final_carbone

Carlos Carbone

7/19/2021

Teoria

Punto 1

1. Un investigador sospecha que el modelo posee heteorscedasticidad. Para esto se deberá realizar el Test de white al modelo $y = \beta_0 + \beta_1 * x_1 + \beta_2 * x2 + u_i$

a. Es cierto que al tener heteroscedasticidad los estimadores son sesgados? Justifique.

Los estimadores con heteroscedasticidad no son sesgados es decir siguen siendo insesgados, son ineficientes. La heteroscedasticidad lo que hace es que la matriz de varianzas y covarianzas, a menos que se corrijan, no se puedan calcular con el MCO, eso implica que σ^2 no va a ser la misma para todas las observaciones y por lo tanto el calculo mediante MCO no es el optimo pusto que debería calcular la verdadera matriz de varianzas y covarianzas, pero no hay un sesgo, sino que los calculos de las estimaciones van a ser menos precisos.

b. Escriba el modelo utilizado en el test de white.

Una forma es hacerla similar a la Breusch-Pagan solo que la regresion inicial se construye multiplicando las variables cruzadas y elevandolas hasta la cantidad de variables que tenga, luego se hacen las mismas pruebas para validar cada uno de los u_i^2.

Para dos variables el algoritmo sería:

1. Estimar el modelo
$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + u$$
y obtener los residuales cuadrados para cada observacion (u_i^2)
2. Hacer la regresion

$$\hat{u}_i^2 = \delta_0 + \delta_1 * x_{1i} + \delta_2 * x_{2i} + \delta_3 * x_{1i}^2 + \delta_4 * x_{2i}^2 + \delta_5 * x_{1i} * x_{2i} + error$$

y conservar $R^2_{u_i^2}$ de la regresion.

3. Formar el estadístico F o el ML y calcular el p-value usando $F_{k,n-k-1}$ o χ^2_k para el estadisitico ML (multiplicadores de lagrange).

4. Si el p-value<α entonces RHN y por lo tanto no hay heteroscedasticidad. Tambien se puede usar el $ML = n * R^2_{u_i^2}$ y contrastarlo contra χ^2_k

Punto 2

Considere el siguiente modelo

$$w_i = \beta_0 + \beta_1 * Ed + \beta_2 * Ed^2 + ui$$

donde:

$$Var(u_i) = \sigma^2 * Ed_i^2$$

a. Que problema encuentra? Solucionole y describa paso a paso. Presente la nueva fórmula del modelo.

Lo que podemos observar es que no se cumple el supuesto de homoscedasticidad dado que la varianza no es igual, sino que depende de la edad al cualdrado. Una forma de solucionarlo sería dividiendo todo por la Edad al cuadrado de la siguiente forma:

dividimos todo por $\sigma_i = Ed_i^2$ lo que nos deja la siguiente ecuacion:

$$\frac{Y}{Ed_i^2} = \frac{\beta_0}{Ed_i^2} + \beta_1 * \frac{Ed}{Ed_i^2} + \frac{\beta_2 * Ed_i^2}{Ed_i^2} + \frac{u_i}{Ed_i^2}$$

Reordenando un poco los términos quedaría.

$$\frac{Y}{Ed_i^2} = \beta_2 + \frac{\beta_0}{Ed_i^2} + \beta_1 * \frac{Ed}{Ed_i^2} + \frac{u_i}{Ed_i^2}$$

b. En caso de no conocer $var(u_i)$ que se le ocurre hacer? Escriba en que consta la solucion.

Si no conozco σ_i^2 lo que puedo hacer es utilizar las varianzas de los errores estimados o sea, reemplazo las varianzas de los errores por los estimados.

$$var(\hat{\beta}) = (X^T * X)^{-1} * X^T * \hat{u}_i * \hat{u}_i^T * X * (X^T * X)^{-1}$$

Practica

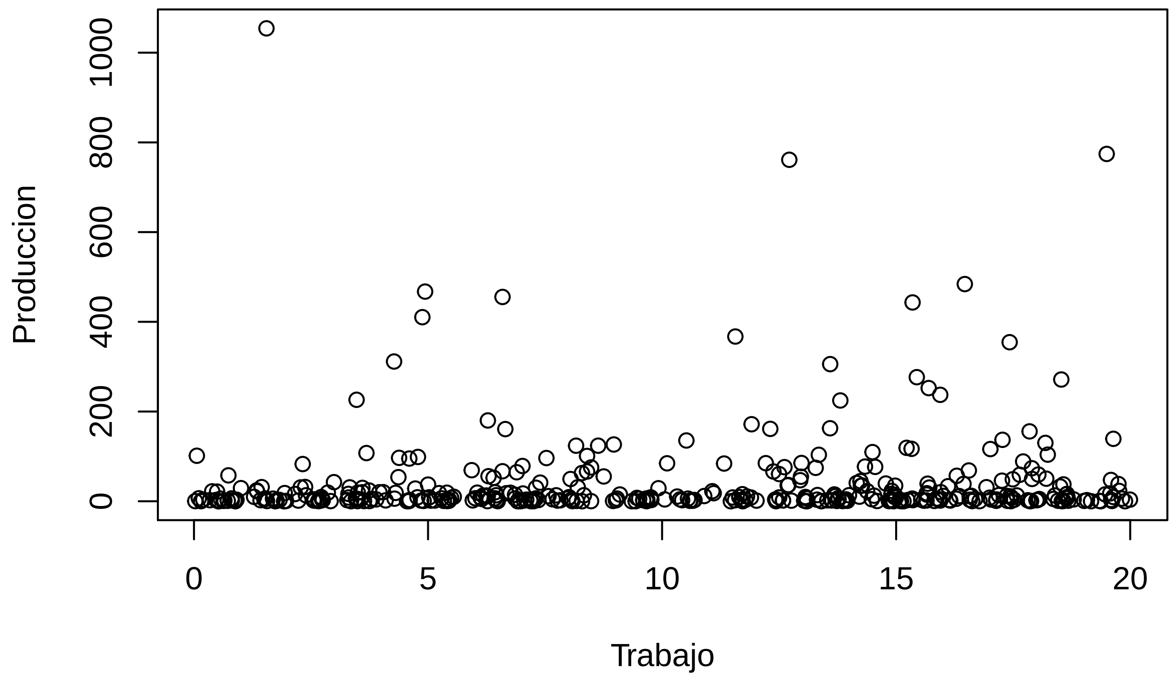
Punto 1

Imagine que usted est ainteresado en estimar la funcion de produccion de un pais. De acuerdo a la teoria microeconomica se sabe que la funcion Cobb Douglas es:

$$y = ak^{\alpha} * l^{\beta}$$

a. Grafique la relacion entre trabajo y produccion. Comente.

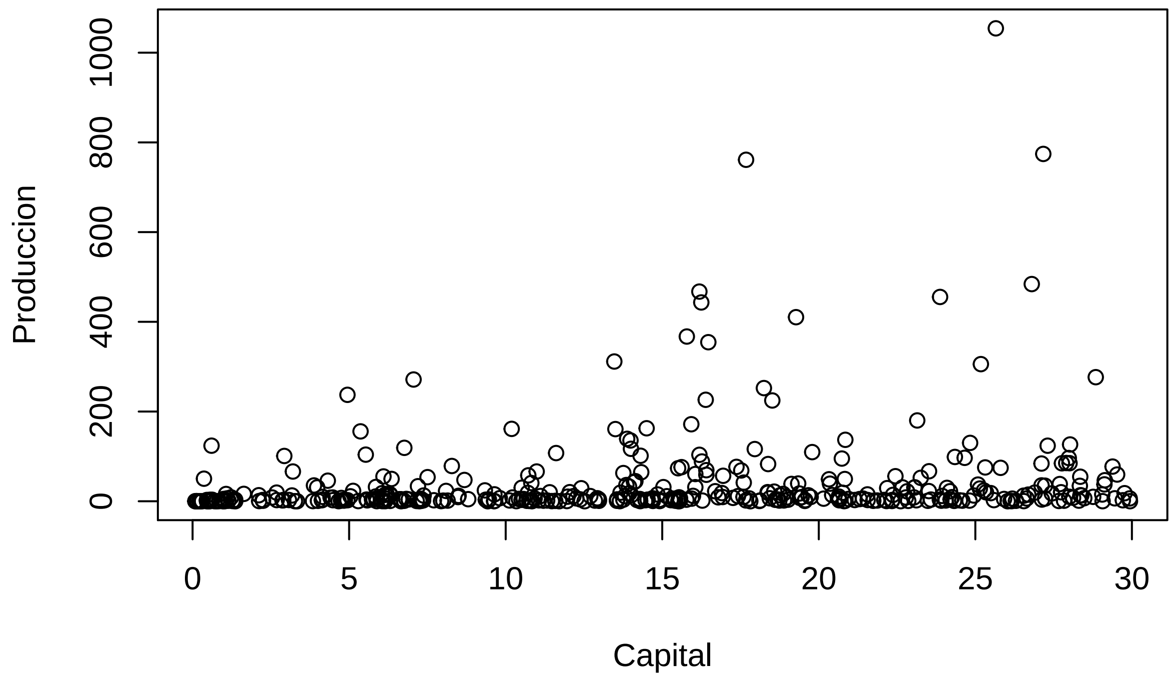
```
library(readxl)
datos_produccion=read_excel("Ejercicio1.xlsx")
plot(x=datos_produccion$l,y=datos_produccion$y,xlab = "Trabajo",ylab = "Produccion")
```



Al ser una función Cobb Douglas se esperaba ver un relación mas parecida a una raíz. Es posible que por la variablidad de algunos puntos la escala de grafico este quedando mal. Pero debería exisitr una relacion creciente entre trabajo y producción, es decir a medida que se incrementa el trabajo se incrementa la produccion, pero no de forma lineal sino decreciente.

b. Relacion entre capital y produccion

```
plot(x=datos_produccion$k,y=datos_produccion$y,xlab = "Capital",ylab = "Produccion")
```



Aqui sucede lo mismo que antes pero lo que deberíamos ver es que medida que se incrementa el capital, se incrementa la producción, pero no linealmente sino mas bien con una forma de funcion de raíz.

c. Estime e interprete cada uno de los coeficientes.

A cada uno de los modelos le vamos a aplicar la función logarítmica para que quede linelizado

Modelo 1

$$Y = aK^{\alpha}$$

$$ln(Y) = ln(aK^{\alpha})$$

$$ln(Y) = ln(a) + \alpha * ln(K)$$

Esto mismo vale para los otros dos modelos y nos quedaría

Modelo 2

$$ln(Y) = ln(a) + \beta * ln(l)$$

Modelo 3

$$ln(Y) = ln(a) + \alpha * ln(K) + \beta * ln(l)$$

Ahora calculamos las regresiones para los 3 modelos.

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
datos_produccion$ln_y<-log(datos_produccion$y)
datos_produccion$ln_l<-log(datos_produccion$l)
datos_produccion$ln_k<-log(datos_produccion$k)


regresor1_log_log<-lm(datos_produccion$ln_y~ ln_k,data = datos_produccion)

regresor2_log_log<-lm(datos_produccion$ln_y~ ln_l,data = datos_produccion)

regresor3_log_log<-lm(datos_produccion$ln_y~ln_l+ln_k,data = datos_produccion)

stargazer(regresor1_log_log,regresor2_log_log,regresor3_log_log,type = "text")
```

```
## 
## ==========================================================================================
##                                              Dependent variable:                          
##                     ----------------------------------------------------------------------
##                                                     ln_y                                  
##                              (1)                     (2)                     (3)          
## ------------------------------------------------------------------------------------------
## ln_k                      0.812***                                        0.842***        
##                            (0.094)                                         (0.094)        
##                                                                                           
## ln_l                                              0.269**                 0.352***        
##                                                   (0.110)                  (0.101)        
##                                                                                           
## Constant                   -0.050                 1.312***                -0.833**        
##                            (0.244)                (0.248)                  (0.329)        
##                                                                                           
## ------------------------------------------------------------------------------------------
## Observations                400                     400                     400           
## R2                         0.157                   0.015                   0.182          
## Adjusted R2                0.154                   0.012                   0.178          
## Residual Std. Error    2.020 (df = 398)        2.184 (df = 398)        1.993 (df = 397)   
## F Statistic         73.859*** (df = 1; 398) 6.003** (df = 1; 398) 44.087*** (df = 2; 397)
## ==========================================================================================
## Note:                                                        *p<0.1; **p<0.05; ***p<0.01
```

Interpretacion de los coeficientes

Modelo 1

Como es un modelo log-Log (Lo usamos para medir la elasticidad. Al aumentar 1% X Y se modifica en regresor %) Al incrementarse 1% el capital se incrementa 0.812% la produccion. El ln_k es decir el regresor es significativo al 1%, 5% y 10%. En este modelo no hay regresor para el trabajo. R es muy poco significativo y le intercepto es -.0050 no significativo y con justa razón dado que no hay productividad negativa.

Modelo 2

Al incrementarse 1% el trabajo se incrementa 0.269% la produccion. El ln_l es decir el regresor es significativo al 5% y 10% pero no al 1%. En este modelo no hay regresor para el capital R es muy poco significativo y el intercepto es 1.312 significativo al 1%, 5% y 10%.

Modelo 3

Aqui ln_k y ln_l son significativos al 1%, 5% y 10% el intercepto lo es solamente al 5% y al 10%. El modelo global es significativo al 1%, 5% y 10% y el R^2 es 0.178 muy bajo al igual que el R de los otros modelos por lo tanto explica muy poco la las estimaciones de la produccion las variables k yl

Test de significacia

a.

Hacemos un test de significancia para α que es el regresor que acompaña al capital a es distinto de 0.8partir del modelo 3.

$$\frac{\hat{\alpha} - 0.8}{ee(\hat{\alpha})}$$

```
k<- coef(summary(regresor3_log_log))[3,1]
stErrorK<-coef(summary(regresor3_log_log))[3,2]

tvalue<-((k-0.8)/stErrorK)
tvalue
```

```
## [1] 0.445829
```

El valor de t es pequeño asi que vamos a calcular el p-value para ver rechazamos o no.

```
## Aca puedo usar lower.tail=FALSE, pero para resptar la ecuacion no lo uso.
## tambien se podría usar pt, es decir una distribucion t-student, pero para seguir la teoría
# que indica que para una mustra grande ya se toman los valores de una normal se elige esta misma.
p_value<-1-pnorm(tvalue)
sprintf("p-value: %f, t-value: %f",p_value,tvalue)
```

```
## [1] "p-value: 0.327860, t-value: 0.445829"
```

```
ifelse(p_value<0.01,"significativo al 1%","no significativo al 1%")
```

```
## [1] "no significativo al 1%"
```

```
ifelse(p_value<0.05,"significativo al 5%","no significativo al 5%")
```

```
## [1] "no significativo al 5%"
```

```r
ifelse(p_value<0.10,"significativo al 10%","no significativo al 10%")
```

```
## [1] "no significativo al 10%"
```

Si el p-value< nivel significacia, entonceses significativo. Esto nos da que nuestro α no es significativo al 1% 5% ni 10%.

b. Vemos si $\alpha + \beta$ es estadísticamente distinto de 1.

Entonces $H_0 : \alpha + \beta = 1$ y $H_1 : \alpha + \beta \neq 1$.

Estimamos el nuevo t.

$$t = \frac{(\alpha + \beta) - 1}{ee(\alpha + \beta)}$$

Recordamos la notación $ee(b)$ es error estándar. Este se pude calcular como $\sqrt{var(\beta)}$. Recordando propiedades de la varianza tenemos que: $var(\alpha + \beta) = var(\alpha) + var(\beta) = 2 * cov(\alpha, \beta)$

Esto quedaria:

$$t = \frac{(\alpha + \beta) - 1}{\sqrt{var(\alpha) + var(\beta) + 2cov(\alpha, \beta)}}$$

```r
l<- coef(summary(regresor3_log_log))[2,1]
stErrorL<-coef(summary(regresor3_log_log))[2,2]

vcov(regresor3_log_log)
```

```
##             (Intercept)          ln_l          ln_k
## (Intercept)  0.10793550 -0.0225313146 -0.0222977509
## ln_l        -0.02253131  0.0101250353  0.0008614573
## ln_k        -0.02229775  0.0008614573  0.0087503136
```

```r
t_no_restringido<-((k+l)-1)/sqrt(((stErrorK^2+stErrorL^2)+2*(-6.5425^-12)))


p_value_no_restringido<-1-pnorm(t_no_restringido)
sprintf("p-value: %f y el t_no_restringido: %f",p_value_no_restringido,t_no_restringido)
```

```
## [1] "p-value: 0.079377 y el t_no_restringido: 1.409272"
```

```r
ifelse(p_value_no_restringido<0.01,"significativo al 1%","no significativo al 1%")
```

```
## [1] "no significativo al 1%"
```

```r
ifelse(p_value_no_restringido<0.05,"significativo al 5%","no significativo al 5%")
```

```
## [1] "no significativo al 5%"
```

```
ifelse(p_value_no_restringido<0.10,"significativo al 10%","no significativo al 10%")
```

```
## [1] "significativo al 10%"
```

Esto nos da que $(\alpha + \beta)$ es estadísticamente disntito de 1 solo con un nivel de significacia del 10%.

Punto 2

El siguiente modelo

$$ln(wi) = \beta_0 + \beta_1 * Ed + \beta_2 * Mujer + u_i$$

El modelo es log-lin eso significa que al variar ED en una unidad Y se modifica en $\beta 2 * 100\%$

a.

Si β_1 es 0.021 signifca que al incrementarse en uno la edad el salario aumenta en 2.1% como el p-value<0.01 es significativo al 1%, 2% y 10%.

b. Si β_2 es -0.12 signifca el salario de las mujeres es menor en 12% como el p-value<0.065 es significativo al 10% pero no al 1% ni al 5$.
c. Afirmar que hay diferencia es correcto en cuanto se tome un nivel de significancia del 10% si tomamos uno mejor, es decir, si somos mas exactos en nuestros intervalos de confianza para predecir nos daria que NO es significativo, con lo cual tomando ya un nivel estandar del 5% nos da que no hay diferencia entre hombres y mujeres en cuanto al salario.

Se podría usar el siguiente modelo

$$ln(wi) = \beta_0 + \beta_1 * Ed + \beta_2 * Mujer + \beta_3 * Mujer * Ed + u_i$$