Deedline: 20:27 Stolyarov BEC193

1. $$y = \beta_0 + \beta_1 \cdot S + \beta_2 \cdot EDUC + \beta_3 \cdot [S \cdot EDUC]$$

So firstly ~~I wrote~~ it must be some function $y(EDUC)$, that's why I decided to choose y and $EDUC$. Since my goal is to estimate ~~this~~ this regression for different genders, I added dummy ~~S to inte~~ (it'll be a part of intercept β_0, if $S=1$) (there are only 2 genders, I hope) also sex can influence earnings in combo with educ, so I put variable SEDUC (product of S and EDUC), to reach my goal. ~~Since~~ (I should've start with this)

As I use ~~dummies th~~ 2 new coefs (with dummy, it's available to provide Chow test; we can do it like this:

Take two regressions (for sex = 1 and sex = 0) separately estimate it, count RSS for them, then count ~~unrestrict~~ pooled regression and find F-stat - k_{pooled}

$$F(k, n-2k) = \frac{RSS_p - [RSS_1 + RSS_2](k)}{(RSS_1 + RSS_2)/(n-2k)}$$ then

compare it with F_{crit} if ~~then~~ $F < F_{crit}$, then ~~I~~ I conclude that there is no significant difference.

2. , $n-k = 100-3 => 97$ dof $t_{crit} \approx 1{,}98$

$\log X_1$: ~~$t = \frac{3,1}{0,3}$~~ ?

Conf int:

$3{,}1 - 0{,}3 \cdot 1{,}98 \leq \beta_1 \leq 3{,}1 + 0{,}3 \cdot 1{,}98$

$2{,}506 \leq \beta_1 \leq 3{,}694$

X_2, conf. int:

interpret. $0{,}4 - 0{,}1 \cdot 1{,}98 \leq \beta_2 \leq 0{,}4 + 0{,}1 \cdot 1{,}98$

$0{,}202 \leq \beta_2 \leq 0{,}598$

as 0 is not in interval, then the ~~estimator~~ coef is significant at the level of 5%.

~~(if X increases by 3,1% the y increases by 3,1%)~~

1 model: intercept: $se < 0$? ~~has no plaus. meaning~~. If $\log X_1 = 0$ and $X_2 = 0$, then $\log y = 0{,}5$

$3{,}1 \log X_1$: if X_1 increases by 1% then y increases by 3,1%

$0{,}4 X_2$: if ~~log~~ X_2 increases by 1 (additional unit) ~~log~~ y increases by 40%.

Model 2:

intercept: if ~~D=~~ all other factors are 0, (including D) then ~~log~~ $\hat{y} = -0{,}5$

$3{,}1 D$: when all factors are 0, but D = 1, then ~~log~~ y = -0,5 + 3. the difference when all factors are

$2X_1$: if D = 0, then increasing in X_1 by 1 causes increasing in y by 2.

$2X_1 + 2{,}1 X_1 \cdot D$: if D = 1, then increasing in X_1 by 1 causes increasing in y by 4,1

$3{,}1 X_2$: if X_2 increases by 1, then y increases by 3,1

3. $n = 300$

$GPA_i = \beta_0 + \beta_1 CLASS + \varepsilon_i$

1) In case of endogeneity problem estimators become inconsistent and it will have bias, so it'll be bad estimators.

2) Maybe, the distance from university is correlated with attendency of classes (it's hard to get to your classes when you live far from university, so ↑ distance ⇒ ↓ attendancy ⇒ ↓ talent (as reasercher thinks). So distance has negative correlation with talent, but won't be correlated with disturbance term, that's what we search for! In this case, it can be used like instrumental variable for talent.

4. $E = \beta_0 + \beta_1 X + \beta_2 Z + u$

we call it like this:

$RSS_1 = 200$ ($n_1 = 50$)

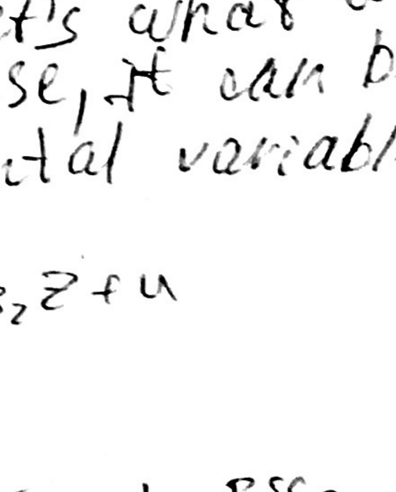
$RSS_2 = 250$ ($n_2 = 50$)

$R^2 = 0,3$

G-Q test: $F(47,47) = \frac{RSS_2}{RSS_1} = \frac{250}{200} = 1,25$

$F_{crit\ 47,47} = 1,6$ (50,50)

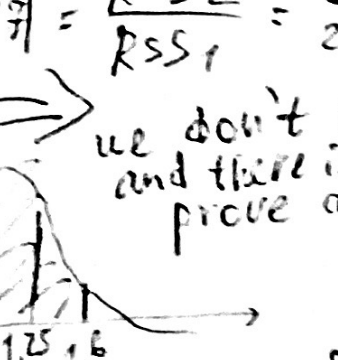
$1,25 < 1,6$ ⇒ we don't reject null-hyp and there is no significant prove about heteroskd



White test:

$nR^2 = 90 \sim \chi^2_{K-1=2}$

$\chi^2_{crit} = 5,99$



5,99 < 90 ⇒ we reject H_0 and say that H_0 is not

ways to fight heteroskedasticity:

1) make logarythmic models

2) divide our equation by σ_{u_i}:

$$\frac{y_i}{\sigma_{u_i}} = \beta_1 \frac{1}{\sigma_{u_i}} + \beta_2 \frac{x_i}{\sigma_{u_i}}$$

or by A_i: $\sigma_{u_i} = \lambda A_i$