Daniel Lesh
ECON 2311Q
10/30/2020

HW #2

1. $\widehat{MJ} = \hat{\beta}_0 + \hat{\beta}_1 \cdot UR$

$\widehat{MJ} = 8.971 + .244 \cdot UR$

$SE = (1.406) \quad (.148)$

a. Null & alt. Hypothesis for UR coefficient $\neq 0$:

$H_0: \beta_1 = 0$; $H_a: \beta_1 \neq 0$

b. Null & alt. Hypothesis for UR coefficient > 0.

$H_0: \beta_1 \leq 0$; $H_a: \beta_1 > 0$

c. Calculate t-stat:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \frac{.244 - 0}{.148} = \boxed{1.649}$$

d. Determine Critical Values for 2-sided hyp. test @ 1%, 5%, 10% levels:

Reject H_0 if $|t| > t_{\alpha/2,\, n-2}$

- 1%: $1.649 > t_{.01/2,\, 51-2} \longrightarrow 1.649 < 2.680$
- 5%: $1.649 > t_{.05/2,\, 51-2} \longrightarrow 1.649 < 2.010$
- 10%: $1.649 > t_{.10/2,\, 51-2} \longrightarrow 1.649 < 1.677$

- No, the coefficient estimate is not statistically significant at each of those levels. Do <u>not</u> reject null hypothesis test.

e. Determine Critical value for 1-sided hyp. test @ 5% sig. level

Reject H_0 if $|t| > t_{\alpha,\, n-2}$

- 5%: $1.649 > t_{.05,\, 51-2} \longrightarrow \boxed{1.649 < 1.677}$

- This coefficient estimate is also not statistically significant at this level. This means we should <u>not reject</u> the Null hypothesis for <u>both of these hypothesis tests.</u>

f. Calculate 95% CI for coefficient est. on UR:

$CI = \hat{\beta}_1 \pm 2.010 \times SE(\hat{\beta}_1)$

$= .244 \pm 2.010 \cdot (.148) \longrightarrow CI = (.541, -.053)$

2. $\widehat{Studenth} = 71.0 - 4.84 \cdot BFemme$, $R^2 = .40$, $SER = 2.0$
$\quad\quad\quad\quad (.3) \quad\quad (.57)$

a. With this line of best fit, we are estimating the relationship X, BFemme (female height, and Y, Studenth (student pop. height). The interpretation of intercept means that, according to the est. line, a university with 0 females would have a (predicted) height of 71 inches.
The interpretatiation of slope means that universities with one more female on average have a student pop. heights that is 4.84in smaller.
Females are 66.16in (71 - 4.84) on average.

b. $H_0: \hat{\beta}_1 \geq 0$; $H_a: \hat{\beta}_1 < 0$

t-stat: $\frac{-4.84 - 0}{.57} = \boxed{-8.49}$

critical val: $|t| > t_{\alpha, n-2}$

used one-sided t-stat @ 100 df

$|-8.49| > t_{.01, 109} \longrightarrow \boxed{8.49 > 2.3647}$

- at the 1% level, we would reject H_0 if $|t| > 2.3647$. In this case, $8.49 > 2.3647$, so we do reject the H_0.

c. No, this error term is not homoskedastic. To be homoskedastic, the error term (u) would not depend on x (BFemme). In this case, the error term does depend on x (BFemme), as it is a binary variable that is either 1 = Female or 0 = Male. Because of this, the error term is heteroskedastic, as u does depend on x in the regression presented in the problem.

3. $\widehat{RelPersInc} = .518 - 18.831 \cdot (n - n_{US})$, $R^2 = .522$, $SER = .197$
$\quad (.056) \quad (3.177)$

a. With this analysis, there is no reason to believe that the variance of the error terms is homoskedastic b/c the SE depends on the difference found between the avg. growth pop rate and the avg. growth pop rate of the U.S. in a specific year.

b. $H_0: \beta_1 = 0$; $H_a: \beta_1 \neq 0$

t-stat: $\frac{-18.831 - 0}{3.177} = \boxed{-5.927}$

critical vale for 2-sided hyp. test: Reject H_0 if $|t| > t_{\alpha/2, n-2}$ $\rightarrow$ >50, using 1.976

$|-5.927| > t_{.05/2, 104-2} \longrightarrow \boxed{5.927 > 1.976}$

- At the 5% level, we would reject H_0 if $|t| > 1.976$. In this case, $5.927 > 1.976$, so we do reject H_0. It is statistically significant as we find it different from zero.

4. $\widehat{RelProd} = -.08 + 2.44 \cdot s_k$, $R^2 = .46$, $SER = .21$
$\quad (.04) \quad (.38)$

a. With this line of best fit, we are estimating the relationship between x, s_k (the avg. inv. share savings of GDP from 1960 - 1990), and y, RelProd (amount of GDP produced per worker relative to the U.S.). The interpretation of the regression finds that with the U.S., one additional worker on average has a $2.44 increase in GDP.

b. $H_0: \beta_1 = 0$; $\beta_1 \neq 0$

t-stat: $\frac{2.44 - 0}{.38} = \boxed{6.421}$

Reject H_0 if $|t| > t_{\alpha/2, n-2}$: $\rightarrow$ >50, using 1.976 for .05 two-side t-test

$|6.421| > t_{.05/2, 104-2} \longrightarrow \boxed{6.421 > 1.976}$

- At the 5% level, we would reject H_0 if $|t| > 1.976$. In this case, $6.421 > 1.660$, so we do reject H_0. More on back $\longrightarrow$

We justify the usage of a two-sided t-test because in the question, it specifies us to determine whether the 2 coefficients are "significently different from zero". With, we use a two-sided t-test to justify usage & find that it is significantly different from zero.

c. The coefficients have not changed in this example because the heteroskedasticity-robust is only accounting for standard errors. The coefficients have not been changed because of this, as the hypothesis test is not affected by this. Yes, the results are also more significant because the SE is smaller, meaning that the dist'n of the data is less spread out and closer along the regression line.

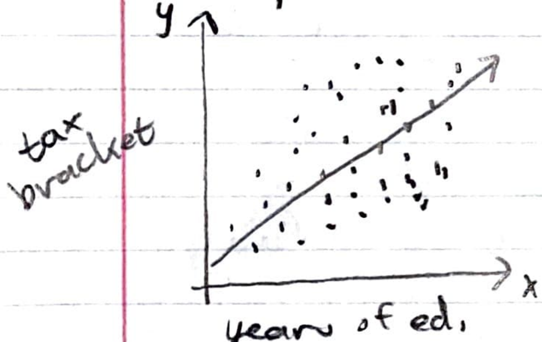
d. The Gauss-Markov Theorem for $\hat{\beta}_1$ consists of 4 assumptions that makeup the preface of how it accounts for distributed data:

1) $E(u|X=x)=0$
2) $(X_i, Y_i), i=1,\ldots,n$, are i.i.d.
3) Large Outliers are rare ($E(y^4)<\infty$, $E(x^4)<\infty$)
4) u is homoskedastic

- OLS is the best with or without assumption 4 holding true b/c it simplifies math calculations and you can prove strong results holding that these assumptions are true. If all 4 hold true, then the OLS est. is the Best (most eff.) Linear conditionally Unbiased Estimator (BLUE).
- With this, it is likely that the error terms would be heteroskedastic in this case, as they did change when running the heteroskedastic-robust command in the statistical software.

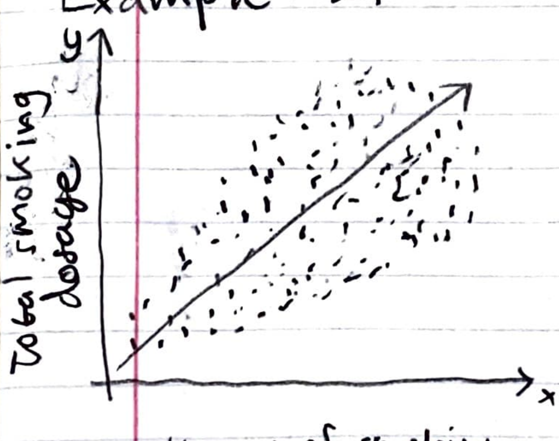
5. The advantages of using heteroskedasticity-robust SE over SE calculated under the assumption of Homoskedasticity is that it lessens the spread of the distributed data, making your results more significant without changing the coefficients. This allows for $\hat{\beta}_1$ to not have to consist of the 4 assumptions that makeup the Gauss-Markov Theorem and its restrictive 4^{th} assumption. Instead it can use OLS properties to exhibit the line of best fit for a regression function that has data exhibit heteroskedastic tendencies.

Example #1:



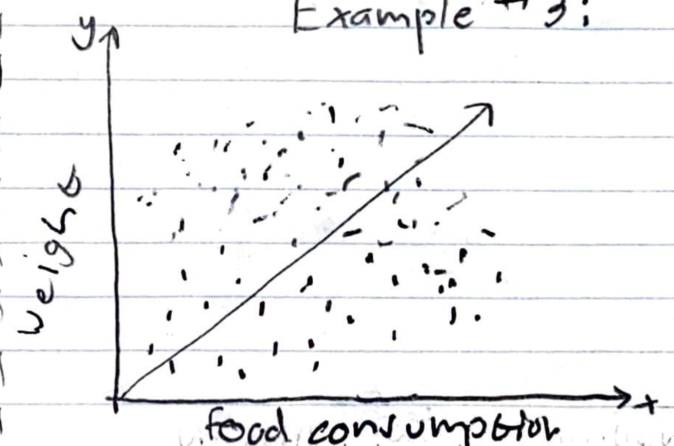
- The higher the years of education one person has completed, the more variability that is found with which tax bracket they fall in.

Example #2:



- The more years of smoking, the more variability in packs you smoke.

Example #3:



- more food you eat, the more variability found in weight of individuals.

7. b. $\hat{\beta}_1 = -253.2284$

iii. 95% CI for $\hat{\beta}_1 = \{\hat{\beta}_1 \pm 1.96 \cdot SE(\hat{\beta}_1)\}$

$\hat{\beta}_1 = (-253.2284 \pm 1.96 \cdot (24.04206))$ ← non-smoker

CI = (-206.106, -300.351)

$\hat{\beta}_1 = (-253.2284 \pm 1.96(11.8890))$ ← smoker

CI = (-229.92596, -276.53084)

c.

iii. 95% CI for $\hat{\beta}_1 = \{\hat{\beta}_1 \pm 1.96 \cdot SE(\hat{\beta}_1)\}$

$\hat{\beta}_1 = (-253.2284 \pm 1.96(26.95149))$

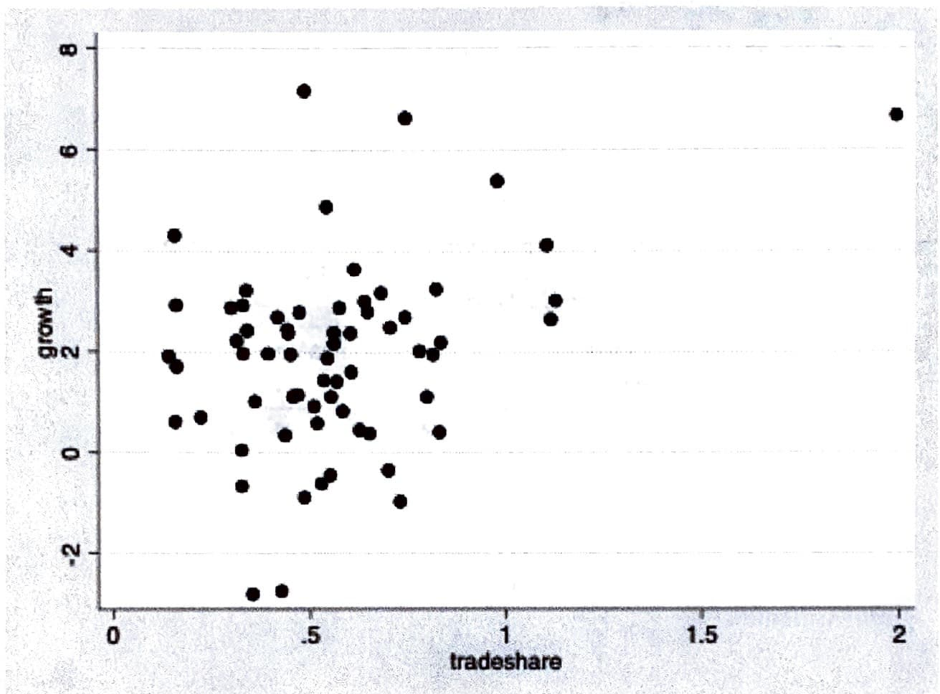
CI = (-200.4035, -306.0533)

d. No, I do not think smoking is uncorrelated w/ other factors that cause low birthweight. For instance, there is a correlation between those that smoke and those that drink. Drinking alcohol while pregnant also has damaging effects on birthweight, as a mom that drinks while pregnant would have unhealthy effects (i.e. low birthweight) on their baby. Therefore, I do not think smoking is uncorrelated w/ other factors, as those that smoke may also exhibit other unhealthy tendencies that, in this case, can affect birthweight.

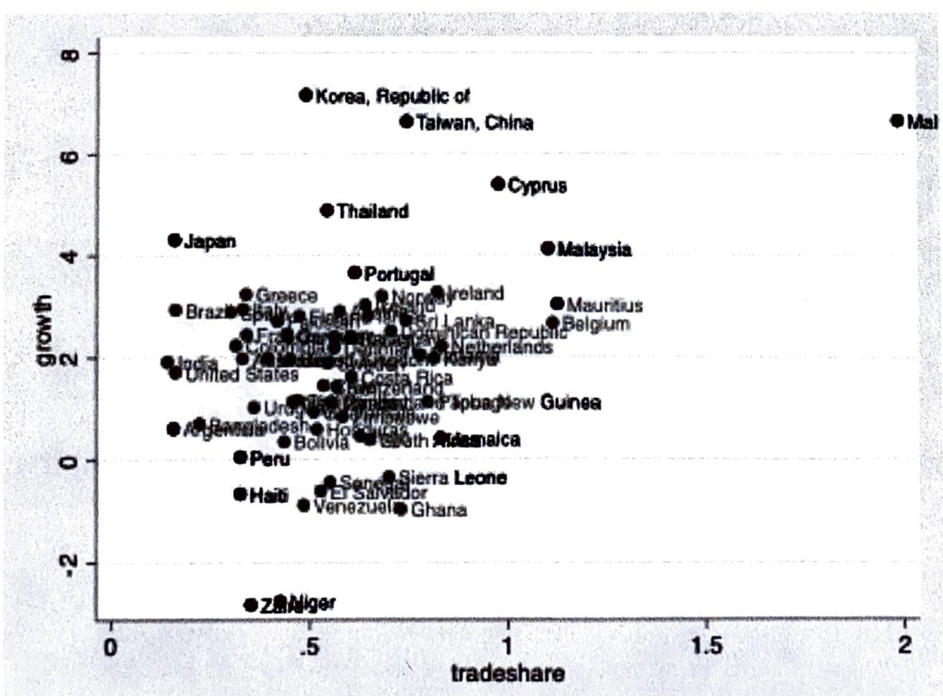
Daniel Lesh
ECON 2311Q
10/30/2020

Graphs for HW #2

6.

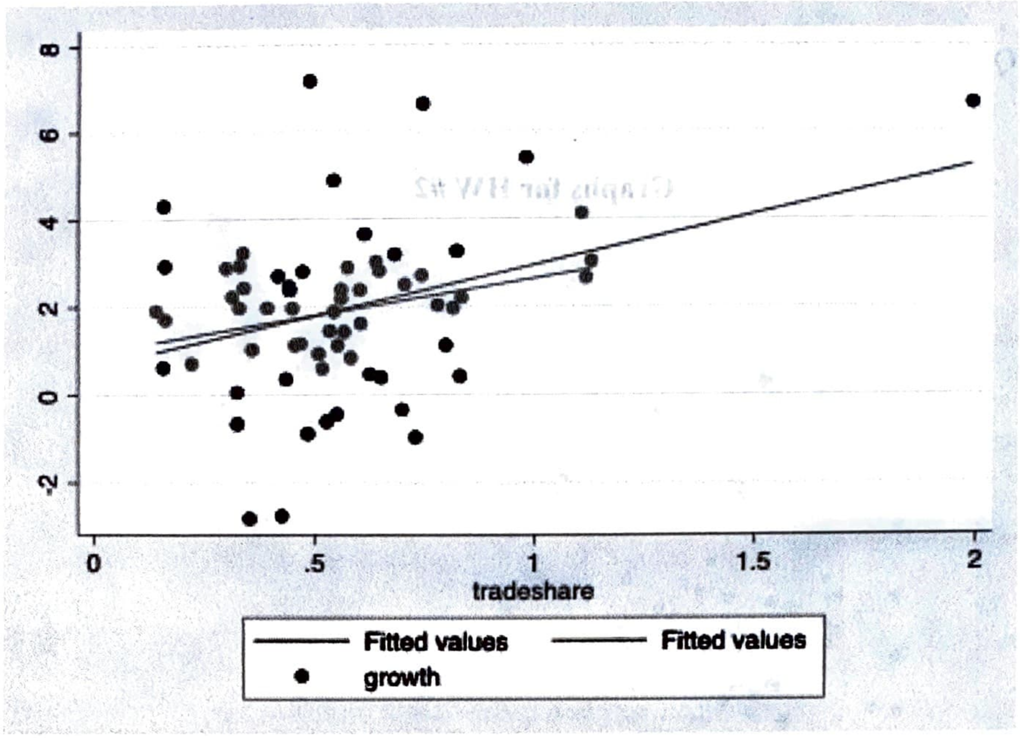


a.



b.

e.



8
6
4
2
0
-2
0
.5
1
1.5
2
tradeshare
Fitted values
Fitted values
growth