education coefficient in this regression is less than the value we found in (a) at an experience level of 4. And the same High School education coefficient was greater at experience levels of 8 and 12 compared to our original

(a) value. For the college education coefficient we see that there is no value for experience level 4. And for experience levels 8 and 12 they are greater than the coefficient we originally found in (a). See table 6 for regression results.

The effects of education are increasing in model two for the same reasons that the increased in model one. It takes a few years for individuals experience to pay off and effect wages. Additionally we see that at a certain age and as experience increases less education matters less with respect to increasing wages.

(g)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 Exper_i + U_i$$

Null Hypothesis: The coefficients for HS_i and $College_i$ from (M1) are equal to the coefficients in this model.

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 Exper_i + U_i$$

Null Hypothesis: The coefficients for HS_i and $College_i$ respectively from (M2) are equal to the coefficients in this model.

In our regression results found on table 7 we can reject our Null Hypothesis because both $\beta_1 HS_i$ and $\beta_1 HS_i$ are greater than the coefficient values we found in (a). When controlling for experience there is a greater return to high school and college education.

These models are more representative of the causal relationship between educational degrees both high school and college and wages. Because education and experience are correlated the more education and individual obtains the less experience they can have.

(h)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 Exper_i + \beta_5 Exper_i * educ_i + U_i$$

Null Hypothesis is that β_5 is 0.

We can reject this hypothesis if we look at the regression results and see that it is not 0. It is 0.003.

(i)

(j)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 Black_i + U_i$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 Black_i + U_i$$

Null Hypothesis is that β_4 is equal to 0. For Both Models.

We can clearly see that β_4 is not equal to 0 in both models. Therefore we can reject our null hypothesis.

It is unlikely that there is a causal relationship between the $\beta_4 Black_i$ and wages. The zero conditional mean assumption is violated. There is a variable in the error term that is correlated with the independent variable of being black. Black people might be victims of discrimination and not relieve the same education or job opportunities compared to people of other races. This might be correlated to lower wages.

(k)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 Black_i + \beta_5 Black_i * educ_i + U_i$$

For (M1) the Null Hypothesis is that the β_5 is 0. Which we can reject if we look at the regression results. Therefore education and its returns to wages is variable depending on what race a individual is.

Often and unfortunately students and individuals of different races have different access to education. Specifically, black students don't always have access to good schools and education. Therefore the effect that education has on their wages is correlated with their race.

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 Black_i + \beta_5 Black_i * HS_i + \beta_5 Black_i * College_i + U_i$$

For (M2) the Null Hypothesis is that β_5 is 0 and β_6 is 0

Because the estimator β_5 is not statistically significant we cannot fully reject the Null hypothesis. However, we can reject part of it as β_6 is not 0 and it is statistically significant. This means that we cannot conclude anything about how high school education might effect the returns to wage across different races.

However, because we reject the second half of the hypothesis. We can conclude that a college education returns to wages does differ by race.

(l)

(m)

Problem 2

(a)

If we left out X2 we would be violating MLR.4. The zero conditional mean assumption would be violated because it would then be placed in the error term. Thus we would know that there is a variable that is correlated with the independent variable income. If this key assumption does not hold then there can be no causal effect of X1 on Y.

(b)

MLR.1-4 need to hold to for $\hat{\beta}_1$ to be an unbiased and consistent estimator. MLR.2 holds because we are pulling the data from a random sample of 7000 households. MLR.3 holds because their is only one independent variable so there is no other variable that could be colinear. MLR.4 is not satisfied because there are variables in the error term that are correlated with the independent variables. For example the parents income is not included as an independent variable and likely correlated with income. I don't find these conditions credible because MLR.4 does not hold.

(c)

I would want to collect various other data that are related to income and potentially correlated with income. I would want to do this in order that the MLR.4 the zero conditional mean assumption holds. I would collect income of the parents, spouse, and dependents.

Problem 3

- Standard of Living and Human Capital comparison among nations
- A Contribution to the Empirics of Economic Growth Mankiw, Romer, Weil
- Data comes from Real National Accounts which is publicly available
- Uses OLS

Table 2: Regression Results

	Dependent variable:		
	lwage		
	M1	M2	M3
	(1)	(2)	(3)
educ	0.021***		0.025***
	(0.003)		(0.004)
HS		0.036**	0.514***
		(0.016)	(0.129)
College		0.116***	−2.162***
		(0.026)	(0.745)
KWW	0.019***	0.021***	0.018***
	(0.001)	(0.001)	(0.001)
educ:HS			−0.037***
			(0.010)
educ:College			0.124***
			(0.042)
Constant	5.351***	5.533***	5.288***
	(0.039)	(0.031)	(0.049)
Observations	2,963	2,963	2,963
R^2	0.196	0.189	0.204
Adjusted R^2	0.195	0.188	0.202
Residual Std. Error	0.396 (df = 2960)	0.398 (df = 2959)	0.394 (df = 2956)
F Statistic	360.893*** (df = 2; 2960)	229.850*** (df = 3; 2959)	125.991*** (df = 6; 2956)
Note:	*p<0.1; **p<0.05; ***p<0.01		
	Standard errors in parentheses		

Table 3: M1 Regression Results by Age

	Dependent variable:			
	lwage			
	All Ages	Age = 24	Age = 28	Age = 32
	(1)	(2)	(3)	(4)
educ	0.021***	−0.0005	0.036***	0.043***
	(0.003)	(0.010)	(0.010)	(0.011)
KWW	0.019***	0.010***	0.011***	0.015***
	(0.001)	(0.003)	(0.003)	(0.004)
Constant	5.351***	5.791***	5.455***	5.319***
	(0.039)	(0.122)	(0.137)	(0.134)
Observations	2,963	388	309	210
R^2	0.196	0.037	0.122	0.269
Adjusted R^2	0.195	0.032	0.116	0.262
Residual Std. Error	0.396 (df = 2960)	0.400 (df = 385)	0.380 (df = 306)	0.380 (df = 207)
F Statistic	360.893*** (df = 2; 2960)	7.372*** (df = 2; 385)	21.236*** (df = 2; 306)	38.071*** (df = 2; 207)
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses			

Table 4: M2 Regression Results by Age

	Dependent variable:			
	lwage			
	All Ages	Age = 24	Age = 28	Age = 32
	(1)	(2)	(3)	(4)
HS	0.125***	0.031	0.248***	0.412***
	(0.022)	(0.063)	(0.082)	(0.080)
College	0.195***	−0.007	0.389***	0.479***
	(0.026)	(0.076)	(0.090)	(0.095)
KWW	0.019***	0.010***	0.010***	0.013***
	(0.001)	(0.003)	(0.003)	(0.004)
Constant	5.514***	5.766***	5.702***	5.608***
	(0.031)	(0.083)	(0.116)	(0.123)
Observations	2,963	388	309	210
R^2	0.199	0.039	0.142	0.313
Adjusted R^2	0.199	0.031	0.133	0.303
Residual Std. Error	0.395 (df = 2959)	0.400 (df = 384)	0.376 (df = 305)	0.369 (df = 206)
F Statistic	245.547*** (df = 3; 2959)	5.148*** (df = 3; 384)	16.765*** (df = 3; 305)	31.232*** (df = 3; 206)
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses			

Table 5: M1 Regression Results by Experience

	Dependent variable:			
	lwage			
	All Levels of Experience	Experience = 4	Experience = 8	Experience = 12
	(1)	(2)	(3)	(4)
educ	0.021***	0.074***	0.052***	0.108***
	(0.003)	(0.026)	(0.012)	(0.019)
KWW	0.019***	0.018***	0.014***	0.009*
	(0.001)	(0.005)	(0.003)	(0.005)
Constant	5.351***	4.498***	5.063***	4.739***
	(0.039)	(0.369)	(0.121)	(0.162)
Observations	2,963	163	308	132
R^2	0.196	0.208	0.283	0.452
Adjusted R^2	0.195	0.198	0.278	0.444
Residual Std. Error	0.396 (df = 2960)	0.384 (df = 160)	0.351 (df = 305)	0.366 (df = 129)
F Statistic	360.893*** (df = 2; 2960)	21.024*** (df = 2; 160)	60.154*** (df = 2; 305)	53.227*** (df = 2; 129)
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Standard errors in parentheses			

Table 6: M2 Regression Results by Experience

	Dependent variable:			
	lwage			
	All Levels of Experience	Experience = 4	Experience = 8	Experience = 12
	(1)	(2)	(3)	(4)
educ	0.021***			
	(0.003)			
College		0.117*	0.362***	0.592***
		(0.069)	(0.088)	(0.136)
HS			0.160**	0.319***
			(0.067)	(0.088)
KWW	0.019***	0.021***	0.015***	0.016***
	(0.001)	(0.005)	(0.003)	(0.004)
Constant	5.351***	5.495***	5.540***	5.551***
	(0.039)	(0.149)	(0.092)	(0.125)
Observations	2,963	163	308	132
R^2	0.196	0.183	0.278	0.412
Adjusted R^2	0.195	0.172	0.271	0.398
Residual Std. Error	0.396 (df = 2960)	0.391 (df = 160)	0.353 (df = 304)	0.381 (df = 128)
F Statistic	360.893*** (df = 2; 2960)	17.861*** (df = 2; 160)	39.020*** (df = 3; 304)	29.861*** (df = 3; 128)
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Standard errors in parentheses			

Table 7: Regression Results for (g)

	Dependent variable:	
	lwage	
	M1	M2
	(1)	(2)
educ	0.056***	
	(0.004)	
HS		0.237***
		(0.024)
College		0.399***
		(0.033)
KWW	0.014***	0.015***
	(0.001)	(0.001)
exper	0.027***	0.022***
	(0.003)	(0.002)
Constant	4.820***	5.309***
	(0.063)	(0.037)
Observations	2,963	2,963
R^2	0.225	0.223
Adjusted R^2	0.225	0.222
Residual Std. Error	0.389 (df = 2959)	0.389 (df = 2958)
F Statistic	286.844*** (df = 3; 2959)	212.510*** (df = 4; 2958)
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table 8: Regression Results for (h)

	Dependent variable:
	lwage
educ	0.031***
	(0.007)
KWW	0.013***
	(0.001)
exper	−0.008
	(0.007)
educ:exper	0.003***
	(0.001)
Constant	5.164***
	(0.092)
Observations	2,963
R^2	0.232
Adjusted R^2	0.231
Residual Std. Error	0.387 (df = 2958)
F Statistic	223.396*** (df = 4; 2958)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Standard errors in parentheses

Table 9: Regression Results for (j)

	Dependent variable:	
	lwage	
	M1	M2
	(1)	(2)
educ	0.020***	
	(0.003)	
HS		0.113***
		(0.021)
College		0.180***
		(0.025)
KWW	0.016***	0.016***
	(0.001)	(0.001)
black	−0.133***	−0.129***
	(0.019)	(0.019)
Constant	5.490***	5.640***
	(0.044)	(0.036)
Observations	2,963	2,963
R^2	0.209	0.211
Adjusted R^2	0.208	0.210
Residual Std. Error	0.393 (df = 2959)	0.392 (df = 2958)
F Statistic	260.601*** (df = 3; 2959)	198.149*** (df = 4; 2958)
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table 10: Regression Results for (k)

	Dependent variable:	
	lwage	
	M1	M2
	(1)	(2)
HS		0.084^{***}
		(0.030)
College		0.138^{***}
		(0.052)
educ	0.016^{***}	0.002
	(0.003)	(0.007)
black:educ		0.017^{**}
		(0.007)
KWW	0.016^{***}	0.016^{***}
	(0.001)	(0.001)
black	-0.366^{***}	-0.345^{***}
	(0.085)	(0.089)
educ:black	0.019^{***}	
	(0.007)	
Constant	5.551^{***}	5.656^{***}
	(0.049)	(0.073)
Observations	2,963	2,963
R^2	0.211	0.213
Adjusted R^2	0.210	0.212
Residual Std. Error	0.392 (df = 2958)	0.392 (df = 2956)
F Statistic	197.861^{***} (df = 4; 2958)	133.532^{***} (df = 6; 2956)