**ECON 6511: Advanced Applied Econometrics**

**Homework 3 January 31, 2018**

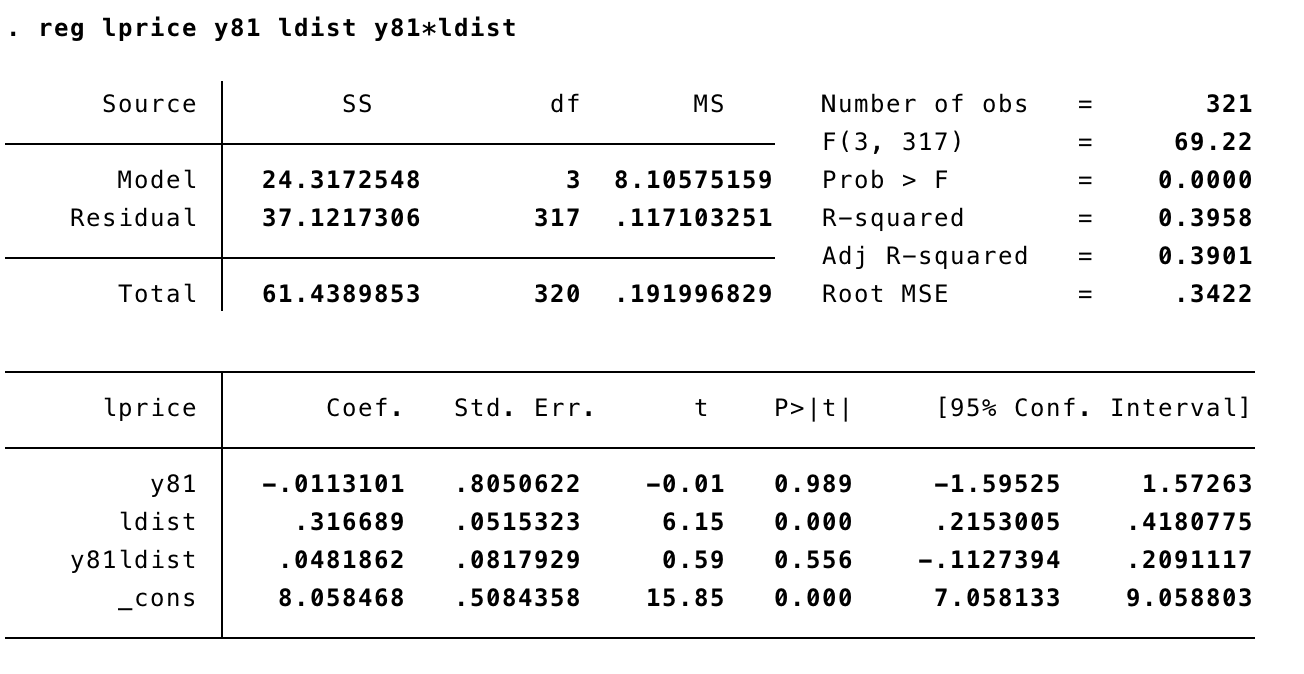
**Submitted By:**

**Surabhi Asati**

**va7892**

1. (Wooldridge, Chapter 13, Problem 3) Use the data in KIELMC.dta for this exercise.
2. The variable dist is the distance from each home to the incinerator site, in feet. Consider the model:  log(price) = β0 + δ0y81 + β1log(dist) + δ1y81 · log(dist) + u If building the incinerator reduces the value of homes closer to the site, what is the sign  ofδ1? What does it mean if β1 >0?

Model: log(price) = β0 + δ0y81 + β1log(dist) + δ1y81 · log(dist) + u



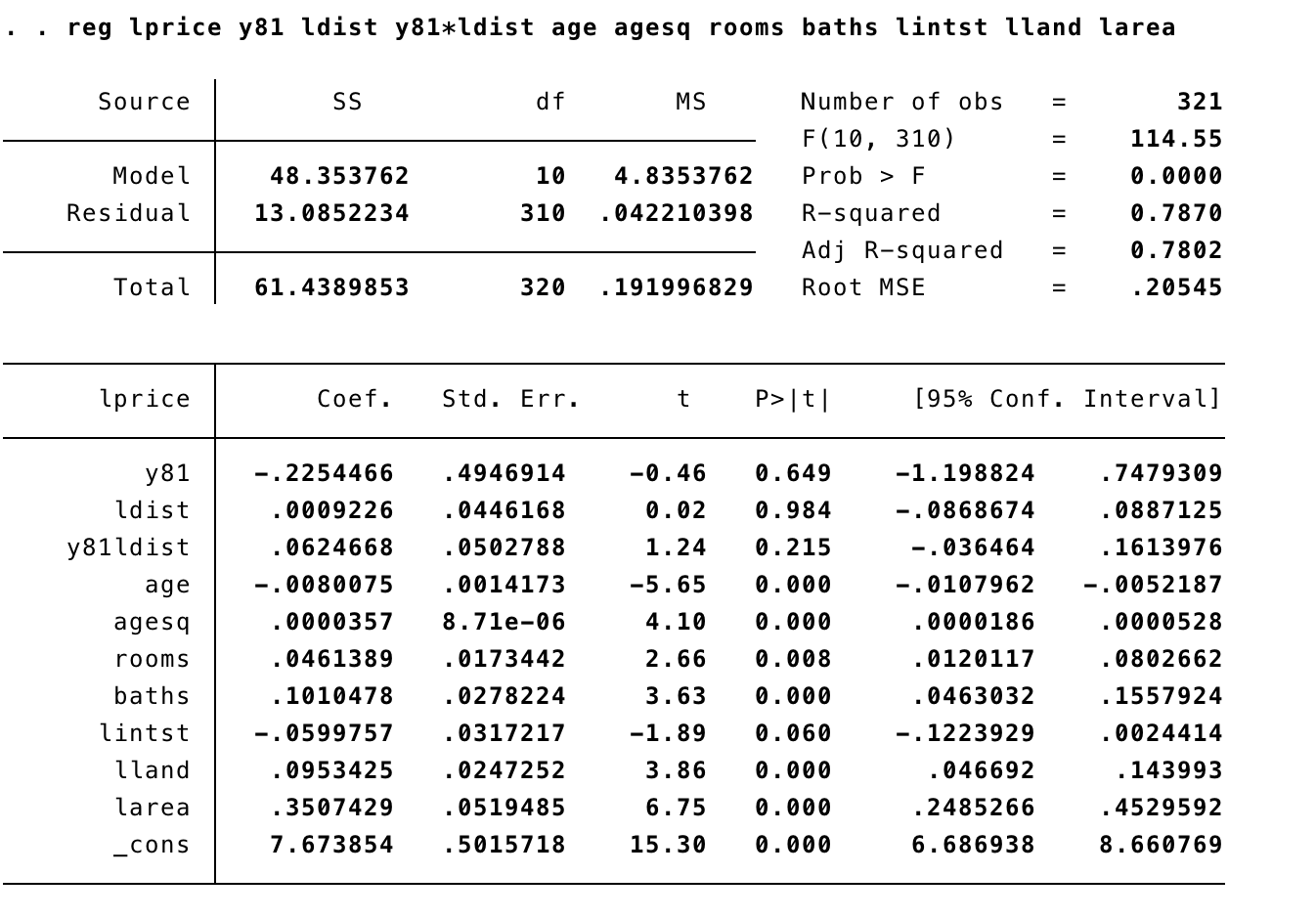
δ1 (= 0.048) > 0. It has a positive sign. As β1 ( = 0.316) > 0, it means price of the house increases by increasing the distance. Homes located farther away from incinerator were be expensive.

1. Estimate the model from part (a) and report the results in equation form. Interpret the  coefficient on y81 · log(dist). What do you conclude?

Estimated model: 8.058 – 0.011y81 + 0.316log(dist) + 0.048y81 · log(dist) + u

δ1 = 0.048 with t value = .59 It is not statistically significant as p value > 5%

1. Add age, age2, rooms, baths, log(instst), log(land), and log(area) to the equation. Now,  what do you conclude about the effect of the incinerator on housing values?



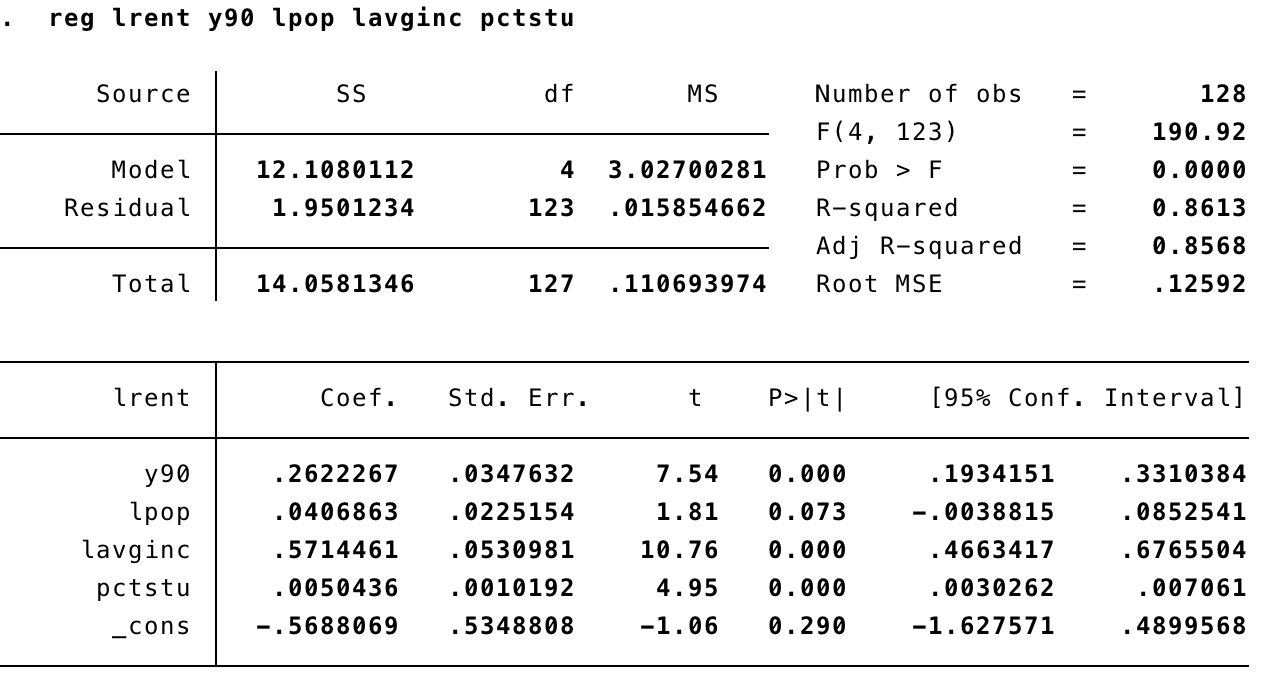
When we add other variables age, age2, rooms, baths, log(instst), log(land), and log(area) to the regression, the coefficient on y81 · log(dist) increases to .062 that means its effect is larger. t statistic = 1.24 which is not significant.

1. Why is the coefficient on log(dist) positive and statistically significant in part (b) but not in part (c)? What does this say about the controls used in part (c)?

The coefficient on log(dist) is positive and statistically significant in part (b) but not in part (c) because the controls used in part c are more significant than distance.

The controls used in part c have a larger effect on the price than distance.

1. (Wooldridge, Chapter 13, Problem 5) Use the data in RENTAL.dta for this exercise. The data for the years 1980 and 1990 include rental prices and other variables for college towns. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is:  log(rentit) = β0 + δ0y90t + β1log(popit) + β2log(avgincit) + β3pctstuit + ai + uit where pop is city population, avginc is average income, and pctstu is student population as  a percentage of city population (during the school year).



1. Estimate the equation by pooled OLS and report the results in equation form. What do  you make of the estimate on the 1990 dummy variable? What do you get for βˆpctstu?

Estimated Equation: log(rentit) = -.568 +.262 y90t + .040 log(popit) + .571 log(avgincit) + .005 pctstuit + ai + uit

The estimate on the 1990 dummy variable is positive d90 = .262 and is very significant with t statistic = 7.54 and p value < 5%. It means that the rents grew by more than 265 in 10 years keeping other factors fixed.

The coefficient on pctstu = .005means that a one percentage point increase in pctstu increases rent by 0.5%. The t statistic = 4.95 means it is very statistically significant.

1. Are the standard errors you report in part (a) valid? Explain.

The standard errors from part (i) are not valid, unless we take ai in the equation. If ai is in the error term, it contains everything that doesn’t change so it validates standard error.

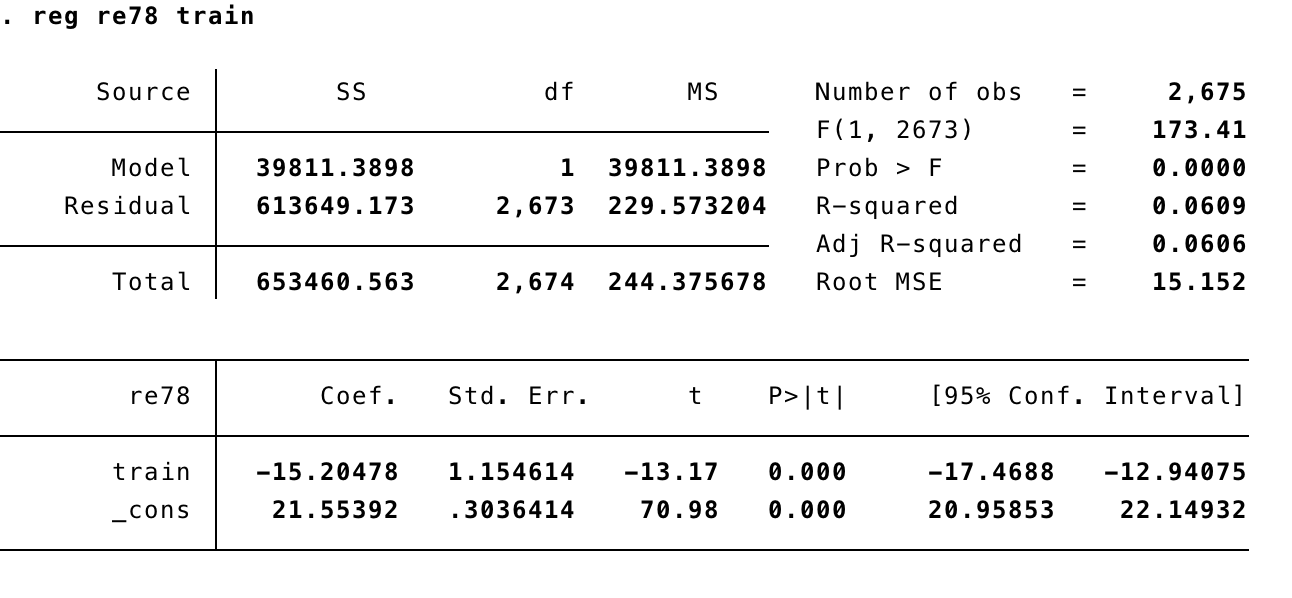
1. Now, difference the equation and estimate by OLS. Compare your estimate of βpctstu with that from part (b). Does the relative size of the student population appear to affect rental prices?

1 percentage point increase in pctstu increases rental rates by about 1.1%. So the effect of pctstu is very large. Estimated in differences = 132

1. Obtain the heteroskedasticity-robust standard errors for the first-differenced equation in part (c). Does this change your conclusions?

The heteroscedasticity-robust standard error for the first-differenced equation in part (c) on pctstu is 0.0027. It is smaller than OLS standard error and pctstu is more significant with t statistic =4

1. (Wooldridge, Chapter 13, Problem 14) Use the data in JTRAIN3.dta for this question. 1

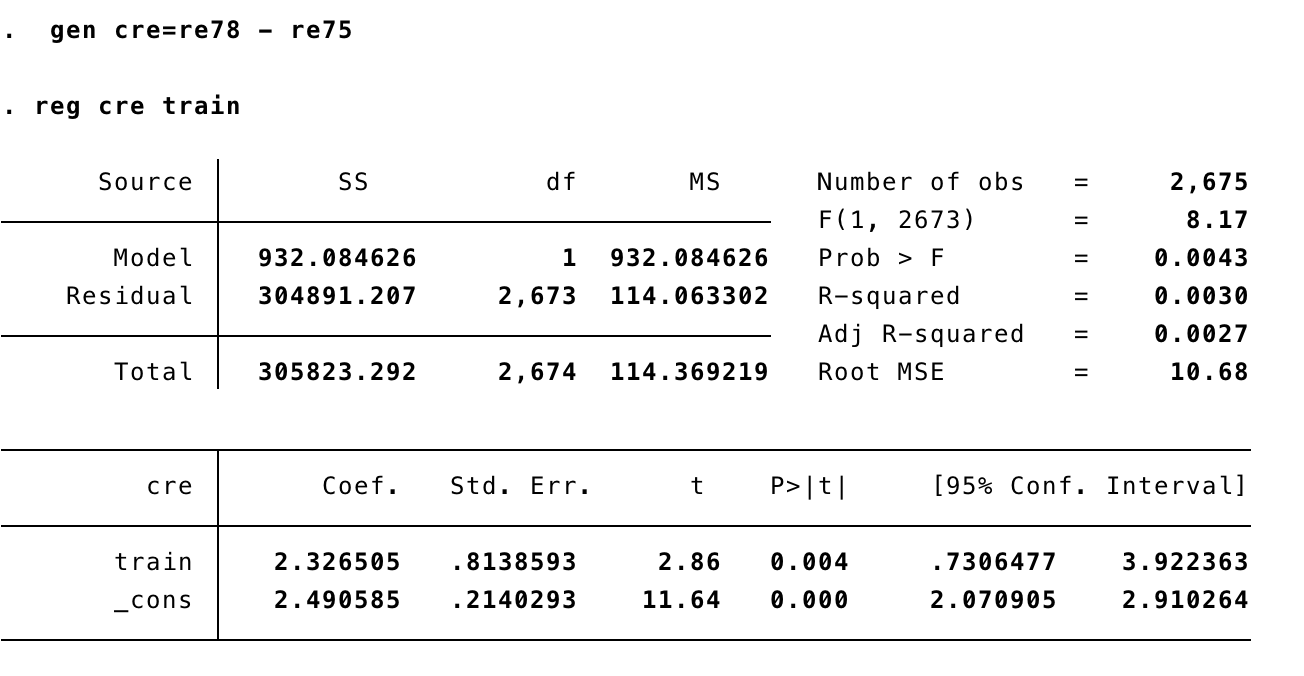


1. Estimate the simple regression model re78 = β0 + β1train + u, and report the results in equation form. Based on this regression, does it appear that job training, which took place in 1976 and 1977, had a positive effect on real labor earnings in 1978?

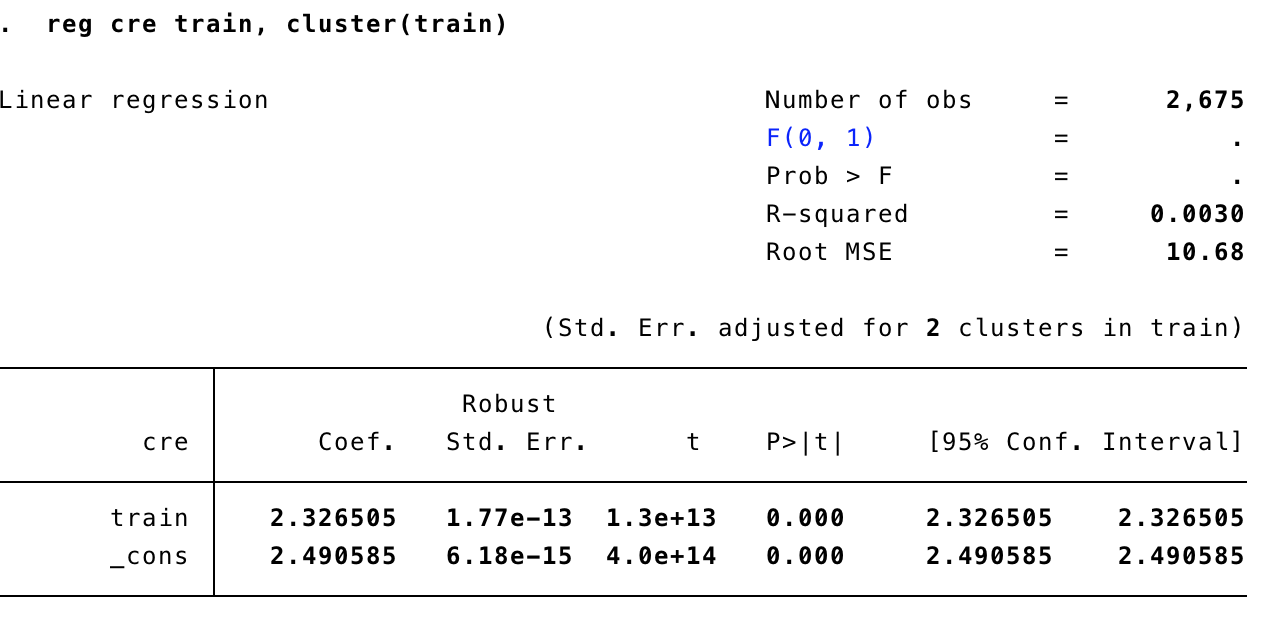
Estimated Model: re78 = 21.55 – 15.205 train + u

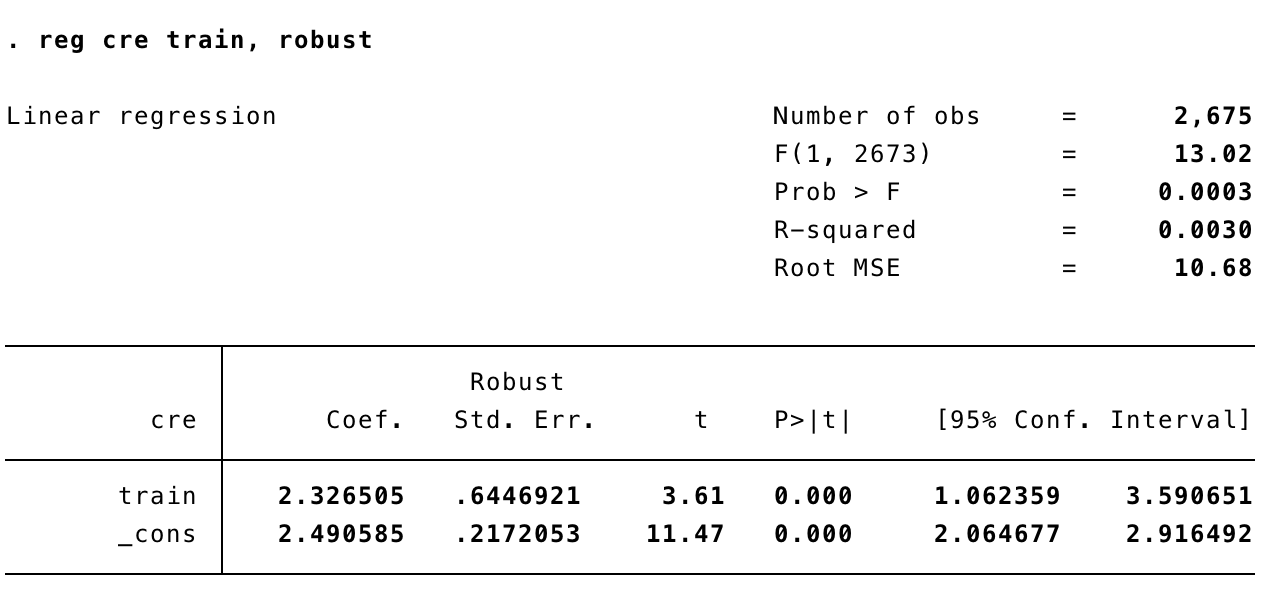
Based on this regression, it appears that job training, which took place in 1976 and 1977, had a negative effect on real labor earnings in 1978.

1. Now use the change in real labor earnings, cre = re78 − re75, as the dependent variable. (We need not difference train because we assume there was no job training prior to 1975. That is, if we define ctrain = train78 − train75 then ctrain = train78 because train75 = 0.) Now what is the estimated effect of training? Discuss how it compares with the estimate in part (a).



1. Find the 95% confidence interval for the training effect using the usual OLS standard error and the heteroskedasticity-robust standard error, and describe your findings.

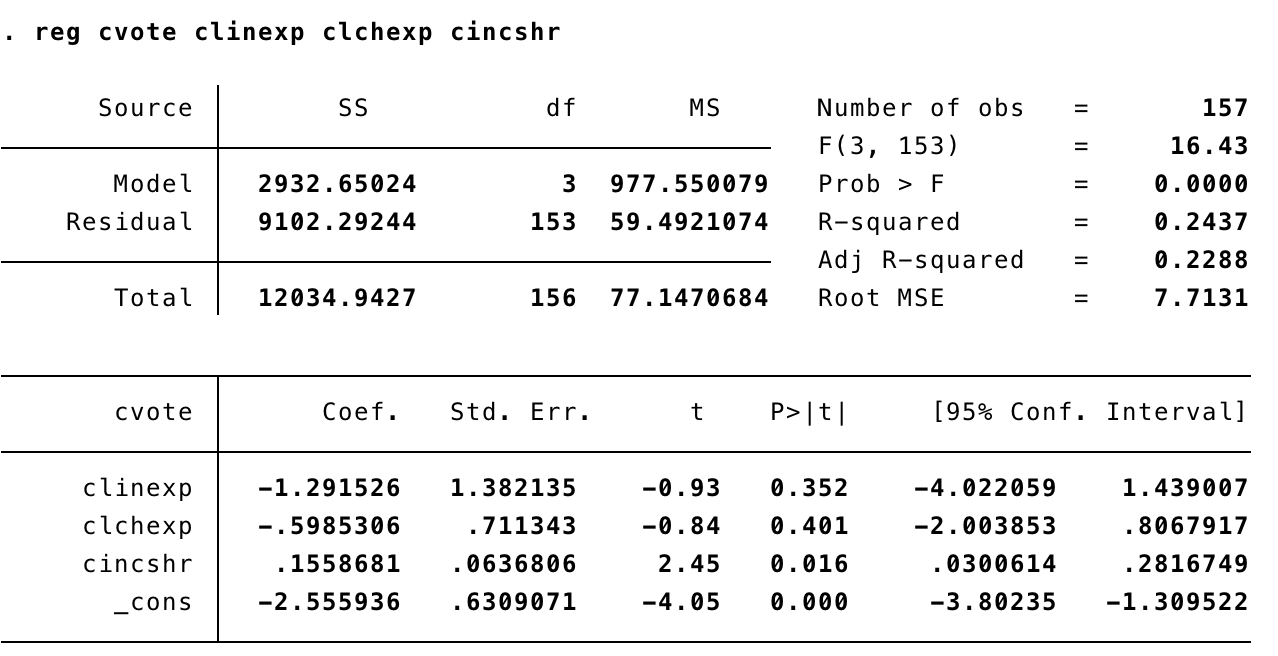




4. (Wooldridge, Chapter 13, Problem 8) VOTE2.data includes panel data on House of Repre- sentatives elections in 1988 an 1990. Only winners from 1988 who are also running in 1990 appear in the sample; these are the incumbents. An unobserved effects model explaining the share of the incumbent’s vote in terms of expenditures by both candidates is:

voteit = β0 + δ0d90t + β1log(inexpit) + β2log(chexpit) + β3incshrit + ai + uit,

where incshrit is the incumbent’s share of total campaign spending (in percentage form). The unobserved effect ai contains characteristics of the incumbent–such as “quality”–as well as things about the district that are constant. The incumbent’s gender and party are con- stant over time, so these are subsumed in ai. We are interested in the effect of campaign expenditures on election outcomes.



(a)  Difference the given equation across the two years and estimate the differenced equation by OLS. Which variables are individually significant at the 5% level against a two-sided alternative?

The estimated equation using differences:

vote = - 2.56 -1.29 log(inexp) - .599 log(chexp) + .156 incshr

Only incshr is statistically significant at the 5% level (t statistic = 2.45, p-value = .016)

The other variables are insignificant with p-values > 5%

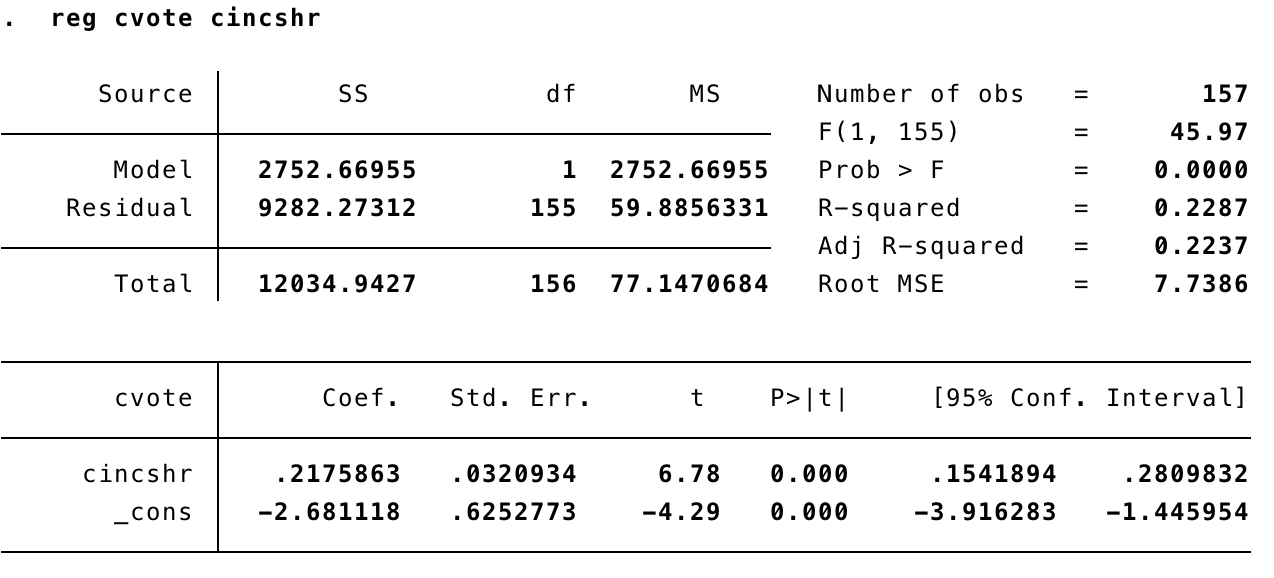
(b)  In the equation from part (a), test for joint significance of ∆log(inexp) and ∆log(chexp). Compute by hand and report the F-statistic.

Fq, N-k = ((RSSR – RSSUR) / q) RSSUR / (N-k)

The F statistic = 1.51 (df =153) with p-value = .224

So, log(inexp) and log(chexp) are jointly insignificant

(c)  Reestimate the equation from part (a) using ∆incshr as the only independent variable. Interpret the coefficient on ∆incshr. For example, if the incumbent’s share of spending increases by 10 percentage points, how is this predicted to affect the incumbent’s share of the vote?



The simple regression equation is

cvote = - 2.68 + .217 cincshr

This means if the incumbent’s share of spending increases by 10 percentage points

Increases the percent of the incumbent’s vote by about 2.17 percentage points.

(d)  Redo part (c), but now use only pairs that repeat challengers. [This allows us to control for characteristics of challengers as well, which would be in ai.]

vote = 2.68 + .218 incshr

t –value is not statistically significant as p- value = .14