Final Exam (100 possible points)
ECNS 561
Fall 2016

Name__________________________

1.) (10 points) Regression through the Origin.

If we force the regression line through the origin, we are constraining the intercept to be zero. This is called a "regression through the origin." In the single-variable case, show that the $\tilde{\beta}$ that minimizes the least squares function is given by

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} x_i \, y_i}{\sum_{i=1}^{n} x_i^2}$$

2.) (15 points) Omitted Variable Bias.

Suppose that a population has the following relationship:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

However, we forget about x_2 and estimate the following misspecified equation:

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1$$

What is the relationship between the true β_1 and $\widehat{\beta_1}$ and how do we quantify the bias? [*Hint*: Consider that you can express the population relationship between x_1 and x_2 as $x_2 = \delta_0 + \delta_1 x_1 + u$.]

3.) (15 points) Rescaling of Data.

Let $\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_k}$ be the OLS estimates from the regression of y_i on $x_{i1}, \dots, x_{ik}$, $i = 1, 2, \dots, n$. For nonzero constants $c_1, \dots, c_k$, argue that the OLS intercept and slopes from the regression of $c_0 y_i$ on $c_1 x_{i1}, \dots, c_k x_{ik}$, $i = 1, 2, \dots$, n, are given by $\widetilde{\beta_0} = c_0\widehat{\beta_0},\ \widetilde{\beta_1} = (c_0/c_1)\widehat{\beta_1}, \dots ,\ \widetilde{\beta_k} = (c_0/c_k)\widehat{\beta_k}$. [*Hint*: Use the fact that the $\widehat{\beta_J}$ solve the OLS first order conditions and the $\widetilde{\beta_J}$ must solve the first order conditions involving the rescaled dependent and independent variables.]

4.) (15 points) OLS Estimator is Consistent.

Consider the following regression model under the first five Gauss-Markov assumptions

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

Show that the OLS estimator $\widehat{\beta_1}$ is a consistent estimator of β_1. Make sure to state any necessary assumptions required in your proof.

5.) (5 points) Functional Form.

As we have discussed in class, a commonly run regression is the log-log model,

$$\ln(y) = \beta_0 + \sum_k \beta_k \ln(x_k) + \varepsilon.$$

Show that β_k measures the percentage change in y associated with a one percent change in x_k.

6.) (15 points total) Deployments, Combat Exposure, and Crime

The following problem is adapted from the paper I gave you all to check out:

Anderson, D. Mark and Daniel Rees. 2015. "Deployments, Combat Exposure, and Crime." *Journal of Law and Economics* 58: 235-267.

To explore the relationship between combat exposure and violent crime near Fort Carson, Colorado, Anderson and Rees (2015) estimated an equation similar to the following:

$$\ln(\textit{Violent Crime}_{jt}) = \alpha_0 + \alpha_1(\textit{Never-Deployed Brigades}_{jt}) + \alpha_2(\textit{Brigades Returned} > \textit{6 Months}_{jt}) + \alpha_3(\textit{Brigades Returned} \leq \textit{6 Months}_{jt}) + \boldsymbol{X_{jt}\beta} + \varepsilon_{jt}.$$

where *Violent Crime* is the number of violent crimes per 1,000 population in police agency j and month t. *Never-Deployed Brigades* is equal to the number of combat brigades at Fort Carson that had not been deployed to Iraq prior to month t, *Brigades Returned > 6 Months* is equal to the number of previously deployed combat brigades at Fort Carson that had not returned from overseas in the past 6 months, and *Brigades Returned ≤ 6 Months* is equal to the number of previously deployed combat brigades at Fort Carson that had returned from overseas in the past 6 months. The vector $\boldsymbol{X}$ includes year indictors, month indicators, agency indicators, and the county unemployment rate.

Combat Exposure and Crime: Returned from a Deployment in the Past 6 Months	
	Violent Crime
Never-Deployed Brigades	.0506*** (.0123)
Brigades Returned > 6 Months	.0022 (.0115)
Brigades Returned ≤ 6 Months	.0026 (.0157)
Observations	2,192

* Statistically significant at 10% level; ** at 5% level; *** at 1% level

Note: Estimated coefficients are from an OLS regression. Coefficient estimates for the constant and variables included in $\boldsymbol{X}$ are omitted for the sake of brevity.

a.) (5 points) Provide an exact interpretation of the statistically significant coefficient estimate.

b.) (5 points) Formally state how you would test whether previously-deployed brigades had the same effect on violent crime as never-deployed brigades. You do not have to carry out any tests, but, by eye-balling the results, what would you conclude?

c.) (5 points) Do you think the variables *Brigades Returned > 6 Months* and *Brigades Returned ≤ 6 Months* are worthwhile measures of combat exposure? Can you think of better measures? Do you recall alternative proxies for combat exposure used by Anderson and Rees (2015)? (Stay brief in your answer and use only the space provided below).

7.) (10 points total) Functional Form

This problem is motivated by the following passage I recently read in an article on the age-productivity profile:

"Mathematicians, theoretical physicists, and poets typically make their most important contributions at younger ages than do astronomers, biologists, and novelists; psychologists have argued that this is because creative ideations can be produced and elaborated more rapidly in disciplines that deal with abstract conceptual entities than in those whose central ideas are more complex and concrete" (Galenson and Weinberg 2001).

Suppose we have data on the ages and productivity of thousands of professionals from the 6 occupations listed above. Suppose further that our measure of "productivity" is a different metric for each occupation. For instance, maybe this could be something like book sales/year for novelists.

a.) (5 points) Specify the regressions you would run if you wanted to estimate the age-productivity profile for these occupations. In general, what would you expect your results to look like? (i.e., what signs would you expect on your coefficients of interest?)

b.) (5 points) Given your regression equations above, can you solve for the peak age of productivity? How would you do this?

8.) (15 points total) Interaction Terms

In their paper, "The Effect of Occupational Licensing on Consumer Welfare: Early Midwifery Laws and Maternal Mortality", Anderson et al. (2016) use state-level data from 1900-1940 on midwifery licensing requirements and maternal mortality rates to estimate the following:

$$\ln(\textit{Maternal Mortality}_{st}) = \beta_0 + \beta_1 \textit{Midwifery License Required}_{st} + \boldsymbol{X_{st}\beta_2} + \varepsilon_{st}, \qquad (1)$$

where *s* indexes states and *t* indexes years. The dependent variable is measured as the maternal mortality rate per 100,000 females in state *s* during year *t*. The independent variable of interest, *Midwifery License Required*, is equal to one if midwives in state *s* were required by state law to be licensed in year *t*. The vector ***X*** contains year indicators, state indicators, state demographics (e.g., %White), and a set of policy controls (e.g., a women's suffrage indicator).

While equation (1) is informative, it masks over possible heterogeneous effects of the laws. In particular, there were three types of licensing laws. In some states, a license could be obtained by simply receiving basic instruction from a public health nurse or county health official. In more stringent states, license applicants had to pass an exam, typically administered by the State Board of Health; and, in the strictest states, license applicants had to be graduates from a recognized school of midwifery. To test for heterogeneous effects, the authors modified equation (1) as follows:

$$\begin{aligned}\ln(\textit{Maternal Mortality}_{st}) = \beta_0 &+ \beta_1 \textit{Midwifery License Required}_{st} \\ &+ \beta_2 \textit{Midwifery License Required}_{st} * \textit{Exam Sufficient}_{st} \\ &+ \beta_3 \textit{Midwifery License Required}_{st} * \textit{Graduation Necessary}_{st} \\ &+ \boldsymbol{X_{st}\beta_4} + \varepsilon_{st},\end{aligned}$$

where *Exam Sufficient* is equal to one if midwives were required to pass an exam, but were not required to have graduated from a recognized school of midwifery and *Graduation Necessary* is equal to one if midwives were required to have graduated from a recognized school of midwifery. The remaining variables are the same as defined above.

The following table on the next page illustrates results based on the estimation of this equation:

Heterogeneous Effects and Midwifery Laws	
	Maternal Mortality
Midwifery License Required	-.036*** (.013)
*Midwifery License Required * Exam Sufficient*	-.026 (.034)
*Midwifery License Required * Graduation Necessary*	-.052 (.054)
Mean of dependent variable	27.9
Observations	1,296
*Statistically significant at 10% level; ** at 5% level; *** at 1% level. Note: Estimated coefficients are from an OLS regression. Coefficient estimates for the constant and variables included in ***X*** are omitted for the sake of brevity.	

a.) (5 points) Suppose $\widehat{\beta_0} = -.014$. Solve for the intercept for states with the most basic licensing requirements.

b.) (5 points) What is the magnitude of the relationship between the strictest form of midwifery licensure and maternal mortality?

c.) (5 points) You are not required to conduct any formal tests here, but, by simply eye-balling the results, do you think we could reject the hypothesis that all three types of midwifery laws had comparable effects on maternal mortality?