Department of Economics
University of Maryland

John C. Chao
Spring 2013

Economics 422

Midterm Examination

This exam contains 5 questions. You should try to do all 5 questions. Please provide explanations for all your answers. Right answers with no explanation will receive no credit. On the other hand, wrong answers with thoughtful explanations may receive some credit. Partial credit will be given. **GOOD LUCK!**

QUESTION 1 (20 POINTS)

Suppose you ran a simple regression of the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i,\ i = 1, ..., 20$$

and obtain the following results

$$\begin{aligned}\text{(unadjusted) } R^2 &= 0.7911 \\ TSS &= 552.36,\end{aligned}$$

where TSS stands for Total Sum of Squares.

(a) Compute the **adjusted** R^2. Show your work.

(b) Compute SER and $RMSE$. Show your work.

QUESTION 2 (15 POINTS)

A soda vendor at University of Maryland football games observes that more sodas are sold the warmer the temperature at game time is. Based on 32 home games covering five years, the vendor estimates the relationship between soda sales and temperature to be

$$\begin{aligned}\widehat{Sodas} &= \widehat{\beta}_0 + \widehat{\beta}_1 \times Temp \\ &= -240 + 8 \times Temp\end{aligned}$$

where $Sodas$ is the number of sodas that the vendor sells and $Temp$ is the temperature in degrees Fahrenheit.

(a) Interpret the estimated slope and intercept. Do the estimates make sense? Why or why not?

(b) On a day when the temperature at game time is forecasted to be 80°F, predict how many sodas the vendor will sell.

(c) Below what temperature are the predicted sales zero?

QUESTION 3 (20 POINTS)

Consider a linear regression which seeks to explain scores on a particular standardized test based on whether one is left-handed, i.e.,

$$TS_i = \beta_0 + \beta_1 LH_i + u_i, \ i = 1, ..., n; \tag{1}$$

where TS_i is the score of individual i on this particular standardized test and LH_i is following dummy variable

$$LH_i = \begin{cases} 1 & \text{if person } i \text{ is left-handed} \\ 0 & \text{if person } i \text{ is right-handed} \end{cases}.$$

(Assume that everyone has a dominant hand and that no one is precisely ambidextrous.) Suppose that the regression above satisfies the standard assumptions for linear regression as discussed in class. In particular, assume that

$$E\left[u_i \middle| LH_i\right] = 0 \text{ with probability one.}$$

(a) Discuss how you would estimate the mean difference in test scores between people who are right-handed and people who are left-handed, i.e., the quantity

$$\tau = E\left[TS_i \middle| \text{ person } i \text{ is right-handed}\right] - E\left[TS_i \middle| \text{ person } i \text{ is left-handed}\right]$$

Write down an explicit formula for an unbiased estimator of τ. Explain why the estimator you have proposed is indeed unbiased.

(b) Consider the alternative regression specification

$$TS_i = \beta_1 LH_i + \beta_2 RH_i + \varepsilon_i, \ i = 1, ..., n; \tag{2}$$

where TS_i and LH_i are as defined above and RH_i is the dummy variable

$$RH_i = \begin{cases} 1 & \text{if person } i \text{ is right-handed} \\ 0 & \text{if person } i \text{ is left-handed} \end{cases}.$$

Suppose that regression (2) also satisfies the standard assumptions, as discussed in class. In particular, assume that

$$E\left[\varepsilon_i \middle| LH_i, RH_i\right] = 0 \text{ with probability one.}$$

Discuss how you would construct an unbiased estimator of the quantity

$$\tau = E\left[TS_i \middle| \text{ person } i \text{ is right-handed}\right] - E\left[TS_i \middle| \text{ person } i \text{ is left-handed}\right]$$

under specification (2). Explain why the estimator you have proposed is indeed unbiased.

QUESTION 4 (25 POINTS)

Consider the linear regression

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, ..., n.$$

Suppose that the following assumptions hold:

A1 (X_i, u_i) are $i.i.d.$ for $i = 1, 2, ..., n$

A2 Both X_i and u_i are continuous random variables

A3 $E[u_i|X_i] = 0$ with probability one for $i = 1, ..., n$

A4 $E[X_i^4] < \infty$ and $E[u_i^4] < \infty$, i.e., large outliers are rare.

A5 $Var(u_i|X_i) = \sigma_u^2$ for every i, i.e., u_i is homoskedastic.

Suppose one collects a large data set with $n = 1000$ observations on (Y_i, X_i), but chooses not to estimate β_1 by OLS. Instead, one uses two pairs of observations to obtain the estimator

$$\widetilde{\beta}_1 = \frac{Y_2 - Y_1}{X_2 - X_1}.$$

(a) Is $\widetilde{\beta}_1$ an unbiased estimator? If so, show that it is. If not, derive the bias of this estimator.

(b) Derive a formula for $Var\left(\widetilde{\beta}_1|X_1, ..., X_n\right)$, the conditional variance of $\widetilde{\beta}_1$; and compare it to $Var\left(\widehat{\beta}_1|X_1, ..., X_n\right)$, the conditional variance of the OLS estimator $\widehat{\beta}_1$. Discuss why one estimator might be more efficient than the other. (**Note:** in arguing for the efficiency of one estimator over the other, you need not give an explicit proof that the conditional variance of one estimator is smaller than that of the other. However, if you do supply such a proof, you will get 10 bonus points.)

QUESTION 5 (25 POINTS)

Suppose that you have collected data on 94 foreign countries to study the determinants of the differences in standard of living among the countries of the world. You recall from your macroeconomic lectures that the neoclassical growth model suggests that output per worker (per capita income) levels are determined by, among other things, the saving rate and population growth rate. To test the predictions of this growth model, you run the following regression

$$\begin{aligned}
\widehat{\text{RelPersInc}}_i &= \widehat{\beta}_0 + \widehat{\beta}_1 \times n_i + \widehat{\beta}_2 \times s_{K,i} \\
&= \underset{(0.068)}{0.339} - \underset{(3.177)}{12.894} \times n_i + \underset{(0.229)}{1.397} \times s_{K,i}, \\
R^2 &= 0.621, SER = 0.177,
\end{aligned}$$

where RelPersInc is GDP per worker of country i relative to the United States, n_i is the average population growth rate of country i from 1980 to 1990, and $s_{K,i}$ is the average investment share of GDP of country i from 1960 to 1990 (noting that investment equals saving). Numbers in parentheses are for heteroskedasticity-robust standard errors.

Suppose further that after you run the above regression, you remember that human capital also plays a role in determining the standard of living of a country. You therefore collect additional data on the average educational attainment in years for 1985, and add this variable ($Educ_i$) to the above regression. This results in the modified regression output:

$$\begin{aligned}
\widehat{\text{RelPersInc}}_i &= \widehat{\beta}_0 + \widehat{\beta}_1 \times n_i + \widehat{\beta}_2 \times s_{K,i} + \widehat{\beta}_3 \times Educ_i \\
&= \underset{(0.079)}{0.046} - \underset{(2.238)}{5.869} \times n_i + \underset{(0.294)}{0.738} \times s_{K,i} + \underset{(0.010)}{0.055} \times Educ_i, \\
R^2 &= 0.775, SER = 0.1377.
\end{aligned}$$

(a) Compare the R^2 values for the two regressions. Does the difference in the R^2 values alone allow you to conclude that $Educ_i$ is a significant variable which should be added into this regression? Explain your answer.

(b) Assuming homoskedasticity and normally distributed errors, use the F-statistic to carry out, at the 5% significance level, the following test:

$$H_0 : \beta_3 = 0 \text{ versus } H_1 : \beta_3 \neq 0$$

i.e,, a test of the null hypothesis that the coefficient on $Educ$ is zero versus the alternative hypothesis that it is different from zero.

(c) Suppose that you are pretty confident about the assumption of homoskedasticity but are unsure about the normality assumption. In this case, would you still use the same F-statistic as in part (b) or would you use a different version of the F-statistic in testing $H_0 : \beta_3 = 0$? What critical value would you now use to test $H_0 : \beta_3 = 0$ at the 5% significance level and why? Without assuming that the errors of the regression are normally distributed, carry out a test of the null hypothesis $H_0 : \beta_3 = 0$ versus the alternative hypothesis $H_1 : \beta_3 \neq 0$ at the 5% significance level. Relative to part (b), do you reach the same decision or a different decision in regard to rejecting or failing to reject the null hypothesis? Explain your findings.