Econometrics
Final Exam
May 27, 2010
Time for completion: 2h

Give your answers in the space provided.
Use draft paper to plan your answers before writing them on the exam paper.
Unless otherwise stated, use 5% for significance level.

Name:_________________________________________________ Number:_________

Group I (9 points, 1.5 for each question)

Give a very concise answer to the following questions. Conciseness will be valued, avoid unnecessary details.

1. The acronym ARCH stands for what in Econometrics?

_______________________________________________________________________

2. In one phrase, describe the meaning of “Inconsistency of the OLS estimator” in a multiple linear regression context.

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________

3. Explain why you would want to control for (or include) seasonal dummies in a multiple linear regression model for time series data.

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
4. Write the expression for the variance (in matrix form) of the Weighted Least Squares Estimator of the parameters of a multiple linear regression model. (Assume the variance of the error term is of known form and that the necessary assumptions hold)

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________

5. Describe succinctly a test aimed at detecting AR(q) serial correlation in the error term of a multiple linear regression model. Assume the strict exogeneity assumption does not fail and all the other necessary assumptions hold.

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________

6. Write a model aimed at testing whether a country having a government budget deficit above 8% of GDP and a public debt above 70% of GDP has an effect on the price of government bonds, controlling for other factors. Describe the variables you use and state the null hypothesis (and alternative) of the test.

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
### Group II (10 points)

1. Consider the following output of the model:

\[
\log(price)_i = \beta_0 + \beta_1 \log(nox)_i + \beta_2 \log(dist)_i + \beta_3 \text{rooms}_i + \beta_4 \text{stratio}_i + \beta_5 \text{crime}_i + u_i
\]

Dependent Variable: LPRICE  
Method: Least Squares  
Sample: 1 506  
Included observations: 506

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>11.00520</td>
<td>0.289504</td>
<td>38.01402</td>
</tr>
<tr>
<td>LNOX</td>
<td>-0.900736</td>
<td>0.106331</td>
<td>-8.471094</td>
</tr>
<tr>
<td>LDIST</td>
<td>-0.214767</td>
<td>0.039989</td>
<td>-5.370700</td>
</tr>
<tr>
<td>ROOMS</td>
<td>0.246468</td>
<td>0.016876</td>
<td>14.60438</td>
</tr>
<tr>
<td>STRATIO</td>
<td>-0.042209</td>
<td>0.005457</td>
<td>-7.734947</td>
</tr>
<tr>
<td>CRIME</td>
<td>-0.014839</td>
<td>0.001446</td>
<td>-10.26302</td>
</tr>
</tbody>
</table>

R-squared 0.656412  
Adjusted R-squared 0.652976  
S.E. of regression 0.241087  
Sum squared resid 29.06141  
Log likelihood 4.869826  
F-statistic 191.0465  
Prob(F-statistic) 0.000000

where \( \log(price) \) is the logarithm of median housing prices (in dollars), \( \log(nox) \) is the logarithm of the amount of nitrogen oxide in the air (in parts per million), \( \log(dist) \) is the logarithm of the weighted distance of the community from five employment centers (in miles), \( \text{rooms} \) is the average number of rooms in houses of the community, \( \text{stratio} \) is the average student-teacher ratio of schools in the community, and \( \text{crime} \) is the number of crimes committed per capita in the community. This model was estimated based in a sample of 506 communities in the Boston area.

**a)** Interpreteach one of the coefficient estimates of the model, \( \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4 \) and \( \hat{\beta}_5 \). Are they statistically significant? Justify. (1 point)

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________

---

3
The following output was also reported, using the residuals as well as the fitted values of the first regression:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>7.884271</td>
<td>3.370715</td>
<td>2.339050</td>
</tr>
<tr>
<td>FITTED_LPRICE</td>
<td>-1.461570</td>
<td>0.681831</td>
<td>-2.143597</td>
</tr>
<tr>
<td>FITTED_LPRICE^2</td>
<td>0.067749</td>
<td>0.034469</td>
<td>1.965496</td>
</tr>
</tbody>
</table>

R-squared 0.079340     Mean dependent var 0.057434
Adjusted R-squared 0.075679     S.D. dependent var 0.150449
S.E. of regression 0.144644     Akaike info criterion -1.023174
Sum squared resid 10.52367     Schwarz criterion -0.998116
Log likelihood 261.8631     Hannan-Quinn criter. -1.013347
F-statistic 21.67346     Durbin-Watson stat 0.988955
Prob(F-statistic) 0.000000

b) What can you conclude from this regression? Identify the test behind it. Does your conclusion affect the validity of the Gauss-Markov assumptions? What are the main consequences of your conclusion over the initial model? (2 points)
c) Given the information above, test the null hypothesis that the effect of one additional room per house equals the effect of one additional unit of the student-teacher ratio over the median housing prices in the Boston area. Formalize your answer, stating the null hypothesis, test statistic and the decision. (2 points)
2. Consider the following output of the model:

\[ r_{\text{sp}500_t} = \beta_0 + \beta_1 p_{\text{cip}_t} + \beta_2 p_{\text{cip}_{t-1}} + \beta_3 p_{\text{cip}_{t-2}} + \beta_4 i3_t + u_t \]

- Dependent Variable: RSP500
- Method: Least Squares
- Sample (adjusted): 4558
- Included observations: 555 after adjustments

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>20.17206</td>
<td>3.331966</td>
<td>6.054102</td>
</tr>
<tr>
<td>PCIP</td>
<td>0.129407</td>
<td>0.140465</td>
<td>0.921274</td>
</tr>
<tr>
<td>PCIP_1</td>
<td>-0.115392</td>
<td>0.148091</td>
<td>-0.779195</td>
</tr>
<tr>
<td>PCIP_2</td>
<td>-0.211874</td>
<td>0.140096</td>
<td>-1.512351</td>
</tr>
<tr>
<td>I3</td>
<td>-1.450676</td>
<td>0.542187</td>
<td>-2.675598</td>
</tr>
</tbody>
</table>

- R-squared: 0.019686
- Mean dependent var: 12.17559
- Adjusted R-squared: 0.012556
- S.D. dependent var: 40.26635
- S.E. of regression: 40.01275
- Akaike info criterion: 10.22524
- Schwarz criterion: 10.26415
- Hannan-Quinn criter.: 10.24044
- Log likelihood: -2832.505
- Durbin-Watson stat: 1.535318

where \( r_{\text{sp}500} \) represents the monthly return on the Standard & Poor’s (S&P’s) 500 stock market index at an annual rate, \( p_{\text{cip}} \) is the percentage change in the industrial production index at an annual rate, and \( i3 \) is the three-month T-bill annualized rate (in percentage points). This regression was obtained using a sample of monthly observations January 1947 through June 1993.

a) Interpret each of the coefficient estimates \( \hat{\beta}_1, \hat{\beta}_2, \) and \( \hat{\beta}_4 \). Construct a 95% confidence interval for the expected change in the monthly return on the S&P’s stock market index caused by an increase of 5 percentage points in the three-month T-bill rate. (1 point)

_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________
_______________________________________________________________________


Now consider the following output where RESIDUALS and RESID_1 denote the residuals of the previous regression in period \( t \) and in period \( t-1 \), respectively:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.796078</td>
<td>3.251655</td>
<td>0.244822</td>
</tr>
<tr>
<td>RESID_1</td>
<td>0.232675</td>
<td>0.041646</td>
<td>5.586942</td>
</tr>
<tr>
<td>PCIP</td>
<td>-0.076767</td>
<td>0.137284</td>
<td>-0.559183</td>
</tr>
<tr>
<td>PCIP_1</td>
<td>0.026234</td>
<td>0.144115</td>
<td>0.182039</td>
</tr>
<tr>
<td>PCIP_2</td>
<td>0.006958</td>
<td>0.136271</td>
<td>0.051063</td>
</tr>
<tr>
<td>I3</td>
<td>-0.107599</td>
<td>0.528586</td>
<td>-0.203561</td>
</tr>
</tbody>
</table>

- **R-squared**: 0.053909
- **Mean dependent var**: 0.102456
- **S.D. dependent var**: 39.83087
- **Akaike info criterion**: 10.17160
- **Schwarz criterion**: 10.21836
- **Hannan-Quinn criter.**: 10.18986
- **Durbin-Watson stat**: 1.966454

**b)** What can you conclude from the regression above? Which Gauss-Markov assumption is probably being violated? What are the consequences of your conclusions over the OLS estimators of the original model? (2 points)
The following output and the following matrix of correlations between the variables of the model were also reported:

Dependent Variable: RSP500
Method: Least Squares
Sample (adjusted): 4 558
Included observations: 555 after adjustments

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>16.78680</td>
<td>3.640013</td>
<td>4.611741</td>
<td>0.0000</td>
</tr>
<tr>
<td>PCIP</td>
<td>0.122710</td>
<td>0.139971</td>
<td>0.876680</td>
<td>0.3810</td>
</tr>
<tr>
<td>PCIP_1</td>
<td>-0.113567</td>
<td>0.147539</td>
<td>-0.769743</td>
<td>0.4418</td>
</tr>
<tr>
<td>PCIP_2</td>
<td>-0.203554</td>
<td>0.139620</td>
<td>-1.457912</td>
<td>0.1454</td>
</tr>
<tr>
<td>I3</td>
<td>-2.697227</td>
<td>0.770859</td>
<td>-3.498988</td>
<td>0.0005</td>
</tr>
<tr>
<td>T</td>
<td>0.034317</td>
<td>0.215642</td>
<td>2.977746</td>
<td>0.0053</td>
</tr>
</tbody>
</table>

R-squared: 0.028775
Mean dependent var: 12.17559
Adjusted R-squared: 0.019930
S.D. dependent var: 40.26635
S.E. of regression: 39.86309
Akaike info criterion: 10.21953
Sum squared resid: 872397.1
Schwarz criterion: 10.26622
Log likelihood: -2829.920
Hannan-Quinn criter.: 10.23777
F-statistic: 3.253095
Durbin-Watson stat: 1.542722
Prob(F-statistic): 0.006639

<table>
<thead>
<tr>
<th></th>
<th>RSP500</th>
<th>PCIP</th>
<th>PCIP_1</th>
<th>PCIP_2</th>
<th>I3</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSP500</td>
<td>1</td>
<td>0.02427687565</td>
<td>0.03838113839</td>
<td>0.039241768021</td>
<td>0.039227754097</td>
<td>0.0042467839</td>
</tr>
<tr>
<td>PCIP</td>
<td>0.02427687565</td>
<td>1</td>
<td>0.39241768021</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>-</td>
</tr>
<tr>
<td>PCIP_1</td>
<td>0.03838113839</td>
<td>0.39241768021</td>
<td>1</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>-</td>
</tr>
<tr>
<td>PCIP_2</td>
<td>0.06646080682</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>1</td>
<td>0.07231366083</td>
<td>-</td>
</tr>
<tr>
<td>I3</td>
<td>0.11024888553</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>0.01055212682</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>0.39227754097</td>
<td>0.07231366083</td>
<td>1</td>
</tr>
</tbody>
</table>
c) Can you conclude from the output above if the initial regression is spurious? Why? Do your conclusions change your answer to question b)? Justify your answer, using the above information. (2 points)
1. Suppose you wanted to estimate the following model for consumption, 
\[ c_i = a + b y_{i, p} + u_i, \] 
satisfying the Gauss Markov assumptions, but you are not able to 
know exactly \( y_{i, p} \), the level of permanent income; you have data only for the level of 
current income, \( y_i \), that is \( y_i = y_{i, p} + \varepsilon_i \) where \( \varepsilon_i \sim \text{Normal}(0, \sigma^2) \). In this situation, 
what can you say about the bias in the OLS estimator of \( b \) when \( y_i \) is used as the 
regressor? (show carefully all the steps and clarify any additional assumptions you 
might need)