Final Exam

You have two and a half hours for this exam. The exam is closed book, but you may use the SPSS help procedure. No scrap paper or notes are permitted. Please do all of your work carefully and check your answers before handing in your work. PLEASE SHOW AS MUCH OF YOUR WORK AS POSSIBLE TO AID IN THE AWARDING OF PARTIAL CREDIT IN THE EVENT OF AN ERROR ON YOUR PART.

*****************************************************************************

Note

Do all problems.

The data have been e-mailed to you.

Please do formal tests rather than rules of thumb in answering all questions.

Export your output and e-mail it to Patti at plkrebeh. Please include your name in the e-mail.

Good luck.
1 (32 points) I gathered some data on the United States for 26 years, 1980-2005 inclusive. The variables are

Year = Year
Imports = Imports in Billions of Dollars
GDP = Gross Domestic Product in Billions of Dollars

a. Please run a regression to determine the effect that GDP has on the level of imports. Please write the regression equation, i.e., the coefficients, below.

$$\text{Imports} = -317.17 + 1.17 \times \text{GDP}$$

b. In terms of the variables, please explain what the slope tells us.

As GDP rises by $1 \text{ (Billions)}$, we predict imports by rise by $1.172 \text{ (Billions)}$

c. Please run a t-test to test the hypothesis:

$$H_0: B_2 = 0$$

Vs.

$$H_A: B_2 \neq 0 \text{ at the .05 level.}$$

Show your work including relevant critical and test values.

$$t_{v-t} = \frac{b - \beta}{s_{b}} = \frac{1.172}{0.03} = 30.71$$

Reject $H_0$ if $|t| > 2.064$

$$30.71 > 2.064 \quad \text{reject } H_0$$

$$p = 0.05 < .05 \quad \text{reject } H_0$$
d. Please run the Durbin-Watson test at the .05 level. Show your work including relevant critical and test values.

\[ n = 26 \quad \text{independent vars} = x \]

\[ d_u = 1.302 \quad d_L = 1.461 \]

\[ \text{If} \quad d < d_L = 1.302 \quad \text{reject } H_0 \]

\[ \text{If} \quad d_L < d > d_u = 1.702 < d < 1.461 \quad \text{unreliable} \]

\[ \text{If} \quad d > d_u = 1.461 \quad d < 2.539 \quad \text{do not reject} \]

\[ \text{If} \quad 4 - d_u < d < 4 - d_L = 2.539 < d < 2.698 \quad \text{unsure} \]

\[ \text{If} \quad 4 - d_L < d \quad 2.698 < d \quad \text{do not reject} \]

\[ d = 4.73 \quad \text{reject} \]

e. If you could reject the null hypothesis in part (d), how would that affect your estimates of the slope and its standard error?

It would not affect the estimate of the slope, i.e., it would not bias the estimated slope. It would cause the variance in the standard error to be less large.

f. What is meant by the term stationary (as it applies to a variable, not what you write on!)

\[ \mu_Y = \mu \]

\[ \text{Var}(Y_t) = \sigma^2 \]

\[ \text{Cov}(Y_t - \mu)(Y_{t-1} - \mu) \text{ is some} \]

or (will state third one)

\[ \text{Cov}(Y_t - \mu)(Y_{t+h} - \mu) \text{ is some} \]

g. What regression and test would you run to see if the variable Imports is stationary? (Note: there are several correct answers.)

\[ \text{A) } \Delta Y_t = \beta_3 Y_{t-1} + \mu_t \]

\[ \text{B) } \Delta Y_t = \beta_1 + \beta_2 Y_{t-1} + \mu_t \]

\[ \text{C) } \Delta Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \mu_t \]
h. Run the regression and perform the relevant test. Show your work including relevant critical and test values.

\[ \Delta Y_A = 0.86 Y_{A-1} + \mu_A \quad t=6.347 \]

\[ \Delta Y_A = -7.987 + 0.94 Y_{A-1} + \mu_A \quad t=3.181 \]

\[ t = 6.477 > 1.96 \quad \text{not stationary} \]

\[ t = 3.177 > 1.96 \quad \text{not stationary} \]

\[ \Delta Y_A = 7.435 + 4.20 \text{Time} + 0.27 Y_{A-1} + \mu_A \quad t=1.219 \]

\[ \Delta Y_A = -532.9 + 4.20 \text{Time} + 0.27 Y_{A-1} + \mu_A \quad t=1.219 \]

i. What do your results in part (h) imply about the appropriateness of running an ordinary least squares regression of imports on GDP?

\[ \text{A or B} \quad \text{Not appropriate} \]

\[ \text{C would be appropriate} \]
2. (32 points) Suppose we decide to estimate the demand and supply equations in the agricultural sector in the United States. I gathered some data on agriculture for the United States for the years 1980-2004.

The variables are:

\[ Q = \text{Quantity of Agriculture production (An index where 1996 =100)} \]
\[ P = \text{Price of Agriculture production (An index where 1996=100)} \]
\[ \text{Inc} = \text{Per capita income in dollars (current dollars)} \]
\[ \text{Prod} = \text{Productivity – an index of agricultural output per unit of labor (1996=100)} \]

Let us assume we know that price and quantity are both endogenous. The model is

Demand: \[ Q_t = B_1 + B_2 P_t + B_3 \text{Inc}_t + \epsilon_{1t} \]

Supply: \[ Q_t = B_4 + B_5 P_t + B_6 \text{Prod}_t + \epsilon_{2t} \]

a. What are the exogenous/pre-determined variables in the model?

\[ \text{Income and Prod.} \]

b. Is it possible to obtain unbiased estimates of all of the parameters in the demand equation? Demonstrate.

\[ D \text{eq. } k' - k^* = 2 - 1 = 1 \]
\[ S \text{eq. } k - k = 2 - 1 = 1 \]

Is it possible to obtain unbiased estimates of all of the parameters in the supply equation? Demonstrate.

c. 

\[ \hat{Q} = 51.85 + 0.01 \text{Inc} + 2.75 \text{Prod} \]
\[ \hat{P} = 91.01 + 0.01 \text{Inc} + 0.01 \text{Prod} \]
e. For the equations that you can find unbiased estimates of the coefficients, please do so. State the name of the technique you used and write the coefficients below.

Demand
\[ Q_d = 77.4.56 - 2.59 \beta_{\text{Price}} + 0.64 L\text{ine} + e_d \]

Supply
\[ Q_s = -62.04 + 1.25 \beta_{\text{Price}} + 1.37 \beta_{\text{Prod}} + e_s \]

f. What is the name of the test we can run to see whether or not the model is in fact simultaneous?

Hausman Test or Pindyck & Rubinfeld

g. Run the test to examine whether or not Quantity is simultaneously determined. Show your work including relevant critical and test values.

\[ \beta_{\text{Price}} = 91.01 + 0.01 \beta_{\text{Line}} - 0.06 \beta_{\text{Prod}} + e_d \]

Hausman version
\[ \beta_{\text{Price}} = 37.4.96 - 3.954 \beta_{\text{Price}} + 0.04 \beta_{\text{Line}} - 0.0043 \beta_{\text{Prod}} + e_d \]

\( t = 1.23 \) 

-2.0 \leq t = 1.23 \leq 2.0 \quad \text{Do not reject } H_0: \beta_{\text{Price}} = 0 \text{ so not simult} }

Pindyck & Rubinfeld version
\[ \beta_{\text{Price}} = -62.04 + 1.25 \beta_{\text{Price}} - 1.29 \beta_{\text{Prod}} + e_s \]

\( t = 1.796 \)

\(-2.0 \leq t = 1.796 \leq 2.0 \text{ crit value} \quad \text{do not reject } H_0 \text{ so not simult} \)
3-6. (8 points each.)

3. I found a model for shrimp consumption in the United States. The data are for 48 quarters and are on the next page.
   a. Please explain, in terms of the variables, what the slope on $X_{2t}$, -0.0056, tells us.

   $\text{All else held equal, shrimp consumption is predicted to be lower when the second quarter.}$

   b. Please explain, in terms of the variables, what the slope on $Z_{2t}$, 0.0038, tells us.

   $\text{All else held equal, the effect of a } \text{10}\% \text{ increase in income will have a } \text{0.38}\% \text{ greater effect on shrimp consumption in the second quarter.}$

   c. Please test the hypothesis that all of the slopes are equal to zero together.

   $H_0: \beta_k = 0 \quad \text{vs} \quad H_a: \not \beta_k = 0$ 

   Reject $H_0$ if $F > F_{0.05, 48, 10} = 2.19$ 

   $23.5 > 2.19$ 

   Reject $H_0$

4. If heteroscedasticity exists in a model, what effect does it have on the slopes?

   a. If heteroscedasticity exists in a model, what effect does it have on the variance of the standard errors of the slopes?

   Will bias them

   b. How might we test for heteroscedasticity? State the name of the test and what we look for.

   Park, although other possible. See if $\beta_2$ signal

   c. How might we test for heteroscedasticity? State the name of the test and what we look for.

   Several forms $\ln e_2^2 = b_1 + b_2 \ln y$

   d. Suppose we find the variance of the errors is proportional to $X$. Explain how we would correct for heteroscedasticity.

   $\sigma_e^2 \text{ prop } X$

   $\sigma_e \text{ prop } \sqrt{X}$

   $\text{Weighed regr }^{\sqrt{X}} = b_1 + b_2 \frac{X}{\sqrt{X}} + \frac{e_2}{\sqrt{X}}$
Indicator Variables for Seasonal Differences

Suppose we are interested in a model that explains and predicts quarterly per capita consumption of shrimp in the United States. After considerable reflection on the factors that may cause variation in shrimp consumption per capita, we decide that the following model should be tested and estimated:

\[ C_t = \alpha + \beta_1 P_t + \beta_2 I_t + \beta_3 X_{2t} + \beta_4 X_{3t} + \beta_5 X_{4t} + \beta_6 Z_{2t} + \beta_7 Z_{3t} + \beta_8 Z_{4t} + u_t \]

where
- \( C_t \) is per capita consumption of shrimp in quarter \( t \)
- \( P_t \) is wholesale price of shrimp in quarter \( t \)
- \( I_t \) is per capita income in thousands of dollars in quarter \( t \)
- \( X_{2t} = 1 \) if the observation is in the second quarter of a given year
  = 0 otherwise
- \( X_{3t} = 1 \) if the observation is in the third quarter of a given year
  = 0 otherwise
- \( X_{4t} = 1 \) if the observation is in the fourth quarter of a given year
  = 0 otherwise
- \( Z_{2t} = I_t \ast X_{2t} \)
- \( Z_{3t} = I_t \ast X_{3t} \)
- \( Z_{4t} = I_t \ast X_{4t} \)

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Regression coefficient</th>
<th>Estimated standard error</th>
<th>Computed t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0621</td>
<td>0.0513</td>
<td>1.21</td>
</tr>
<tr>
<td>Price ( (P_t) )</td>
<td>-0.1448</td>
<td>0.0326</td>
<td>4.43</td>
</tr>
<tr>
<td>Income ( (I_t) )</td>
<td>0.1670</td>
<td>0.0262</td>
<td>6.38</td>
</tr>
<tr>
<td>( X_{2t} )</td>
<td>-0.0056</td>
<td>0.0716</td>
<td>0.08</td>
</tr>
<tr>
<td>( X_{3t} )</td>
<td>0.0383</td>
<td>0.0708</td>
<td>0.54</td>
</tr>
<tr>
<td>( X_{4t} )</td>
<td>0.0372</td>
<td>0.0697</td>
<td>0.53</td>
</tr>
<tr>
<td>( Z_{2t} )</td>
<td>0.0038</td>
<td>0.0339</td>
<td>0.11</td>
</tr>
<tr>
<td>( Z_{3t} )</td>
<td>-0.0001</td>
<td>0.0333</td>
<td>0.00</td>
</tr>
<tr>
<td>( Z_{4t} )</td>
<td>0.0008</td>
<td>0.0326</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\[ F = \frac{0.8282/8}{0.1718/39} = 23.5 \]
5. a. In a logit regression, what is the exact form of the dependent variable that we estimate?

\[ \ln \left( \frac{\hat{y}}{1-\hat{y}} \right) = \ln \left( \frac{p}{1-p} \right) \]

b. A Dutch economist, J.S. Cramer, estimated a logit model of car ownership as a function of income. In space to the right, show the approximate shape of a logit regression. Put Car Ownership on the vertical axes and income on the horizontal.

c. Cramer found that the estimated logit equation was \(-2.772 + 0.3476 \ln \text{Inc}\). What would be the probability that someone with an income of 20,000 will own a car? (Hint: look at part a to understand the dependent variable.)
6. a What is meant by the term panel data set?

Panel data is a pooling of cross-section and time series data. We follow a cross-section of observations over time.

b. Give an example of a panel data set, either real or hypothetical. Explain what an observation is.

Many examples say so states \( x = 1 \ldots 50 \) 3 years \( t = 1, 2, 3 \)

\[ \text{Taxes}_{xt} = b_0 + b_2 \text{Income}_{xt} + e_{xt} \]

c. Given your model in b state a regression that we might run on it

Gather data on income \( \text{Income}_{xt} \) and taxes for the 50 states for the three years.

d. Please add fixed effects to your model in c and state what they show us.

Could see if there was a difference between the years.

\[ \text{Taxes}_{xt} = b_0 + b_2 \text{Income}_{xt} + b_3 \text{Time 1}_{xt} + b_4 \text{Time 2}_{xt} + \epsilon_{xt} \]

\( \text{Time 1} = \{0 \text{ if } \text{Time 1} \} \quad \text{Time 2} = \{0 \text{ if } \text{Time 2} \} \)