Name: _____ Spring 2018

EC200 Econometrics and Applications

Unit 1 Quiz - SOLUTIONS

You have 75 minutes to complete this quiz. There are 36 total points. Please round to three decimal places when necessary. Note that some sub-questions build on previous sub-questions; if you get stuck on one part, at least make a guess so have you have values to apply to the following parts - you won't be penalized a second time if the numbers you carry over are incorrect.

1. *(8 points)* We've spent a decent amount of time talking about the zero conditional mean assumption behind our ordinary least squares estimates, that $E[u_i|X_i] = 0$. Consider one of two research questions:

   - A: What is the relationship between city-level crime rates and city-level median household income in 2015?
   - B: What is the relationship between quarterly CO_2 emissions and quarterly GDP in the United States, from 1980-2014.

   (a) For *each* research question, identify the dependent variable and independent variable. *[2 points]*

   A: Dependent variable: Crime rates. Independent variable: Median HH income

   B: Dependent variable: CO_2 emissions. Independent variable: Quarterly GDP

   (b) Select question A *or* B and provide an example of a variable that would be captured in the error term but *would not* violate the zero conditional mean assumption. Explain your reasoning in one sentence. *[3 points]*

   Some possible answers:
   A: A correct answer would be something that affects crime rates but is not correlated with household income. One possibility would be temperature (higher temperatures associated with higher crime).
   B: A correct answer would be something that affects CO_2 emissions but is not correlated with quarterly GDP. One possibility is policy changes that restrict CO_2 emissions (so long as the implementation of policy wasn't influenced by GDP)

(c) Select question A *or* B (same or different from your choice in part b) and provide an example of a variable that would be captured in the error term but *would* violate the zero conditional mean assumption. Explain your reasoning in one sentence. *[3 points]*

Some possible answers:
A: A correct answer would be something that affects crime rates and is correlated with household income. One possibility would be police officers per capita, if areas with more income can afford more police, and if areas with more police lead to more (or fewer) crimes .
B: A correct answer would be something that affects CO_2 emissions but is correlated with quarterly GDP. One possibility would be major macroeconomic shocks, like the 2008 financial crisis or Covid-19 pandemic, which affects both emissions and GDP

2. *(14 points)* Consider the relationship between annual per-capita cheese consumption (*cheese*, dependent variable) and unemployment rates (*unemp*, independent variable). You collect the following data:

Year	2005	2006	2007	2008
Per-capita cheese consumption (pounds):	31.7	32.6	33.1	32.7
Unemployment rate (percent):	4.9	4.4	5.0	7.3

(a) Write a population model for the relationship between per-capita cheese consumption and unemployment rates. *[2 points]*

$cheese_i = \beta_0 + \beta_1 unemp_i + u$

(b) Estimate $\widehat{\beta}_0$ and $\widehat{\beta}_1$. *[4 points]*

$\widehat{cheese_i} = 32.0517 + 0.08765 unemp_i$

(c) Interpret $\widehat{\beta}_0$ and $\widehat{\beta}_1$, making sure to include appropriate units. *[2 points]*

$\widehat{\beta}_0$ is the intercept - in an imaginary world with 0% unemployment, we would expect per-capita cheese consumption to be 32.05 pounds. $\widehat{\beta}_1$ is the slope coefficient. A one percentage point increase in unemployment rates is associated with an additional 0.088 pounds of cheese consumption per person.

(d) Calculate the residual for 2008. What does it mean? *[2 points]*

$$\begin{aligned} \hat{u_{2008}} &= y_{2008} - \hat{y}_{2008} \\ &= 32.7 - (32.05 + 0.088 * 7.3) \\ &= .0085 \end{aligned}$$

(e) Calculate R^2. How much does unemployment explain per-capita cheese consumption? *[4 points]*

$$\begin{aligned} R^2 &= \frac{SSR}{TSS} \\ &= \frac{1.008934263}{1.0475} \end{aligned}$$

3. *(14 points)* Consider the following summary statistics and regression results from a nationally representative random sample of 1,172 mothers surveyed in the 2016 General Social Survey.

   where variables are defined as follows:

   - `hrs1` = hours worked last week
   - `_doesnthurt` = 1 if agrees that it doesn't hurt children for their mothers to work outside of home, 0 otherwise
   - `childs` = number of children
   - `educ` = years of completed education
   - `age` = age in years

```
.
. regress _hrs1  _doesnthurt childs  educ age if _female == 1 & childs > 0
      Source |       SS           df       MS      Number of obs   =     1,172
-------------+----------------------------------   F(4, 1167)      =     70.12
       Model |  102760.201         4  25690.0502   Prob > F        =    0.0000
    Residual |  427534.475     1,167  366.353449   R-squared       =    0.1938
-------------+----------------------------------   Adj R-squared   =    0.1910
       Total |  530294.676     1,171  452.856256   Root MSE        =     19.14

------------------------------------------------------------------------------
       _hrs1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_doesnthurt |    .5572165   1.120283     0.50   0.619    -1.640778     2.75521
      childs |   -.8703755    .426837    -2.04   0.042    -1.707829   -.0329217
        educ |    1.146882   .1924429     5.96   0.000     .7693097    1.524455
         age |   -.4831164   .0339084   -14.25   0.000    -.5496446   -.4165883
       _cons |     31.4043   3.456793     9.08   0.000     24.62208    38.18653
------------------------------------------------------------------------------
```

   (a) What share of respondents believe that it doesn't hurt children for their mothers to work? *[2 points]*

   49.7 %
   answer in percents (0%-100%)

(b) Interpret $\widehat{\beta}_{childs}$, the coefficient on the number of children. Exactly what does it mean, in words? Is it statistically significant? Is it large? *[2 points]*

> $\widehat{\beta}_{childs} = -0.87$, which means that among women with at least one child, each child is associated with working 0.87 hours per week less, and it is statistically significant at the 5% level (with a p-value of 0.042). It is fairly modest in magnitude: on average, women in this sample work 19.4 hours per week, so it is a 4% decrease in work hours per child.

(c) Fill out the following table based on the regression results above. *[2 points]*

R^2	0.1938	TSS	530294.676
ESS	102760.201	SSR	427534.475
d.f.	1167	SER	$\sqrt{(427534.475/1167)} = 19.14$

(d) Under what assumptions would it be the case that $\widehat{\beta}_{childs}$ is BLUE? Which assumptions are likely to hold, which are not, and for which would you need more information? Give your reasoning for *each* assumption you list. *[4 points]*

> i. Zero conditional mean assumption: nothing in error term correlated with number of children and work hours. This is not likely to hold, as there may be multiple factors correlated with both, such as marital status. If we think of it as a binary variable equal to 1 if married, that might be positively correlated with number of children and negatively correlated with work hours. You could imagine occupation might be correlated with both as well.
>
> ii. Observations are iid. This is likely to hold, since the sample is drawn from a representative sample of respondents at one point in time.
>
> iii. No extreme outliers (finte kurtosis): this is likely to hold, as we can see from summary statistics (plus, there are fairly natural limits to all variables included)
>
> iv. Homoskedasticity: would need more information about whether variance changes with number of children (but often does not hold)

(e) Consider two friends, Drithi and Ayako. They are both 33-year-old mothers who believe that when mothers work outside of the home, it could hurt their children. However, Drithi is a high school graduate (12 years completed) with 3 children, and Ayako is a college graduate (16 years completed) with 2 children. What is the *difference* in predicted number of work hours between them? *[2 points]*

$$\widehat{hours}_D = 31.404 + 0.557 - 0.870 * 3 + 1.147 * 12 - 0.483 * 33$$
$$\widehat{hours}_A = 31.404 + 0.557 - 0.870 * 2 + 1.147 * 16 - 0.483 * 33$$
$$\widehat{Dif} = -0.870(3 - 2) + 1.147(12 - 16)$$
$$\widehat{Dif} = -5.458$$

We expect that Ayako works 5.5 hours more per week than Drithi.

(f) Marius says that since older people likely have had more children, $\widehat{\beta}_{childs}$, the impact of number of children on hours of work, is biased. Do you agree? Explain. *[2 points]*

It may be the case that older people have had more children on average. However, the coefficient $\widehat{\beta}_{childs}$ reflects the relationship between an increase in number of children and hours worked, holding *all else equal*. Since age is already in the regression, no bias is introduced.