Economics 475: Econometrics
Midterm

Student ID#: ___________________________

Please answer the following questions to the best of your ability. Remember, this exam is intended to be closed books, notes, programmable calculators, and neighbors. If you have any questions, please raise your hand. Be sure to show your work if you want partial credit. For any hypothesis tests you perform, full points will be given only if the null and alternative hypothesis are stated, the critical value of your test statistic is stated, and your test statistic is correctly computed. All hypothesis test should be done at the 95% level. Good Luck!

1. Over the course of 375 and 475, as I have been performing a number of Monte Carlo experiments to demonstrate how OLS works. As I've done this, I have noticed that for individual regressions when the estimated slope is a large number the estimated intercept is usually small number and vice versa. In other words, it appears that if OLS estimates too steep a slope, it "compensates" by producing too small an intercept (and vice versa). This has led me to hypothesize that in an univariate regression model, there is a negative covariance between $\hat{B}_0$ and $\hat{B}_1$. Using the fact that $\hat{B}_0 = \overline{Y} - \hat{B}_1 \overline{X}$ and any other facts that you might know, find the covariance between $\hat{B}_0$ and $\hat{B}_1$. Is it negative? (15)

$$Cov(\hat{B}_0, \hat{B}_1) = E\left[\left[\hat{B}_0 - E[\hat{B}_0]\right]\left[\hat{B}_1 - E[\hat{B}_1]\right]\right]$$

$$= E\left[\left(\bar{Y} - \hat{B}_1\bar{X} - \bar{Y} + B_1\bar{X}\right)\left(\hat{B}_1 - E[\hat{B}_1]\right)\right]$$

$$= E\left[-\bar{X}(\hat{B}_1 - B_1)(\hat{B}_1 - B_1)\right]$$

$$= E\left[-\bar{X}(\hat{B}_1 - B_1)^2\right]$$

$$= -\bar{X}\, E\left[(\hat{B}_1 - B_1)^2\right]$$

$$= -\bar{X}\, Var(\hat{B}_1)$$

$$= \frac{-\bar{X}\sigma^2}{\sum x_i^2}$$

2. A large amount of macroeconomic data is measured quarterly, that is 4 times per year. Imagine that you are interested in estimating the regression:

$$y_t = B_0 + B_1 x_t + \varepsilon_t$$

but are fearful that autocorrelation exists not only between the current period and the immediately previous period but also between the current period and the period four time periods earlier. In other words, you believe:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-4} + v_t$$

where ρ_1 and ρ_2 are parameters that measure the correlation between past error terms and the current error term and v is an independent random normal error term with mean zero and variance σ_v^2.

a. Assuming $\rho_1 \neq 0$ and $\rho_2 \neq 0$, if you estimated $y_t = B_0 + B_1 x_t + \varepsilon_t$ with OLS, what would happen? (5)

$\hat{B}_0$ & $\hat{B}_1$ are unbiased but inefficient.

b. Describe a method for estimating B_0 and B_1 that accounts for $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-4} + v_t$. Be as specific as possible—I'm looking for an equation you can estimate. (10)

(1) $Y_t = B_0 + B_1 X_t + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-4} + V_t$

(2) $\rho_1 Y_{t-1} = \rho_1 B_0 + \rho_1 B_1 X_{t-1} + \rho_1 \varepsilon_{t-1}$

(3) $\rho_2 Y_{t-4} = \rho_2 B_0 + \rho_2 B_1 X_{t-4} + \rho_2 \varepsilon_{t-4}$

Subtract (2) & (3) from (1) and get

(4) $Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-4} = (1 - \rho_1 - \rho_2) B_0 + B_1 (X_t - \rho_1 X_{t-1} - \rho_2 X_{t-4}) + V_t$

(1) Regress e_t on e_{t-1} and e_{t-4} and obtain estimates of ρ_1 & ρ_4

(2) Create NEW dependent and independent variables using $\hat{\rho}_1$ & $\hat{\rho}_4$

$Y_t - \hat{\rho}_1 Y_{t-1} - \hat{\rho}_2 Y_{t-4}$ $\quad$ $X_t - \hat{\rho}_1 X_{t-1} - \hat{\rho}_2 X_{t-4}$

(3) Regress NEW Y on NEW X

3. David Romer (1993) proposes theoretical models of inflation that imply that countries with more open international trade policies have lower inflation rates. His empirical analysis explains average annual inflation rates (since 1973) in terms of the average share of imports in gross domestic product—which is his measure of openness. His equation is

$$(1) \qquad \text{Inflation}_i = B_0 + B_1\text{Open}_i + B_2\text{Pcinc}_i + \varepsilon_{1i}$$

where inflation is a nation's annual inflation rate, open is imports/GDP, and pcinc is the per capita national income of the country (pcinc is assumed to be exogenous for the purposes of this problem).

Romer was skeptical of his results. He argues that the amount of openness in a country may be a function of the inflation rate (after all, citizens may be willing to import more if domestic prices are rising fast). Romer hypothesizes the equation:

$$(2) \qquad \text{Open}_i = a_0 + a_1\text{Inflation}_i + a_2 z_i + \varepsilon_{2i}$$

where z_i represents a variable which will be described later. Romer suggests a_1 should be positive.

a. If Romer's suspicions about a_1 are correct, what would happen to the OLS estimates of B_1 if equation (1) were estimated with OLS? (6)

$\text{Cov}(\varepsilon_1, \text{Open}) > 0$ and $\hat{B}_1$ would be biased upward

b. In attempting to estimate B_1, Romer suggests estimating the reduced form equation for Open and then substituting the predicted values from this equation into (1). For ease, I've written the correct reduced form equation for Open:

$$(3) \qquad Open_i = \Gamma_0 + \Gamma_1 z_i + \Gamma_1 Pcinc_i + v_i$$

Unfortunately, it is often difficult for researchers to actually find an appropriate z. What characteristics of z are required for Romer to use this procedure? For each characteristic you list, describe what happens to the estimates of B_1 if z does not have that characteristic but is used anyway. (10)

$Cov(z, \varepsilon_1) = 0$

$Cov(z, open) \neq 0$

c. Provide one suggestion for z. What makes your suggestion attractive? (5)

d. After finding the correct Z and using the procedure discussed in part b, , how would you go about testing for heteroskedasticity? (5)

1. Regress Inflation = $B_1 + B_2 \hat{Open} + B_3 Pcinc$ & save residuals e
2. Square residuals
3. Regress e^2 on $\hat{Open}$, Pcinc, squares & interactions
4. F-test or Chisquared test.

e. How would you correct any heteroskedasticity you found? (5)

1. Use squared residuals from #2 in part d
2. take $\ln e^2$ & regress on X's, squares & interactions
3. get predicted values of $\ln e^2$ from (2)
4. take exponential of predicted values from (3) call these $\hat{e}_i^2$
5. Take inverse square root of $\hat{e}_i^2$ and weight the original regression by this:

$$\frac{Inflation}{\sqrt{\hat{e}_i^2}} = B_0 \frac{1}{\sqrt{\hat{e}_i^2}} + B_1 \frac{Inflation}{\sqrt{\hat{e}_i^2}} + B_2 \frac{Pcinc}{\sqrt{\hat{e}_i^2}}$$

4. A neophyte econometrics student is interested in estimating the marginal propensity to consume for college students. Assume that the true population regression function describing consumption for individual college students is given by:

$$(1) \qquad c_i = \beta_0 + \beta_1 y_i + \varepsilon_i$$

where c is monthly spending and y is monthly income and i represents individual college students. Assume that each individual's error term is homoskedastic, uncorrelated with y, and is not correlated with any other individual's error term—in other words this error term fulfills the basic requirements of the Gauss-Markov proof.

Since our student is a neophyte, he collects data on 10 individuals and then aggregates the data by summing consumption and income. This generates one observation for our neophyte. Our neophyte then repeats this process each of the next 14 months yielding 15 observations (each of which is based upon 10 individual students). However, as the neophyte takes some econometrics courses, he realizes that larger sample sizes produce more precise results so, starting after month 15, the neophyte collects data on 50 individuals per month rather than 10. The neophyte continues this sampling process for 15 additional months generating a total of 30 aggregated observations.

a. Using the 30 months worth of aggregated data, our student estimates:

$$(2) \qquad C_t = a_0 + a_1 Y_t + v_i$$

Where, just to be clear, $C_t = \sum_{i=1}^{10} c_i$ for the first t = 1 through 15 observations and $C_t = \sum_{i=1}^{50} c_i$ for observations t = 16 through 30. Likewise, $Y_t = \sum_{i=1}^{10} y_i$ for the first 15 observations and $Y_t = \sum_{i=1}^{50} y_i$ for the last 15 observations

If (2) is estimated with OLS, are a_0 and a_1 unbiased and efficient estimates of β_0 and β_1? (10)

With a little work, it can be shown that (2) is heteroskedastic. With more work, it can be shown $E[\hat{a}_1] = \beta_1$ and $E[\hat{a}_0] \neq \beta_0$. So, one can use this process to discover an estimate for β_1 but not β_0.

<u>Heteroskedasticity</u>

For periods 1 - 15

$$V_i = \sum_{i=1}^{10} \varepsilon_i$$

$$Var(V_i) = E\left[\left(V_i - E[V_i]\right)^2\right] = E\left[(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \cdots + \varepsilon_{10})^2\right] = 10\sigma_\varepsilon^2$$

But for periods 16 - 30

$$V_i = \sum_{i=1}^{50} \varepsilon_i$$

$$Var(V_i) = E\left[(\varepsilon_1 + \varepsilon_2 + \cdots \varepsilon_{50})^2\right] = 50\,\sigma_\varepsilon^2$$

The variance of V is 5 times larger in the later periods relative to the earlier ones.

$E[\hat{a}_1] = \beta_1$?

$$\hat{a}_1 = \frac{Cov(C_t, Y_t)}{Var(Y_t)} \qquad C_t = \sum_{i=1}^{N} c_i \quad Y_t = \sum_{i=1}^{N} y_i$$

$$c_i = \beta_0 + \beta_1 y_i + \varepsilon_i$$

so $C_t = N\beta_0 + \beta_1 \sum y_i + \sum \varepsilon_i$

$$Cov(C_t, Y_t) = Cov\left(n\beta_0 + \beta_1 \sum y_i + \sum \varepsilon_i, \sum y_i\right)$$

$$\mathrm{Cov}(C_t, Y_t) = E\left[\left(nB_0 + B_1\Sigma y_i + \Sigma \varepsilon_i - nB_0 - E\left[B_1 \Sigma y_i + \Sigma\varepsilon_i\right]\right)\left(\Sigma y_i - E\left[\Sigma y_i\right]\right)\right]$$

$$= E\left[\left(\beta_1 \Sigma\left(y_i - \bar{y}\right) + \Sigma\varepsilon_i\right)\left(\Sigma y_i - \Sigma\bar{y}\right)\right]$$

$$= E\left[\left(\beta_1 \Sigma\left(y_i - \bar{y}\right) + \Sigma\varepsilon_i\right)\left(\Sigma\left(y_i - \bar{y}\right)\right)\right]$$

$$= E\left[\beta_1\left(\Sigma(y_i - \bar{y})\right)^2\right]$$

$$= \beta_1 E\left[\left(\Sigma(y_i - \bar{y})\right)^2\right]$$

$$\mathrm{Var}(y_t) = E\left[\left(\Sigma y_i - E\left[\Sigma y_i\right]\right)^2\right]$$

$$= E\left[\left(\Sigma(y_i - \bar{y})\right)^2\right]$$

So $E\left[\hat{a}_1\right] = \dfrac{\mathrm{cov}(C_t, Y_t)}{\mathrm{Var}(Y_t)} = \dfrac{\beta_1 E\left[\left(\Sigma(y_t - \bar{y})\right)^2\right]}{E\left[\left(\Sigma(y_t - \bar{y})\right)^2\right]} = \beta_1$