

ECONOMETRIA COD. 6089

SECONDO PROBLEM SET

SOLUZIONI

Esercizio 1

Considerate il modello

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1, \dots, 100$$

dove assumiamo che valgano tutte le assunzioni standard del modello di regressione lineare, tranne quella di omoschedasticità e

$$\text{var}(\varepsilon_i) = c \log x_i.$$

a) Costruite lo stimatore GLS di β per questo modello.

Costruiamo una matrice tale che:

$$P^{-1}\Omega P'^{-1} = I$$
$$P^{-1} = \frac{1}{\sqrt{c}} \begin{bmatrix} \frac{1}{\sqrt{\log x_1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\log x_N}} \end{bmatrix}$$

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$$

Possiamo scrivere questa soluzione considerando i Weigthed Least Squares (WLS) estimators.

Infatti,

$$\frac{y_i}{\sqrt{c \log x_i}} = \beta \frac{x_i}{\sqrt{c \log x_i}} + \frac{\varepsilon_i}{\sqrt{c \log x_i}}$$

b) Costruite la sua matrice di varianza-covarianza.

$$\text{var}(\hat{\beta}_{GLS}) = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\Omega\Omega^{-1}X (X'\Omega^{-1}X)^{-1} = (X'\Omega^{-1}X)^{-1}$$

c) Costruite il test di Breusch-Pagan per questo modello.

Vedi appunti

Esercizio 2

La seguente tabella riporta la regressione del salario, *wage*, sull'istruzione *educ*, sull'esperienza in livelli ed al quadrato e su una variabile dummy, *female* che assume valore uno in caso in cui la persona intervistata nel campione sia di sesso femminile. Sotto trovate anche le statistiche descrittive delle variabili della regressione:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 female + \beta_4 exper^2 + \varepsilon$$

```
reg wage educ exper female expersq
```

Source	SS	df	MS	Number of obs = 526		
Model	2506.95576	4	626.738939	F(4, 521) =	70.17	
Residual	4653.45854	521	8.93178222	Prob > F =	0.0000	
-----				R-squared =	0.3501	
Total	7160.41429	525	13.6388844	Adj R-squared =	0.3451	
-----				Root MSE =	2.9886	
wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.5562848	.0502875	11.06	0.000	.4574937	.6550759
exper	.2551276	.0348671	7.32	0.000	.1866302	.323625
female	-2.114035	.2625501	-8.05	0.000	-2.629822	-1.598248
expersq	-.0044396	.0007762	-5.72	0.000	-.0059645	-.0029148
_cons	-2.319204	.7388254	-3.14	0.002	-3.770647	-.8677608

```
sum wage educ exper expersq female
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	526	5.896103	3.693086	.53	24.98
educ	526	12.56274	2.769022	0	18
exper	526	17.01711	13.57216	1	51
expersq	526	473.4354	616.0448	1	2601
female	526	.4790875	.500038	0	1

a) Come interpretate il coefficiente associato ad *educ*?

In questo caso di regressione multipla, possiamo interpretare il coefficiente associato ad educ, come derivata marginale sul salario. Però nella sua interpretazione dobbiamo osservare la tabella delle statistiche descrittive. La variabile educ assume valori continui da 1 a 18, che indicano anni di istruzione, quindi all'aumentare degli anni di istruzione, il salario dell'individuo aumenta di 0.556 unità.

b) Qual è potrebbe essere lo scopo di usare una variabile dummy tipo *female*?

Lo scopo di usare una variabile dummy è quello di osservare se ci sono differenze nella regressione che dipendono da alcune caratteristiche particolari, in questo caso che la persona intervistata sia di sesso femminile. Quindi la dummy va ad aggiungersi o togliersi alla costante.

c) Come interpretate il coefficiente stimato per la variabile? Scrivete i valori predetti di *wage* nel caso in cui la dummy sia uguale a 0 e nel caso sia uguale ad 1?

In questo caso se l'individuo è femmina, il suo salario diminuisce di -2.11 unità, quindi ha un impatto negativo!

Nel caso la dummy assuma valore 0:

$$\widehat{wage} = -2.319 + 0.556 * educ + 0.255 * \exp er - 0.0044 * \exp er^2$$

Nel caso la dummy assuma valore 1:

$$\widehat{wage} = -(2.319 + 2.11) + 0.556 * educ + 0.255 * \exp er - 0.0044 * \exp er^2$$

L'introduzione della variabile dummy permette di evidenziare un differenziale salariale tra uomini e donne, negativo per le donne, a parità di istruzione ed esperienza

d) Se nel dataset ci fosse stata anche la dummy *male*, che assume valori pari ad 1 in caso di intervistato di sesso maschile, ci sarebbero stati problemi nell'includere questa dummy nella regressione precedente?

Nel caso avessimo inserito sia la dummy *male* che *female*, si sarebbero verificato il problema di multicollinearità, chiamato nello specifico *DUMMY VARIABLE TRAP*.

e) Come interpretate i coefficienti di *exper* ed *esper²*?

Partendo dall'effetto di *exper* su *wage*,

$$\frac{\partial wage}{\partial \exp er} = \beta_2 + 2\beta_4 \exp er$$

Osservando i segni di β_2 e β_4 , si può dire che l'effetto dell'esperienza prima è crescente e poi è decrescente, per trovare il valore di *exper* che permette questo cambiamento, poniamo

$$\begin{aligned} \widehat{\beta}_2 + 2\widehat{\beta}_4 \exp er &= 0 \\ 0.255 - 2 * 0.0044 * \exp er &= 0 \\ \exp er &= \frac{0.255}{0.0088} = 28.977 \end{aligned}$$

Notate che il punto in cui cambia da effetto crescente ad effetto decrescente è pari a 28.977 e se osserviamo il range della variabile esperienza, notiamo che è tra 1 e 51, quindi sembra che dopo 29 circa anni di esperienza c'è questo cambiamento, da effetto totale positivo ad effetto totale negativo.

Esercizio 3

```
reg colgpa hsperc sat
```

Source	SS	df	MS			
Model	490.606706	2	245.303353	Number of obs =	4137	
Residual	1303.58897	4134	.315333567	F(2, 4134) =	777.92	
Total	1794.19567	4136	.433799728	Prob > F =	0.0000	
				R-squared =	0.2734	
				Adj R-squared =	0.2731	
				Root MSE =	.56155	

colgpa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hsperc	-.0135192	.0005495	-24.60	0.000	-.0145965	-.012442
sat	.0014762	.0000653	22.60	0.000	.0013482	.0016043
_cons	1.391757	.0715424	19.45	0.000	1.251495	1.532018

Osservate l'output della regressione:

$$\text{colgpa} = \beta_0 + \beta_1 \text{hsperc} + \beta_2 \text{sat} + \varepsilon$$

dove *colgpa* misura su una scala da 0 a 4 i voti degli studenti di un college americano, *hsperc* esprime in termini percentili la promozione alla scuola superiore (*hsperc*=5 significa che lo studente era nel 5% dei bravi della sua classe) e *sat* rappresenta il punteggio di un test attitudinale per l'ingresso in università, in cui vengono testate le abilità matematiche e verbali.

a) Ecco l'output del test di Ramsey-Reset. Spiegate il suo utilizzo e commentate il risultato ottenuto.

```
estat ovtest
```

```
Ramsey RESET test using powers of the fitted values of colgpa
Ho: model has no omitted variables
      F(3, 4131) =      40.74
      Prob > F =      0.0000
```

Questo output si riferisce al test di Ramsey-RESET. Questo test è utile per constatare la presenza di mispezificazioni nel modello. Ramsey-RESET test svolge una regressione aumentata che include i regressori originali, le potenze dei valori predetti della regressione originale (vengono inserite le prime 3 potenze di \hat{y} (la seconda, la terza e la quarta potenza), da cui dipendono dai gradi di libertà del numeratore della statistica F , invece al denominatore contiamo $N - k - \text{num.potenze} = 4137 - 3 - 3 = 4131$). Sotto ipotesi nulla di non mispezificazione, i coefficienti di questi regressori addizionali sono pari a zero. In questo caso, il

pvalue è inferiore a qualsiasi livello di significatività, quindi rifiutiamo l'ipotesi nulla, di conseguenza ci sono problemi di misspecificazione.

b) Introducendo la dummy *female* (come visto nell'esercizio precedente), cosa possiamo capire dalla stima?

```
reg colgpa hsperc sat female
```

Source	SS	df	MS			
Model	511.874076	3	170.624692	Number of obs =	4137	
Residual	1282.3216	4133	.310264118	F(3, 4133) =	549.93	
Total	1794.19567	4136	.433799728	Prob > F =	0.0000	
				R-squared =	0.2853	
				Adj R-squared =	0.2848	
				Root MSE =	.55701	

colgpa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hsperc	-.012594	.0005564	-22.64	0.000	-.0136848	-.0115032
sat	.001589	.0000662	24.00	0.000	.0014593	.0017188
female	.1489008	.0179848	8.28	0.000	.1136409	.1841607
_cons	1.190771	.0750024	15.88	0.000	1.043726	1.337816

Come nell'esercizio precedente, il ruolo della dummy female è un ruolo additivo alla costante, quindi quando l'individuo è femmina,

$$\widehat{\text{col gpa}} = (1.19 + 0.14) - 0.012 * \text{hsperc} + 0.001 * \text{sat}$$

La sua performance universitaria sarà maggiore di 0.14 punti!

c) Se avessimo generato una variabile interagita tra *sat* e *female* e svolto la seguente regressione, come avremmo interpretato la variabile dummy? Scrivete il valore predetto di *colgpa* quando *female* = 1 e quando *female* = 0.

```
g satf=sat*female
```

```
. reg colgpa hspec sat satf
```

Source	SS	df	MS			
Model	512.005618	3	170.668539	Number of obs =	4137	
Residual	1282.19006	4133	.31023229	F(3, 4133) =	550.13	
Total	1794.19567	4136	.433799728	Prob > F =	0.0000	
				R-squared =	0.2854	
				Adj R-squared =	0.2848	
				Root MSE =	.55699	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hsperc	-.0126411	.0005552	-22.77	0.000	-.0137295	-.0115527
sat	.0015305	.0000651	23.51	0.000	.0014028	.0016581
satf	.0001443	.0000174	8.31	0.000	.0001103	.0001784
_cons	1.25364	.072884	17.20	0.000	1.110748	1.396532

In questo caso la variabile dummy assume un ruolo moltiplicativo, nel senso che non cambia l'intercetta dei due modelli con female = 1 e female = 0, ma cambia l'effetto della variabile sat su colgpa:

con female = 0

$$\widehat{\text{colgpa}} = 1.25 - 0.012 * \text{hsperc} + 0.00153 * \text{sat}$$

con female = 1

$$\widehat{\text{colgpa}} = 1.25 - 0.012 * \text{hsperc} + (0.00153 + 0.0001) * \text{sat}$$

Anche in questo caso l'impatto della variabile female è positivo nella stima!

Esercizio 4

Consideriamo il seguente modello:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

in formato matriciale:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

Ma il vero modello è

$$y = \beta_0 + \beta_1 x + \beta_2 z + u$$

in formato matriciale:

$$\mathbf{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \mathbf{U} \quad (2)$$

a) Se stimiamo il modello 1, invece del modello 2, in che problema incorriamo. Lo stimatore OLS sarà ancora non distorto?

In questo caso, incorriamo nel problema di variabile omessa, lo stimatore OLS sarà ora distorto:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X'X)^{-1} X'y = (\text{sostituisco il vero modello}) = \\ &= (X'X)^{-1} X'(X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \mathbf{U}) = \\ &= \boldsymbol{\beta} + (X'X)^{-1} X'Z\boldsymbol{\gamma} + (X'X)^{-1} X'\mathbf{U} \\ E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + (X'X)^{-1} X'Z\boldsymbol{\gamma} \text{ se } X'Z \neq 0 \end{aligned}$$

b) Se il modello vero fosse stato il modello 1 e quello stimato il modello 2, in che problema incorriamo. Lo stimatore OLS di $\hat{\boldsymbol{\beta}}$ sarà ancora non distorto?

In questo caso, incorriamo nel problema di variabile irrilevante, lo stimatore OLS è ancora non distorto, invece avrà varianza maggiore:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X'M_zX)^{-1} X'M_zy = (\text{sostituisco il vero modello}) = \\ &= (X'M_zX)^{-1} X'M_z(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \\ &= (X'M_zX)^{-1} X'M_zX\boldsymbol{\beta} + (X'M_zX)^{-1} X'M_z\boldsymbol{\varepsilon} \text{ dato } M_z\boldsymbol{\varepsilon} = \mathbf{0} \\ E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} \end{aligned}$$

Invece lo stimatore $\hat{\boldsymbol{\gamma}}$ sarà pari a zero.

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= (z'M_xz)^{-1} z'M_xy = \\ &= (z'M_xz)^{-1} z'M_x(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{0} \text{ dato } M_z\boldsymbol{\varepsilon} = \mathbf{0} \text{ } M_xX = 0 \end{aligned}$$

Esercizio 5

Il seguente output riporta il risultato della regressione:

$$lprice = \beta_0 + \beta_1 lproptax + \varepsilon$$

dove il logaritmo del prezzo di un appartamento, $lprice$, è regredito su il logaritmo della tassa di proprietà del quartiere dove è ubicato l'appartamento, $lproptax$

```
reg lprice lproptax
```

Source	SS	df	MS	Number of obs =	506
Model	26.0826748	1	26.0826748	F(1, 504) =	224.71
Residual	58.4995961	504	.116070627	Prob > F =	0.0000
Total	84.5822709	505	.167489645	R-squared =	0.3084
				Adj R-squared =	0.3070
				Root MSE =	.34069

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lproptax	-.5733679	.0382489	-14.99	0.000	-.6485147 - .4982211
_cons	13.34193	.2273745	58.68	0.000	12.89522 13.78865

```
. sum lprice lproptax
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lprice	506	9.941057	.409255	8.517193	10.8198
lproptax	506	5.931405	.3963666	5.231109	6.566672

```
sum price proptax
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	506	22511.51	9208.856	5000	50001
proptax	506	40.82372	16.85371	18.7	71.1

a) Commentate l'effetto di $lproptax$ su $lprice$.

In questo caso abbiamo di fronte un modello logaritmico con sia regressori che variabile dipendente espressi in logaritmi. Quindi β_1 rappresenta non l'effetto marginale, ma l'elasticità, ovvero se la tassa di proprietà del quartiere varia dell'1%, il prezzo in logaritmi della casa varia di -0.57%

$$\beta_1 = \frac{\partial lprice}{\partial lproptax} = \frac{\partial lprice \frac{dprice}{dprice}}{\partial lproptax \frac{dproptax}{dproptax}} = \frac{\frac{1}{price} dprice}{\frac{1}{proptax} dproptax} = \epsilon_{yx}$$

b) Come potete ricavare l'effetto di $proptax$ su $price$ a partire dai risultati riportati?

In questo caso ci riferiamo all'effetto marginale di $proptax$ su $price$, considerando costanti le altre variabili esplicative:

$$\beta_1 = \frac{\partial price}{\partial proptax} \frac{proptax}{price} = \epsilon_{yx}$$

$$\text{Effetto Marginale} = \frac{\partial \text{price}}{\partial \text{proptax}} = \beta_1 \frac{\text{price}}{\text{proptax}} = \text{nel punto medio} = \beta_1 \frac{\overline{\text{price}}}{\overline{\text{proptax}}}$$

$$\text{Quindi } \frac{\partial \text{price}}{\partial \text{proptax}} = -0.57 * \frac{22511.51}{40.82372} = -314.32$$

c) Qual è l'utilità di considerare variabili espresse in logaritmi?

Il logaritmo è una trasformazione monotona di una variabile, spesso in economia vengono usati per avere unità di misura comparabili, per costruire tassi di crescita e quindi per motivi di convenienza e pratica numerica. Come in questo caso, notiamo dalla tabella delle statistiche descrittive che l'uso dei logaritmi ci permette di avere numeri piccoli e se non conosciamo le unità di misura delle variabili, tale trasformazione ci permette di compararle meglio.

Esercizio 6

Considerate la stessa regressione di prima, l'unica differenza è che il regressore e la variabile dipendente sono in livelli e non in logaritmi.

$$\text{price} = \beta_0 + \beta_1 \text{proptax} + \varepsilon$$

reg price proptax

Source	SS	df	MS			
Model	9.3433e+09	1	9.3433e+09	Number of obs =	506	
Residual	3.3482e+10	504	66433096.2	F(1, 504) =	140.64	
Total	4.2826e+10	505	84803032	Prob > F =	0.0000	
				R-squared =	0.2182	
				Adj R-squared =	0.2166	
				Root MSE =	8150.7	

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
proptax	-255.2159	21.52043	-11.86	0.000	-297.4967	-212.9351
_cons	32930.37	950.3317	34.65	0.000	31063.27	34797.47

a) Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

In questo caso la pendenza è rappresentata dal coefficiente stimato β_1 , quindi è pari a -255.2159. Invece se volessimo calcolare l'elasticità, partiamo dal risultato del precedente esercizio:

$$\frac{\partial \text{price}}{\partial \text{proptax}} \frac{\text{proptax}}{\text{price}} = \varepsilon_{yx}$$

$$\beta_1 = \frac{\partial \text{price}}{\partial \text{proptax}}$$

$$\varepsilon_{yx} = \frac{\partial \text{price}}{\partial \text{proptax}} \frac{\text{proptax}}{\text{price}} = \beta_1 \frac{\overline{\text{proptax}}}{\overline{\text{price}}}$$

$$\text{Quindi elasticità} = -255.2159 * \frac{40.82372}{22511.51} = -0.46282$$

In STATA c'è la possibilità di usare dopo la regressione il comando `mfx, eyex` per calcolare il valore dell'elasticità:

`mfx, eyex`

Elasticities after regress

y = Fitted values (predict)
= 22511.51

variable	ey/ex	Std. Err.	z	P> z	[95% C.I.]	X
proptax	-.4628238	.03973	-11.65	0.000	-.540695 -.384952	40.8237

b) Ipotizzate di avere la variabile dipendente espressa in logaritmi e la variabile indipendente espressa in livelli:

$$lprice = \beta_0 + \beta_1 proptax + \varepsilon$$

`reg lprice proptax`

Source	SS	df	MS	Number of obs =	506
Model	26.493825	1	26.493825	F(1, 504) =	229.87
Residual	58.0884459	504	.115254853	Prob > F =	0.0000
Total	84.5822709	505	.167489645	R-squared =	0.3132
				Adj R-squared =	0.3119
				Root MSE =	.33949

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
proptax	-.0135904	.0008964	-15.16	0.000	-.0153515 -.0118293
_cons	10.49587	.0395834	265.16	0.000	10.4181 10.57364

Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

In questo caso, abbiamo un modello log-lineare e quindi il coefficiente riportato nella tabella rappresenta una semi-elasticità, $\beta_1 = -0.013$.

Per trovare il valore della pendenza dobbiamo applicare la seguente formula:

$$\beta_1 = \frac{\partial lprice}{\partial proptax} = \frac{\partial lprice \frac{dprice}{dprice}}{\partial proptax} = \frac{1}{price} \frac{dprice}{dproptax}$$

$$ef\ fetto\ marginale = \frac{\partial price}{\partial proptax} = \beta_1 price = \beta_1 \overline{price}$$

pendenza = $-0.0135904 * 22511.51 = -305.94$

Invece, l'elasticità si ricava usando la seguente formula:

$$\beta_1 = \frac{\partial \text{price}}{\partial \text{proptax}} = \text{semi-elasticità}$$

$$\beta_1 * \text{proptax} = \frac{\frac{1}{\text{price}} d\text{price}}{\frac{1}{\text{proptax}} d\text{proptax}} = \varepsilon_{yx}$$

$$\varepsilon_{yx} = \beta_1 * \overline{\text{proptax}}$$

elasticità = $-0.0135904 * 40.82372 = -0.55481$

c) Ipotizzate di avere ora la variabile dipendente in livelli e la variabile indipendente in logaritmi:

$$\text{price} = \beta_0 + \beta_1 \ln \text{proptax} + \varepsilon$$

reg price lnproptax

Source	SS	df	MS			
Model	9.5881e+09	1	9.5881e+09	Number of obs =	506	
Residual	3.3237e+10	504	65947363.2	F(1, 504) =	145.39	
Total	4.2826e+10	505	84803032	Prob > F =	0.0000	
				R-squared =	0.2239	
				Adj R-squared =	0.2223	
				Root MSE =	8120.8	

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnproptax	-10993.16	911.7085	-12.06	0.000	-12784.38	-9201.946
_cons	87716.41	5419.75	16.18	0.000	77068.33	98364.5

Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

In questo caso ci troviamo di fronte ad una regressione lineare-logaritmica, è il coefficiente stimato rappresenta la semielasticità, $\beta_1 = -0.13$.

Per trovare il valore della pendenza dobbiamo applicare la seguente formula:

$$\beta_1 = \frac{\partial \text{price}}{\partial \ln \text{proptax}} = \frac{\partial \text{price}}{\partial \text{proptax}} \frac{d\text{proptax}}{d\text{proptax}} = \frac{\partial \text{price}}{\frac{1}{\text{proptax}} d\text{proptax}}$$

$$\text{ef fetto marginale} = \frac{\partial \text{price}}{\partial \text{proptax}} = \beta_1 \frac{1}{\text{proptax}}$$

pendenza = $-10993.16 * \frac{1}{40.82372} = -269.28$

Invece, l'elasticità si ricava usando la seguente formula:

$$\beta_1 * \frac{1}{price} = \frac{\frac{1}{price} \partial price}{\frac{1}{proptax} dproptax} = \varepsilon_{yx}$$

$$\varepsilon_{yx} = \beta_1 * \frac{1}{price}$$

$$elasticità = -10993.16 * \frac{1}{22511.51} = -0.48834$$

Esercizio 7

Consideriamo il seguente modello:

$$C_i = c_0 + c_1 Y d_i + \varepsilon_i$$

C =consumo; c_0 = consumo autonomo; c_1 =propensione marginale del consumo al reddito disponibile; Yd =reddito disponibile.

Sono stati intervistati $N=50$ consumatori, di cui $n_1=20$ consumatori europei ed $n_2=30$ consumatori americani.

a) Scrivete un modello in cui si tengano conto delle differenze tra i due gruppi di consumatori (sia per il consumo autonomo che per la propensione marginale del consumo al reddito), usando una variabile dummy. Usate il parametro δ come coefficiente nel caso dell'interazione con la variabile dummy.

Consideriamo una variabile:

$$D = 0 \text{ nel caso di consumatori europei}$$

$$D = 1 \text{ nel caso di consumatori americani}$$

Il modello completo usando la variabile dummy è:

$$C_i = c_0 + c_1 Y d_i + \delta_0 D_i + \delta_1 D_i Y d_i + \varepsilon_i$$

$$i = 1, \dots, N$$

b) Scrivete il modello nel caso $D = 0$

$$C_i = c_0 + c_1 Y d_i + \varepsilon_i$$

$$i = 1, \dots, n_1$$

c) Scrivete il modello nel caso $D = 1$

$$C_i = (c_0 + \delta_0) + (c_1 + \delta_1) * Y d_i + \varepsilon_i$$

$$i = n_1 + 1, \dots, N \text{ dove } n_2 = N - n_1$$

d) Scrivete la statistica del Chow test nel caso in cui testiamo l'ipotesi nulla:

$$H_0 : \delta_0 = \delta_1 = 0$$

ovvero che non ci sia un cambiamento tra i due gruppi di consumatori.

La statistica F del test di Chow é:

$$F = \frac{(RSS_R - RSS_M)/j}{RSS_M/(N - 2k)} \sim F(j, N - 2k)$$

dove RSS_R = somma del quadrato dei residui del modello ristretto sotto ipotesi nulla.

RSS_M = somma del quadrato dei residui del modello completo

e) In questo caso abbiamo i seguenti valori nella stima del modello completo e ristretto:

MODELLO COMPLETO

$$\begin{aligned}\hat{C} &= 8 + 0.8 * Yd \\ i &= 1, \dots, n_1\end{aligned}$$

con $RSS = 200$

$$\begin{aligned}\hat{C} &= 20 + 0.2 * Yd \\ i &= n_1 + 1, \dots, N\end{aligned}$$

con $RSS = 150$

con RSS totale del modello completo = $200 + 150 = 350$

MODELLO RISTRETTO

$$\hat{C} = 10 + 0.4 * Yd$$

con $RSS = 500$

Calcolate la statistica F . Consideriamo un valore critico, in questo caso , pari a 3.18. Possiamo rifiutare l'ipotesi nulla?

$$F = \frac{(RSS_R - RSS_M)/j}{RSS_M/(N - 2k)} = \frac{(500 - (200 + 150))/2}{350/(50 - 4)} = 9.86$$

In questo caso, la statistica calcolata è maggiore al livello critico, quindi possiamo rifiutare l'ipotesi nulla di conseguenza possiamo affermare che c'è un cambiamento nel modello tra i consumatori europei e quelli americani.

Esercizio 8

a) Provate a scrivere il modello completo nel caso in cui vogliate spiegare la differenza nel prezzo di un biglietto aereo tra due compagnie, Alfa e Beta, considerando come regressori il costo del carburante ed il costo del servizio di catering a bordo, oltre ai costi fissi di manutenzione che rappresentano la costante.

Possiamo considerare il seguente modello:

$$\text{prezzo} = \beta_0 + \beta_1 \text{carburante} + \beta_2 \text{catering} + \varepsilon$$

dove β_0 = costi fissi di manutenzione.

Il modello completo ottenuto usando la variabile dummy, $D=1$ se per esempio la compagnia è Alfa, è il seguente:

$$\text{prezzo} = \beta_0 + \beta_1 \text{carburante} + \beta_2 \text{catering} + \delta_0 D + \delta_1 D \text{carburante} + \delta_2 D \text{catering} + \varepsilon$$

b) Che test potete utilizzare per testare tale differenza?

Possiamo usare il test di Chow, usando la statistica F con la seguente formula:

$$F = \frac{(RSS_R - RSS_M)/j}{RSS_M/(N - 2k)} \sim F(j, N - 2k)$$

c) Considerate la stima del modello completo:

```
. reg prezzo carburante catering
```

Source	SS	df	MS			
Model	2647166.85	2	1323583.42	Number of obs = 1149		
Residual	3137332.44	1146	2737.63738	F(2, 1146) = 483.48		
Total	5784499.28	1148	5038.76244	Prob > F = 0.0000		
				R-squared = 0.4576		
				Adj R-squared = 0.4567		
				Root MSE = 52.322		

prezzo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
carburante	.08822	.0029673	29.73	0.000	.0823981	.0940419
catering	73.7979	9.343216	7.90	0.000	55.46617	92.12962
_cons	42.26237	7.916568	5.34	0.000	26.72978	57.79496

```
reg prezzo carburante catering
```

Source	SS	df	MS			
Model	2686377.72	2	1343188.86	Number of obs = 1149		
Residual	3690823.28	1146	3220.61368	F(2, 1146) = 417.06		
Total	6377200.99	1148	5555.05313	Prob > F = 0.0000		
				R-squared = 0.4212		
				Adj R-squared = 0.4202		
				Root MSE = 56.75		

prezzo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
carburante	.0899637	.0032647	27.56	0.000	.0835582	.0963691
catering	79.34504	10.20206	7.78	0.000	59.32823	99.36186
_cons	40.93595	8.548038	4.79	0.000	24.16439	57.70752

e la stima del modello ristretto:

reg prezzo carburante catering

Source	SS	df	MS			
Model	5470133.56	2	2735066.78	Number of obs =	2298	
Residual	7000151.5	2295	3050.17494	F(2, 2295) =	896.69	
				Prob > F =	0.0000	
				R-squared =	0.4387	
				Adj R-squared =	0.4382	
Total	12470285.1	2297	5428.9443	Root MSE =	55.228	

prezzo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
carburante	.08986	.0022122	40.62	0.000	.085522	.0941981
catering	76.69622	6.896794	11.12	0.000	63.17162	90.22082
_cons	38.33582	5.831495	6.57	0.000	26.90027	49.77137

Calcolate la statistica del test precedentemente descritto, coi dati degli output.

$$F = \frac{(RSS_R - RSS_M)/j}{RSS_M/(N - 2k)} = \frac{(7000151.5 - (3137332.44 + 3690823.28))/3}{(3137332.44 + 3690823.28)/(2298 - 6)} = 19.245$$

Esercizio 9

Considerate il modello

$$Y_i = X_i\beta + u_i$$

$$var(u_i) = \sigma_1^2 \quad \text{per } i \text{ dispari}$$

$$var(u_i) = \sigma_2^2 \quad \text{per } i \text{ pari}$$

con σ_1^2 e σ_2^2 note.

a) Come trasformereste il modello in modo da avere errori con varianza costante?

Abbiamo

$$var(u_i) = \begin{bmatrix} \sigma_1^2 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \sigma_2^2 & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \vdots & & & & & \sigma_1^2 \\ 0 & \dots & \dots & \dots & \dots & \sigma_2^2 \end{bmatrix} = \Omega$$

Se vogliamo avere errori con varianza costante dobbiamo dividere le variabili con i dispari per σ_1 e le variabili con i pari per σ_2 .

Costruiamo una matrice P^{-1} :

$$P^{-1} = \begin{bmatrix} \sigma_1^{-1} & \dots & \dots & \dots & \dots & 0 \\ \vdots & \sigma_2^{-1} & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \sigma_1^{-1} & \sigma_2^{-1} \end{bmatrix}$$

Riscriviamo il modello come:

$$P^{-1} Y_i = P^{-1} X_i \beta + P^{-1} u_i$$

$$\text{con } \text{var}(u_i) = P^{-1} \Omega P^{-1} = I$$

b) Qual è la varianza dello stimatore OLS? Quale è la varianza dello stimatore OLS del modello trasformato?

Lo stimatore OLS è:

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$$

$$\text{var}(\hat{\beta}_{OLS}) = (X'X)^{-1} X' \text{var}(Y_t) X (X'X)^{-1} =$$

$$= (X'X)^{-1} X' \text{var}(u_i) X (X'X)^{-1} = (X'X)^{-1} X' \Omega X (X'X)^{-1}$$

Lo stimatore GLS invece è:

$$\hat{\beta}_{GLS} = (X'P^{-1}P^{-1}X)^{-1} X'P^{-1}P^{-1}Y$$

$$\text{Poichè } P^{-1}\Omega P^{-1} = I \implies \Omega = PP' \implies P^{-1}P^{-1} = \Omega^{-1}$$

riscriviamo lo stimatore $\hat{\beta}_{GLS}$ come:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y.$$

Ne calcoliamo la varianza:

$$\text{var}(\hat{\beta}_{GLS}) = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} \text{var}(Y_i) \Omega^{-1} X (X'\Omega^{-1}X)^{-1} =$$

$$= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} \text{var}(u_i) \Omega^{-1} X (X'\Omega^{-1}X)^{-1} = (X'\Omega^{-1}X)^{-1}$$

Esercizio 10

Considerate il seguente modello:

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \epsilon$$

dove *wage* rappresenta il salario percepito da un lavoratore, *educ* sono gli anni di istruzione, *exper* gli anni di esperienza lavorativa ed *exper*² sono gli anni di esperienza al quadrato.

a) Qual è l'impatto di *exper* sulla variabile *wage*?

$$\frac{\partial \text{wage}}{\partial \text{exper}} = \beta_2 + 2\beta_3 \text{exper}$$

b) Quale sarà il segno di β_3 ?

Ci troviamo di fronte ad una funzione concava del salario all'esperienza, maggiore è l'esperienza, maggiore sarà il salario, ma attenzione, maggiore sono gli anni di esperienza, minore sarà l'impatto dell'esperienza sul salario. Di conseguenza, ci aspettiamo un segno negativo per β_3 .

Ecco il modello stimato in STATA

```
reg wage educ exper expersq
```

Source	SS	df	MS			
Model	1927.87673	3	642.625576	Number of obs =	526	
Residual	5232.53756	522	10.0240183	F(3, 522) =	64.11	
Total	7160.41429	525	13.6388844	Prob > F =	0.0000	
				R-squared =	0.2692	
				Adj R-squared =	0.2650	
				Root MSE =	3.1661	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.5953429	.0530251	11.23	0.000	.4911741	.6995118
exper	.268287	.0368969	7.27	0.000	.1958023	.3407717
expersq	-.0046123	.000822	-5.61	0.000	-.006227	-.0029975
_cons	-3.96489	.7521526	-5.27	0.000	-5.442508	-2.487272

c) Calcolate l'impatto dell'esperienza sul salario, considerando che $exper = 1$

$$\beta_2 + 2\beta_3 \text{ exper} = 0.268287 - 2 * 0.0046123 * 1 = 0.25906$$

d) Scrivete l'ipotesi nulla per testare che il coefficiente dell'istruzione sia pari a 0.6. Scrivete l'ipotesi alternativa nel caso in cui il coefficiente è maggiore di 0.6 e nel caso in cui è diverso a 0.6. Testate il test a due code.

$$H0 : \beta_1 = 0.6 \quad \text{vs} \quad H1 : \beta_1 > 0.6$$

$$H0 : \beta_1 = 0.6 \quad \text{vs} \quad H1 : \beta_1 \neq 0.6$$

$$t = \frac{\widehat{\beta}_1 - 0.6}{\widehat{\sigma}_{\beta_1}} = \frac{0.5953429 - 0.6}{0.053251} = -0.0874556$$

Usando come livello di significatività il 5%, il cui valore critico è 1.96, possiamo non rifiutare l'ipotesi nulla.

e) Scrivete l'ipotesi nulla in cui testate $\beta_2 + \beta_3 = 0.25$. Scrivete il test t che potete implementare in questo caso. Avete tutti i dati nella tabella di STATA?

$$H0 : \beta_2 + \beta_3 = 0.25 \quad \text{vs} \quad H1 : \beta_2 + \beta_3 \neq 0.25$$

$$t = \frac{\widehat{\beta}_2 + \widehat{\beta}_3 - 0.25}{\sqrt{\text{var}(\widehat{\beta}_2 + \widehat{\beta}_3)}}$$

$$\text{var}(\widehat{\beta}_2 + \widehat{\beta}_3) = \text{var}(\widehat{\beta}_2) + \text{var}(\widehat{\beta}_3) + 2\text{cov}(\widehat{\beta}_2, \widehat{\beta}_3)$$

In questo caso non conosciamo la covarianza tra $\widehat{\beta}_2$ e $\widehat{\beta}_3$

f) Scrivete l'ipotesi nulla in cui testate $\beta_2 = -\beta_3$. Scrivete il test t che potete implementare in questo caso. Avete tutti i dati nella tabella di STATA? C'è un modo per riscrivere il modello in modo da testare H_0 guardando direttamente alla significatività di un coefficiente di regressione?

$H_0 : \beta_2 + \beta_3 = 0$. vs $H_1 : \beta_2 + \beta_3 \neq 0$

$$t = \frac{\widehat{\beta}_2 + \widehat{\beta}_3 - 0}{\sqrt{\text{var}(\widehat{\beta}_2 + \widehat{\beta}_3)}}$$

$$\text{var}(\widehat{\beta}_2 + \widehat{\beta}_3) = \text{var}(\widehat{\beta}_2) + \text{var}(\widehat{\beta}_3) + 2\text{cov}(\widehat{\beta}_2, \widehat{\beta}_3)$$

Anche in questo caso non conosciamo la covarianza tra $\widehat{\beta}_2$ e $\widehat{\beta}_3$

Possiamo riscrivere il modello tenendo conto della restizione nella seguente forma:

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + (\beta_2 + \beta_3) \exp \text{er} + \beta_3 (\exp \text{er}^2 - \exp \text{er}) + \varepsilon$$