

**ECONOMETRIA – Cod. 6089**  
**Mock Exam 2 Prova Parziale**  
**PRIMA PARTE (“TEORICA”)**

**Domanda 1**

Considerate la seguente equazione:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

- (a) Supponete che la variabile  $x_3$  sia misurata con errore. Mostrate perché  $x_3$  può essere endogena.
- (b) Di quali proprietà gode lo stimatore OLS in questo caso? Assumete di ottenere una stima  $\hat{\beta}_3 = 125$ : che cosa potete concludere sul vero parametro  $\beta_3$  nella popolazione?
- (c) E' disponibile uno strumento per la variabile  $x_3$ :  $z_1$ . Quali condizioni deve soddisfare la variabile  $z_1$  per essere considerata uno strumento valido?
- (d) Supponete ora di avere due strumenti disponibili:  $z_1$  e  $z_2$ . Spiegate brevemente come è possibile stimare il modello sopra riportato utilizzando il metodo dei minimi quadrati a due stadi (2SLS).

**Domanda 2**

Considerate il seguente modello di regressione con serie storiche ( $t = 1, \dots, T$ ), in cui gli errori seguono un processo AR(1):

$$\begin{aligned} y_t &= \underline{x}_t' \beta + u_t \\ u_t &= \theta u_{t-1} + \varepsilon_t \end{aligned}$$

dove  $\underline{x}_t$  raggruppa  $k$  regressori,  $\theta < 1$  e  $\varepsilon_t \sim N(0, \sigma^2)$ .

- (a) Dimostrate che:  $Corr(u_t, u_{t-k}) = \theta^k$ .
- (b) Di quali proprietà gode lo stimatore OLS in questo caso?
- (c) Come potete trasformare il modello affinché gli errori abbiano media zero, siano omoschedastici e non autocorrelati?
- (d) In generale, in presenza di autocorrelazione, si può far ricorso agli stimatori GLS. Si mostri come ricavare una formula per la matrice di varianza e covarianza di questa classe di stimatori.

## SECONDA PARTE (“APPLICATA”)

### Domanda 3

Avete un dataset che contiene osservazioni su 2725 individui, 755 dei quali sono stati arrestati nel 1986. Nel dataset sono contenute le seguenti informazioni:

*arr86* = variabile dummy uguale a 1 se l'individuo è stato arrestato nel 1986;

*tottime* = mesi spesi in prigione prima del 1986;

*inc86* = reddito (in dollari) nel 1986;

*black* = variabile dummy uguale a 1 se l'individuo è di colore.

Supponete di stimare la seguente relazione lineare con OLS:

$$arr86 = \beta_1 + \beta_2 tottime + \beta_3 inc86 + \beta_4 black + \varepsilon$$

(a) Commentate il risultato del seguente test basato sul residuo della regressione:

estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of arr86

chi2(1) = 62.33

Prob > chi2 = 0.0000

(b) Supponete ora di stimare la stessa relazione con il metodo logit, che vi fornisce il risultato riportato nella tabella sottostante. Quali sono le ipotesi su cui il modello è costruito? Che proprietà ha lo stimatore impiegato?

Logistic regression

Number of obs = 2725

LR chi2(3) = 143.67

Prob > chi2 = 0.0000

Log likelihood = -1536.3479

Pseudo R2 = 0.0447

arr86	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
totttime	-.011618	.0094845	-1.22	0.221	-.0302073	.0069713
inc86	-.0073647	.0008442	-8.72	0.000	-.0090192	-.0057102
black	.6607814	.1111238	5.95	0.000	.4429826	.8785801
_cons	-.7222556	.0617762	-11.69	0.000	-.8433347	-.6011764

- (c) Spiegate come si valuta nel modello logit la bontà della stima. Che conclusioni potete trarre a questo proposito dall'output della regressione?
- (d) Basandosi sul valore del coefficiente stimato della variabile *inc86* è possibile sostenere che un dollaro in più di reddito diminuisce la probabilità che l'individuo sia arrestato dello 0.7 per cento circa. Siete d'accordo? Motivate la vostra risposta.

#### Domanda 4

Abbiamo un dataset con 14 osservazioni annuali dal 1980 al 1993 riferite a un determinato paese. Per ogni anno osserviamo le seguenti variabili:

*criv* = numero di crimini riportati ogni mille abitanti;

*polpc* = numero di poliziotti ogni mille abitanti;

*incpc* = reddito pro capite in termini reali;

*unem* = tasso di disoccupazione.

Supponete di stimare la seguente relazione lineare con OLS:

$$criv = \beta_1 + \beta_2 polpc + \beta_3 incpc + \beta_4 unem + \varepsilon$$

- (a) Commentate il risultato dei due test riportati qui di seguito (Durbin-Watson e Breusch-Godfrey; tenendo conto che in questo caso  $d_L = 0.82$  e  $d_U = 1.75$  per un livello di significatività del 5%). Qual è l'ipotesi nulla dei due test? Quali i rispettivi vantaggi e svantaggi? Che cosa ci suggeriscono in questo caso?

```
. estat dwatson
Durbin-watson d-statistic( 4, 14) = 1.480504
```

```
. estat bgodfrey
```

```
Breusch-Godfrey LM test for autocorrelation
```

lags( $\rho$ )	chi2	df	Prob > chi2
1	0.091	1	0.7630

H0: no serial correlation

- (b) Qui sotto trovate le stime dei coefficienti del modello OLS. Commentate il coefficiente stimato per la variabile “polpc”. Come possiamo interpretarlo? Ci aspettiamo un problema di endogeneità? In quale direzione?

criv	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
polpc	.0337294	.0084843	3.98	0.003	.0148251 .0526337

- (c) Assumiamo adesso di usare uno strumento per la variabile “polpc”: una dummy (“govelec”) uguale ad uno se ci troviamo in un anno pre-elettorale. L’idea è che, prima delle elezioni, il governo aumenta spesa e impiego pubblico per incrementare le chance di essere rieletto, e questo aumenta anche il numero di poliziotti. Usando questo strumento per stimare il modello con 2SLS, al primo stadio, il coefficiente della variabile “govelec” (che misura il suo impatto su “polpc”) ha una statistica t pari a 5.89: che cosa concludete circa l’eventuale debolezza dello strumento? Quali problemi possono sorgere nel caso di strumenti deboli?

- (d) Qui sotto trovate il secondo stadio della stima 2SLS del modello con “govelec” come strumento. Commentate il coefficiente stimato per la variabile “polpc”. Assumendo che lo strumento sia valido, il risultato è coerente con quanto vi aspettavate al punto (b)?

criv	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
polpc	-.0531264	.0223426	-2.38	0.017	-.096917 -.0093357