Name: _____
EC200 Econometrics and Applications

# Unit 3 Practice Exam

There are 42 total points. Show all your work to receive full credit.

1. *[12 points]* A researcher investigating the determinants of crime in the United Kingdom has data for 42 police regions over 22 years. She estimates by OLS the following regression:

$$ln(cmrt)_{it} = \beta_1 unrtm_{it} + \beta_2 proyth_{it} + \beta_3 ln(pp)_{it} + \alpha_i + \phi_t + u_{it}$$
$$i = 1, ..., 42; \quad t = 1, ..., 22$$

where *cmrt* is the crime rate per head of population, *unrtm* is the unemployment rate of males, *proyth* is the proportion of youths, *pp* is the punishment rate, measured as (number of convictions)/(number of crimes reported). $\alpha$ and $\phi$ are area and year fixed effects, where $\alpha_i$ equals one for area $i$ and is zero otherwise for all $i$, and $\phi_t$ is one in year $t$ and zero for all other years for $t = 2, ..., 22$. Note that there is no intercept and that $\phi_1$ is not included.

   (a) What is the purpose of excluding $\phi_1$? What are the terms $\alpha$ and $\phi$ each likely to pick up? Discuss the advantages of using panel data for this type of investigation. *[4 points]*

   i. Purpose of excluding $\phi_1$:

   ii. What does $\alpha$ pick up?

   iii. What does $\phi$ pick up?

   iv. Advantages of panel data:

(b) The results of the previous estimation are reported below, where the fixed effects are included but the coefficients are not reported. Heteroskedasticity and autocorrelation-consistent standard errors are reported below each coefficient.

$$\widehat{ln(cmrt)}_{it} = \underset{(0.109)}{0.063 unrtm_{it}} + \underset{(0.179)}{3.739 proyth_{it}} - \underset{(0.024)}{0.588 ln(pp)_{it}}$$
$$R^2 = 0.904$$

Discuss the meaning of the three reported coefficients and their statistical significance. What is the effect of a ten-percent increase in the punishment rate? What assumptions, if any, are necessary in order to interpret this as the *causal* impact of the punishment rate on crime rates? *[5 points]*

i. Discuss coefficients:

ii. Effect of 10% increase in punishment rate?

iii. Assumptions for causal impact?

(c) You now want to analyze what happens to the coefficients and their standard errors when the equation is re-estimated without fixed effects. In the resulting regression, $\widehat{\beta}_2$ and $\widehat{\beta}_3$ do not change by much, although their standard errors roughly double. However, the coefficient on $unrtm$, $\widehat{\beta}_1$, is now 1.340 with a standard error of 0.234. Why do you think that is? *[3 points]*

2. *(15 points)* Consider the following model of education and fertility, using a 1988 survey of women in Botswana.

$$children_i = \beta_0 + \beta_1 edu_i + \beta_2 age_i + \beta_3 age_i^2 + u_i$$

Define *children* as the number of living children that women $i$ has, *edu* is her years of completed education, and *age* is her reported age in years.

```
. regress children educ age agesq
```

| Source | SS | df | MS | | Number of obs = | 4361 |
|---|---|---|---|---|---|---|
| | | | | | F( 3, 4357) = | 1915.20 |
| Model | 12243.0295 | 3 | 4081.00985 | | Prob > F = | 0.0000 |
| Residual | 9284.14679 | 4357 | 2.13085765 | | R-squared = | 0.5687 |
| | | | | | Adj R-squared = | 0.5684 |
| Total | 21527.1763 | 4360 | 4.93742577 | | Root MSE = | 1.4597 |

| children | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| educ | -.0905755 | .0059207 | -15.30 | 0.000 | -.102183 | -.0789679 |
| age | .3324486 | .0165495 | 20.09 | 0.000 | .3000032 | .364894 |
| agesq | -.0026308 | .0002726 | -9.65 | 0.000 | -.0031652 | -.0020964 |
| _cons | -4.138307 | .2405942 | -17.20 | 0.000 | -4.609994 | -3.66662 |

(a) Interpret the coefficient on *edu*, including units. *[2 points]*

(b) Does *age* have a non-linear relationship with the number of live children? How do you know? *[2 points]*

(c) Ronalda hypothesizes that we can use the variable $frsthalf$, which is a dummy equal to one if the woman was born during the first six months of the year, as an instrument for education. Her intuition is that in some countries, students can drop out at a certain age, so some students born earlier in the year may complete less schooling. What assumption(s) need to hold in order for her instrument to be valid? Which, if any, can be tested? *[4 points]*

(d) Compare the OLS results (above) and instrumental variable results (below). How do the returns to education differ between the two specifications? Explain why you think that is. *[3 points]*

```
. ivregress 2sls children (educ = frsthalf) age agesq
Instrumental variables (2SLS) regression              Number of obs =    4361
                                                      Wald chi2(3)  = 5300.22
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.5502
                                                      Root MSE      =    1.49
```

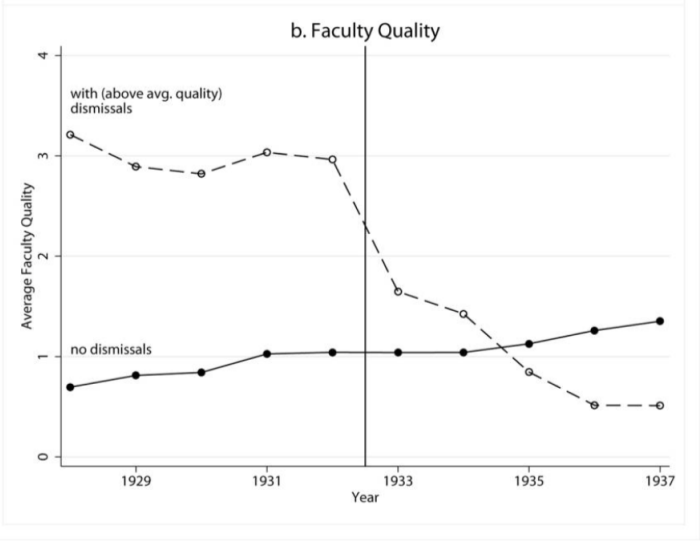
| children | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| educ | -.1714989 | .0531553 | -3.23 | 0.001 | -.2756813 | -.0673165 |
| age | .3236052 | .0178514 | 18.13 | 0.000 | .2886171 | .3585934 |
| agesq | -.0026723 | .0002796 | -9.56 | 0.000 | -.0032202 | -.0021244 |
| _cons | -3.387805 | .5478988 | -6.18 | 0.000 | -4.461667 | -2.313943 |

```
Instrumented:  educ
Instruments:   age agesq frsthalf
```

(e) Name one specific example of a potential threat to the validity of the instrument $frsthalf$. Explain why it is a threat. *[2 points]*

(f) Provide a specific example of reverse causality that might arise in this model. *[2 points]*

3. *(15 points)* Waldinger (2010) looks at the impact of faculty quality on the outcomes of PhD students. He uses a natural experiment caused by the dismissal of scientists in Nazi Germany. Upon gaining power in 1933, the new Nazi government fired all Jewish and "politically unreliable" scholars from German universities. Depending on the composition of their faculty, some universities had to dismiss more than half of their faculty, while others were not affected at all. The following figure summarizes the impact of this policy on faculty quality between universities with high levels of dismissals (dashed line) and those with no dismissals (solid line).




(a) Suppose you have panel data on graduate productivity and faculty quality at the university-year level. Aside from potential measurement error, provide one specific reason why the following equation would be unlikely to measure the causal impact of faculty quality on PhD outcomes. *[3 points]*

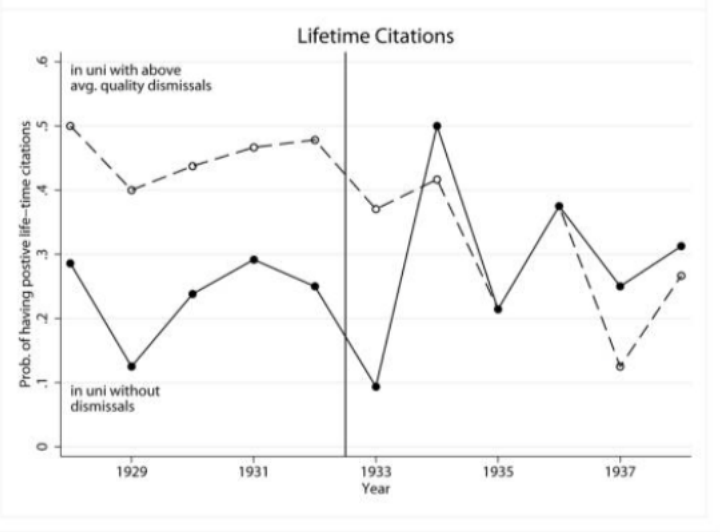
$$PhDCitations_{uni,y} = \beta_0 + \beta_1 FQual_{uni,y} + u_{uni,y}$$

$PhDCitations_{uni,y}$ is the number of lifetime citations the graduating $y$ cohort from university $uni$ receives. $FQual_{uni,y}$ is the number of citation-weighted publications per faculty member at university $uni$ in year $y$.

(b) Based on the description above, propose a difference-in-differences model that you could

use to estimate the causal impact of faculty quality on graduate student productivity. Define any new variables you use. Indicate which coefficient(s) will show the causal impact. *[5 points]*

(c) Name what, if any, assumptions you need for the model you write in part (b) to reflect a causal relationship. *[2 points]*




(d) The above figure shows the average level of *PhDCitations* by year separately for universities with a high number of dismissals and those without dismissals. Based on that figure, does faculty quality affect graduate student productivity? Explain how you know. *[3 points]*

(e) What, if any, concerns do you have about the *external validity* of your model? Explain. *[2 points]*