Econ 400

B. Brown Spring 2008

EXERCISE 6: Answers

Consider the following model of expeditures on health

$$E_i = \beta_1 + \beta_2 Y_i + \beta_3 S_i + u_i \quad \text{(i)}$$

where E_i is aggregate personal healthcare expediture in state i, Y_i is aggregate personal income, S_i is aggregate number of seniors (over 65) in 1993, and u_i is an unknown disturbance. The data is given in the dataset ECON400_EXR6.DAT, which I will provide on my website or via e-mail.

1. Assume that the errors are homoskedastic.
   a. Making explicit all your assumptions, what are the properties of OLS-based estimates and inferences? *Assume* $\mathrm{E}[u_i] = 0$, $\mathrm{E}[u_i^2] = \sigma^2$, $\mathrm{E}[u_i u_j] = 0$ *for* $i \neq j$, (Y_i, S_i) *nonstochastic,* (Y_i, S_i) *linear independent. The the OLS estimates are unbiased, consistent, and BLUE. If we add the assumption that* $u_i \sim N(o, \sigma^2)$ *then OLS will be MLE and BUE.*
   b. Run the indicated regression and report your results. *Running OLS on this model yields:* $\widehat{\beta}_1 = 0.3442\ (0.1326)$, $\widehat{\beta}_2 = 0.04480\ (0.002521)$, *and* $\widehat{\beta}_3 = 0.000147\ (0.1394)$.
   c. Test the hypothesis that $\beta_2 = 0$. Construct a confidence interval for β_2? What is the relationship between the hypothesis test and confidence interval? *Under* $H_0 : \ \beta_2 = 0$ *and the assumptions made,* $(\widehat{\beta}_2 - 0)/s.e.(\widehat{\beta}_2) = (0.04481 - 0)/0.002521) = 17.7677$ *is a realization from* t_{48}*. Cosulting the t-tables we find this to be to the left of the .005 LHS tail value of -2.68. So we reject at the 1% level. For a 95% interval we have* $\{\widehat{\beta}_2 \pm t_{48,.025} \times s.e.(\widehat{\beta}_2)\} = \{.04480 \pm 2.01 \times .002521\} = \{.03974, .04988\}$*. This interval represents the set of values that are acceptable as a null hypothesis with a choice of* $\alpha = .05$.
   d. Test the overall significance of the regression. That is, test whether $\beta_2 = \beta_3 = 0$. *The* F*-statistic provided by the regression is appropriate for testing all the slope coefficients being zero, which is the null here. In the present case this should have a* F*-distribution with 48 and 2 degrees of freedom, for which the critical-value for a 1% test is 5.08. Since the realization of 1539.62 exceeds this critical- value, we reject the null hypothesis.*

2. Suppose that the errors might be heteroskedastic
   a. How will the properties of OLS-based estimates and inferences be impacted? *If we suppose that* $\mathrm{E}[u_i^2] = \lambda_i^2 = \lambda^2(x_i)$ *is nonconstant, the OLS remains unbiased and consistent but is no longer BLUE. More importantly, the usual ratios such as* $(\widehat{\beta}_2 - 0)/s.e.(\widehat{\beta}_2)$ *will not have a* t*-distribution in finite samples nor a standard normal asymptotically, since we are estimating the wrong covariance matrix. Thus inferences will increase or decrease the probability of type I errors depending on whether the standard errors are under- or over-stated, so the size will be wrong.*
   b. Does there seem to be evidence of heteroskedasticity in this model? *We can utilize the White test to test for inferential stability resulting from possible heteroskedasticity. For the present case this statistic will have a chi-squared distribution with* $k(k+1)/2 - 1 = 5$ *degrees of freedom for which the RHS .01 critical-value is 15.086. Since the realization of 30.3828 exceeds the critical value, we reject the null hypothesis of homoskedasticity.*
   c. Discuss how the problem can be avoided using heteroskedastic consistent covariance (HCC) estimates. Test the hypothesis that $\beta_2 = 0$.using HCC estimated standard errors. *The HCC estimates consistently estimate the covariance matrix of OLS in the presence of heteroskedasticity and are hence asymptotically appropriate for forming* t*-type ratios, which will be asymptotically standard normal under the hypothesis being tested. In the*

present case we have $(.04481 - 0)/.00269 = 16.66$ *which is far out in the RHS tail of the standard normal, so we reject the null. Note that although we rejected heteroskedaticity, it did't make much difference for testing this statistic.*

d. Show how the problem can be corrected through the use of WLS. Calculate feasible WLS estimates using the White weights. *We can write* $\lambda_i^2 = z_i'\gamma$ *where* $z_i =< x_i \otimes x_i >$ *are the unique elements of the squares and cross-products of the regressors. We then estimate* γ *by regressing* e_i^2 *on* z_i*. We then use the estimated* $\widehat{\lambda}_i^2 = z_i'\widehat{\gamma}$ *to do feasible WLS on the original regression. Such a procedure yields:* $\widehat{\beta}_1 = 0.2746\ (0.0438)$, $\widehat{\beta}_2 = 0.04511\ (0.00273)$, *and* $\widehat{\beta}_3 = 0.1609\ (0.1103)$. *Note that the estimates are about the same for the first and second coefficient but the s.e. is rather differenct for the first.*

3. Suppose that we reformulate the model in the form

$$E_i/P_i = \beta_1 + \beta_2 Y_i/P_i + \beta_3 S_i/P_i + u_i. \qquad \text{(ii)}$$

where P_i is population is state i.

a. Run this regression and report your results. Running OLS on the so transfromed model yields $\widehat{\beta}_1 = 0.05723\ (0.0561)$, $\widehat{\beta}_2 = 0.04215\ (0.003356)$, *and* $\widehat{\beta}_3 = 0.0170\ (0.3424)$.

b. Does there seem to be heteroskedasticity in this formulation? *The realization of the White statistic of 0.9310 is not in the RHS tail of the chi-squared distribution with 5 degrees of freedom for any likely choice of* α*. Thus there does not seem to be heteroskedaticity in the so transfromed model.*

c. How are the estimates and inferences impacted? *The inferences on* β_2 *are essentially unchanged, but the estimates and* t*-ratios for* β_1 *and* β_3 *are somewhat different.*

d. If (ii) is the correct model, discuss why (i) would exhibit heteroskedasticity. Discuss the role that population might play in the heteroskedasticity. *If (ii) is the correct model with homoskedastic errors then the disturbances for (i) will have* $\mathrm{E}[u_i^2] = P_i^2\sigma^2$ *and the disturbances will be heteroskedatic. Since we are using state data the values for the aggregate variables* E_i, Y_i, *and* S_i *will be sums over the* P_i *individuals in state* i*. Suppose the relationship (i) holds at the individual level, then (i) for the aggregate variables will have* $\mathrm{E}[u_i^2] = P_i\sigma^2$*. Thus, strictly speaking we should divide by* $\sqrt{P_i}$ *to correct the heteroskedasticity.*