

ECONOMETRIA – Cod. 6089
PRIMA PROVA PARZIALE (MOCK VERSION 2)

Si risponda soltanto a una domanda della prima parte e a una domanda della seconda parte.
Il tempo a disposizione è di 60 minuti.

PRIMA PARTE (“TEORICA”)

Domanda 1

Si consideri il seguente modello lineare multivariato con N osservazioni e K regressori (inclusa una costante): $\underline{y} = X \underline{\beta} + \underline{u}$ (le variabili sottolineate indicano vettori).

- a) Si dimostri che, sotto determinate assunzioni, gli stimatori dei minimi quadrati ordinari $\hat{\underline{\beta}}$ sono stimatori non distorti di $\underline{\beta}$.
- b) Si assuma, adesso, che il vero modello valido per la popolazione sia: $\underline{y} = X \underline{\beta} + P \underline{\lambda} + \underline{u}$, dove P è una matrice [NxJ] contenente altre J variabili che hanno un impatto sulla variabile dipendente, impatto catturato dai parametri del vettore $\underline{\lambda}$ di dimensione [Jx1]. Gli stimatori OLS derivati nel punto precedente sono ancora stimatori non distorti di $\underline{\beta}$? Argomentate la vostra risposta.
- c) Tralasciando il problema al punto b) e tornando al modello originario con i regressori X, come potremmo stimare l'ipotesi nulla di corretta specificazione funzionale? Si discuta brevemente la procedura inferenziale suggerita.
- d) Si supponga di voler aggiungere un termine quadratico della variabile x_{k-1} e un termine quadratico della variabile x_k nell'insieme dei regressori, di modo che il modello stimato diventa: $\underline{y} = \tilde{X} \underline{\beta} + \underline{v}$, dove $\tilde{X} = [X \quad \underline{x}_{k-1}^2 \quad \underline{x}_k^2]$. Si scrivano (in forma matriciale) le restrizioni sul nuovo modello e il test F che valuta la significatività congiunta dei due termini quadratici.

Domanda 2

Si consideri il modello lineare multivariato $y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_K x_{Ki} + \varepsilon_i$, dove tutte le variabili sono misurate a livello di impresa (i), ovvero il modello è stimato su un campione di N imprese.

- a) Se il modello è correttamente specificato, qual è la varianza (in forma matriciale) dello stimatore OLS del vettore di parametri $\underline{\beta}$? Come la stimate?
- b) Potete confrontare la varianza dello stimatore OLS con quella di un altro generico stimatore lineare dello stesso vettore $\underline{\beta}$? Specificate le ipotesi su cui basate le vostre affermazioni.
- c) Supponete ora che, per motivi di riservatezza, di tutte le variabili disponiate solo della media a livello di settore (s) e che possiate stimare con OLS il modello $\bar{y}_s = \beta_1 + \beta_2 \bar{x}_{2s} + \dots + \beta_K \bar{x}_{Ks} + \bar{\varepsilon}_s$ sul campione di S settori, dove $\bar{y}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} y_i$ (ovvero la media di y_i calcolata sul numero di imprese che operano nel settore s, N_s) e tutte le altre variabili sono definite in modo analogo. Dimostrate che l'errore di questo secondo modello, $\bar{\varepsilon}$, è eteroschedastico.
- d) Spiegate come potete trasformare il modello al punto c) per stimare i coefficienti con il metodo dei minimi quadrati ponderati (WLS)?

SECONDA PARTE (“APPLICATA”)

Domanda 3

Considerate la seguente funzione di domanda:

$$ECOLBS_i = \beta_1 + \beta_2 ECOPRC_i + \beta_3 REGPRC_i + \beta_4 FAMINC_i + \varepsilon_i,$$

dove ECOLBS è la quantità di mele biologiche (in kg), ECOPRC è il loro prezzo in euro, REGPRC è il prezzo delle mele “normali” (non biologiche) e FAMINC è il reddito familiare in euro. La stima è effettuata su un campione di 660 famiglie a cui sono stati sottoposti un prezzo per le mele biologiche ed un prezzo per le mele normali ed a cui è stato chiesto quante mele di ciascun tipo avrebbero acquistato ai dati prezzi. La tabella sottostante riporta la stima OLS della funzione di domanda:

Source	SS	df	MS			
Model	161.057604	3	53.685868	Number of obs =	660	
Residual	4043.07922	656	6.16323052	F(3, 656) =	8.71	
Total	4204.13682	659	6.3795703	Prob > F =	0.0000	
				R-squared =	0.0383	
				Adj R-squared =	0.0339	
				Root MSE =	2.4826	

ecol bs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ecoprc	-2.906232	.588067	-4.94	0.000	-4.060953	-1.751512
regprc	3.029023	.7106725	4.26	0.000	1.633556	4.42449
fami nc	.0030818	.0027098	1.14	0.256	-.0022391	.0084027
_cons	1.778726	.4138961	4.30	0.000	.9660053	2.591447

- a) I coefficienti stimati di ECOPRC e di REGPRC hanno il segno atteso? Rispondete fornendo un’interpretazione economica e utilizzate le informazioni riportate di seguito nella tabella delle statistiche descrittive per fornire una stima dell’elasticità della domanda di mele biologiche al proprio prezzo (calcolata per valori medi delle due variabili).

Variable	Obs	Mean	Std. Dev.	Min	Max
ecol bs	660	1.47399	2.525781	0	42
ecoprc	660	1.081515	.295573	.59	1.59
regprc	660	.8827273	.2444687	.59	1.19
fami nc	660	53.40909	35.74122	5	250

- b) Supponete che il reddito sia ora misurato in migliaia di euro: quale effetto vi attendete sul coefficiente stimato e sulla significatività della variabile FAMINC nella regressione precedente? Motivate la vostra risposta.
- c) Una volta osservato l’output della regressione decidete di testare l’ipotesi nulla $\beta_2 = -\beta_3$ (contro l’ipotesi alternativa $\beta_2 \neq -\beta_3$) al 5%. Utilizzate la stima della matrice di varianza dello stimatore OLS sotto riportata per effettuare il test t corrispondente, sapendo che il valore critico di una distribuzione t_{656} per un livello di significatività del 5% è pari a 1.96.

	ecoprc	regprc	fami nc	_cons
ecoprc	.3458228			
regprc	-.34703252	.50505533		
fami nc	.00004821	2.626e-07	7.343e-06	
_cons	-.07025243	-.07051921	-.00044456	.17130997

- d) Spiegate come costruireste il test della stessa ipotesi specificata al punto c) utilizzando il test F. Quale regressione dovreste svolgere per il modello ristretto?

Domanda 4

Un gruppo di lavoratori viene sottoposto a un corso di formazione nel 1976 e si intende valutare l'effetto di questo corso sulla probabilità di avere un lavoro un anno più tardi (cioè, all'inizio del 1978). La variabile dipendente è una dummy uguale a 1 se il lavoratore è occupato nel 1978 (*em78*). Tra i regressori, includiamo una dummy uguale a 1 se il lavoratore ha seguito il corso di formazione (*train*) più una serie di variabili di controllo che possono influire sulla probabilità di trovare lavoro: età (*age*), anni di studio (*educ*), una dummy per gli sposati (*married*), e due variabili binarie indicanti origini afro-americane (*black*) o ispaniche (*hisp*). Si stima quindi il modello:

$$em78_i = \beta_1 + \beta_2 train_i + \beta_3 educ_i + \beta_4 age_i + \beta_5 married_i + \beta_6 black_i + \beta_7 hisp_i + u_i.$$

```
. reg em78 train age educ black hi sp married
```

Source	SS	df	MS			
Model	10.6219085	6	1.77031809	Number of obs =	2675	
Residual	279.420708	2668	.1047304	F(6, 2668) =	16.90	
Total	290.042617	2674	.108467695	Prob > F =	0.0000	
				R-squared =	0.0366	
				Adj R-squared =	0.0345	
				Root MSE =	.32362	

em78	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
train	-.137391	.0288294	-4.77	0.000	-.1939212	-.0808607
age	-.0045605	.0006433	-7.09	0.000	-.0058218	-.0032991
educ	.0061592	.0022766	2.71	0.007	.0016952	.0106232
black	.023404	.0158165	1.48	0.139	-.0076098	.0544177
hi sp	.0485345	.0350614	1.38	0.166	-.0202158	.1172848
married	.0538545	.0186921	2.88	0.004	.017202	.090507
_cons	.9153492	.043394	21.09	0.000	.83026	1.000438

- Si supponga di avere individuato un problema di eteroschedasticità nel modello appena stimato (magari con un test di Breusch-Pagan). Si discuta formalmente quali problemi crea la presenza di eteroschedasticità agli stimatori OLS.
- Si spieghi perché può essere utile applicare lo stimatore della varianza Huber-White (anche semplicemente detto stimatore della varianza di White) in presenza di eteroschedasticità. Si commentino quindi i risultati riportati qui sotto.

```
. reg em78 train age educ black hi sp married, vce(robust)
```

```
Linear regression
```

```
Number of obs = 2675
F( 6, 2668) = 14.54
Prob > F = 0.0000
R-squared = 0.0366
Root MSE = .32362
```

em78	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
train	-.137391	.0347639	-3.95	0.000	-.2055578	-.0692241
age	-.0045605	.0006686	-6.82	0.000	-.0058715	-.0032495
educ	.0061592	.0022221	2.77	0.006	.0018019	.0105165
black	.023404	.0141089	1.66	0.097	-.0042614	.0510694
hi sp	.0485345	.0312085	1.56	0.120	-.0126608	.1097298
married	.0538545	.0189236	2.85	0.004	.016748	.0909609
_cons	.9153492	.0406758	22.50	0.000	.8355899	.9951086

Si assuma adesso di poter disporre di un'ulteriore variabile di controllo: il fatto di essere stati disoccupati prima di frequentare il corso di formazione (*unem75*). Visto che questa variabile potrebbe influire sia sulla probabilità di trovare un impiego nel 1978, sia sul fatto di ricevere l'offerta formativa nel 1976, la includiamo tra i regressori e stimiamo questo secondo modello:

$$em78_i = \delta_1 + \delta_2 train_i + \delta_3 educ_i + \delta_4 age_i + \delta_5 married_i + \delta_6 black_i + \delta_7 hisp_i + \delta_8 unem75_i + v_i .$$

```
. reg em78 train age educ black hisp married unem75, robust
```

Linear regression

```
Number of obs = 2675
F( 7, 2667) = 62.19
Prob > F = 0.0000
R-squared = 0.2774
Root MSE = .28032
```

em78	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
train	.1273581	.0411615	3.09	0.002	.0466464	.2080697
age	-.001999	.0005588	-3.58	0.000	-.0030947	-.0009034
educ	.0045106	.001898	2.38	0.018	.0007889	.0082323
black	.0147119	.0126151	1.17	0.244	-.0100244	.0394483
hisp	.063269	.0274561	2.30	0.021	.0094316	.1171065
married	.0257958	.0176307	1.46	0.144	-.0087754	.0603669
unem75	-.5175423	.0282081	-18.35	0.000	-.5728543	-.4622303
_cons	.9238178	.0350535	26.35	0.000	.855083	.9925525

- c) Si scriva la formula e si discuta l'interpretazione dell' R^2 in un modello di regressione. Cosa ci dice il confronto dell' R^2 nei due modelli appena stimati?
- d) Confrontando la stima di β_2 nel primo modello con la stima di δ_2 nel secondo modello, cosa possiamo concludere circa l'impatto del corso di formazione?