Department of Economics
University of Maryland

John C. Chao
Spring 2009

Economics 422

Final Examination Solution Sheet

QUESTION 1

(a) Here, the large sample distribution of the F-statistic under H_0 is the $F_{3,\infty}$ distribution. Hence, using table 4 in the back of the exam, we find that at the 5% significance level the critical value is 2.60, while at the 1% significance level, the critical value is 3.78. Since in our case, $F = 3.21 > 2.6$, we reject H_0 at the 5% level; but since $3.21 < 3.78$, we fail to reject H_0 at the 1% level.

(b) Here, the large sample distribution of the F-statistic under H_0 is the $F_{7,\infty}$ distribution. Using table 4 in the back of the exam, we find that at the 5% significance level the critical value is 2.01, while at the 1% significance level, the critical value is 2.64. Since $F = 4.92 > 2.01$, we reject H_0 at the 5% level. Also, since $4.92 > 2.64$, we reject H_0 at the 1% level as well.

(c) Here, the large sample distribution of the F-statistic under H_0 is a $F_{5,\infty}$ distribution. Using table 4 in the back of the exam, we find that at the 5% significance level the critical value is 2.21, while at the 1% significance level, the critical value is 3.02. Since $F = 1.82 < 3.02$, we fail to reject H_0 at the 1% significance level. Also, since$1.82 < 2.21$, we fail to reject H_0 at the 5% level as well.

QUESTION 2

Note that in this case

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{D},$$

where $\overline{Y}$ is the average wage for the whole sample; and, from results given in the review session, we know that

$$\widehat{\beta}_1 = \overline{Y}_f - \overline{Y}_m$$

with $\overline{Y}_f$ denoting the average wage for the females in the sample and $\overline{Y}_f$ denoting the average wage for the males in the sample. Next, let

$$\mathbb{I}\{i = female\} = \begin{cases} 1 & \text{if } i^{th} \text{ individual is female} \\ 0 & \text{if } i^{th} \text{ individual is male} \end{cases}$$

and note that

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{i = female\} = \frac{n_f}{n},$$

where n_f is the number of females in the sample. Note also that the number of males in the sample $n_m = n - n_f$.

Hence,

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{Y} - \widehat{\beta}_1\overline{D} \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i - \left(\overline{Y}_f - \overline{Y}_m\right)\frac{n_f}{n} \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n} Y_i - \left(\frac{1}{n_f}\sum_{i=1}^{n} Y_i\mathbb{I}\{i = female\} - \frac{1}{n_m}\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right]\right) n_f\right\} \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i\mathbb{I}\{i = female\} + \frac{n_f}{n_m}\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right]\right\} \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n} Y_i\mathbb{I}\{i = female\} + \sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right] - \sum_{i=1}^{n} Y_i\mathbb{I}\{i = female\}\right. \\
&\quad \left. + \frac{n_f}{n_m}\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right]\right\} \\
&= \frac{1}{n}\left\{\left(\frac{n_m + n_f}{n_m}\right)\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right]\right\} \\
&= \frac{1}{n}\left\{\left(\frac{n}{n_m}\right)\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right]\right\} \\
&= \frac{1}{n_m}\sum_{i=1}^{n} Y_i\left[1 - \mathbb{I}\{i = female\}\right] \\
&= \overline{Y}_m
\end{aligned}$$

QUESTION 3

(a) Let β_0, β_1, and β_2 denote the intercept parameter, the coefficient on $teamera$, and the coefficient on *ops*, respectively. Note that the regression estimates implies that lowering the team ERA by one results in a winning percentage increase of roughly 10 percent. Moreover, increasing OPS by 0.1 generates a higher winning percentage of approximately

15 percentage. Note also that the regression explains 92 percent of the variation in winning percentages. Now, the t-statistics for the two slope coefficients are

$$
\begin{aligned}
t_{\widehat{\beta}_1} &= \frac{-0.099}{0.008} = -12.375, \\
t_{\widehat{\beta}_2} &= \frac{1.490}{0.126} = 11.825.
\end{aligned}
$$

Note that since $-12.375 < -1.645$, we reject the null hypothesis $H_0 : \beta_1 = 0$ in favor of the alternative hypothesis $H_1 : \beta_1 < 0$ at the 5% level. Also, since $11.825 > 1.645$, we reject the null hypothesis $H_0 : \beta_2 = 0$ in favor of the alternative hypothesis $H_1 : \beta_2 > 0$ at the 5% level. Hence, both slope coefficients are found to be statistically significant; and they are also important given the small differences in the winning percentage across different percentiles.

(b) Let β_3, β_4, and β_5 denote the coefficient on DAL, $DAL \times teamera$, and $DAL \times ops$, respectively. Note that the AL regression is one where the binary variable $DAL = 1$, so that the intercept, the marginal effect of *teamera*, the marginal effect of *ops* are, respectively, $\beta_0 + \beta_3$, $\beta_1 + \beta_4$, and $\beta_2 + \beta_5$ instead of β_0, β_1, and β_2, as they are for the NL regression. Hence, based on the results given, the two estimated regressions are as follows

$$
\begin{aligned}
\text{NL} &: \widehat{Winpct}_i = -0.29 - 0.100 \times teamera + 1.622 \times ops. \\
\text{AL} &: \widehat{Winpct}_i = -0.19 - 0.092 \times teamera + 1.435 \times ops.
\end{aligned}
$$

Now, the t-statistics are the AL coefficients (i.e., the t-statistics on all variables involving DAL) are

$$
\begin{aligned}
t_{\widehat{\beta}_3} &= \frac{0.10}{0.24} = 0.42, \\
t_{\widehat{\beta}_4} &= \frac{0.008}{0.018} = 0.44, \\
t_{\widehat{\beta}_5} &= \frac{-0.187}{0.160} = -1.17.
\end{aligned}
$$

Since the absolute value of all three t-statistics are less than 1.645, none of these coefficients are statistically significant at the 5% level.

(c) Using a large sample approximation, the distribution of the F-statistic under H_0 is an $F_{3,\infty}$ distribution. From table 4, the critical value at the 1% level is 3.78. Since in this case $F = 0.35 < 3.70$, we cannot reject the null hypothesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ at the 1% level. We might worry about the small sample here since the sample size is only 30, and we are making a large sample approximation.

QUESTION 4

The t-statistics are

$$\begin{array}{ll} \text{Data Set 1:} & t = \dfrac{-1.76}{\sqrt{0.37}} = -2.89, \\ \text{Data Set 2:} & t = \dfrac{0.0025}{\sqrt{0.000003}} = 1.44, \\ \text{Data Set 3:} & t = \dfrac{2.85}{\sqrt{117.5}} = 0.26, \\ \text{Data Set 4:} & t = \dfrac{-0.00014}{\sqrt{0.0000013}} = -0.123. \end{array}$$

From table 1 on the back of the exam, the associated p-values for testing against a two-sided alternative are

$$\begin{array}{ll} \text{Data Set 1:} & p\text{-value} = 0.0019 \times 2 = 0.0038, \\ \text{Data Set 2:} & p\text{-value} = (1 - 0.9251) \times 2 = 0.1498, \\ \text{Data Set 3:} & p\text{-value} = (1 - 0.6026) \times 2 = 0.7948, \\ \text{Data Set 4:} & p\text{-value} \approx 0.4522 \times 2 = 0.9044. \end{array}$$

Since all p-values are greater than 0.05 except for the first data set, we will reject H_0 only for data set 1.

BONUS QUESTION

Note that running OLS on equation (1), we get

$$\widehat{\pi}_1 = \frac{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)^2}$$

and running OLS on equation (2), we get

$$\widehat{\pi}_2 = \frac{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)^2}.$$

It follows that

$$\begin{aligned} \widehat{\beta}_1 &= \frac{\widehat{\pi}_1}{\widehat{\pi}_2} \\ &= \frac{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(Y_i - \overline{Y}\right) / \sum_{i=1}^n \left(Z_i - \overline{Z}\right)^2}{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(X_i - \overline{X}\right) / \sum_{i=1}^n \left(Z_i - \overline{Z}\right)^2} \\ &= \frac{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^n \left(Z_i - \overline{Z}\right)\left(X_i - \overline{X}\right)} \\ &= \widehat{\beta}_1^{TSLS}. \end{aligned}$$