ECONOMETRIA
TERZO PROBLEM SET

Soluzioni della prima parte

Un processo autoregressivo di ordine uno, AR(2) può essere scritto:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t$$

b) Scrivete la formula della $var(\varepsilon_t)$.
Si noti che: $var(\varepsilon_t) = var(\rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t) =$

$= \rho_1^2 var(\varepsilon_t) + 2\rho_1\rho_2 cov(\varepsilon_{t-1}, \varepsilon_{t-2}) + \rho_2^2 var(\varepsilon_t) + \sigma_u^2$
Per trovare la variaza è necessario calcolare la covarianza $cov(\varepsilon_{t-1}, \varepsilon_{t-2})$*:*
$cov(\varepsilon_{t-1}, \varepsilon_{t-2}) = cov(\rho_1\varepsilon_{t-2} + \rho_2\varepsilon_{t-3} + u_{t-1}, \varepsilon_{t-2})$
ma si noti che $\rightarrow \quad \varepsilon_{t-2} = \rho_1\varepsilon_{t-3} + \rho_2\varepsilon_{t-4} + u_t$
quindi: $cov(\varepsilon_{t-1}, \varepsilon_{t-2}) = \rho_1 var(\varepsilon_t) + \rho_1\rho_2 var(\varepsilon) + \rho_1\rho_2^2 var(\varepsilon_t) + ... =$
$= \frac{\rho_1}{1-\rho_2} var(\varepsilon_t)$
Inserendo questo risultato nella formula iniziale sulla varianza:
$var(\varepsilon_t) = \rho_1^2 var(\varepsilon_t) + 2\rho_1\rho_2(\frac{\rho_1}{1-\rho_2} var(\varepsilon_t)) + \rho_2^2 var(\varepsilon_t) + \sigma_u^2$
quindi: $(1 - \rho_1^2 - 2\rho_1\rho_2(\frac{\rho_1}{1-\rho_2}) - \rho_2^2) var(\varepsilon_t) = \sigma_u^2$
$var(\varepsilon_t) = \sigma_u^2 / (1 - \rho_1^2 - 2\rho_1\rho_2(\frac{\rho_1}{1-\rho_2}) - \rho_2^2)$
c) Scrivete $cov(\varepsilon_t, \varepsilon_{t-1})$ e $cov(\varepsilon_t, \varepsilon_{t-s})$.

Partendo da: $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t$

moltiplicando per ε_t *e prendendo il valore atteso:*
$E(\varepsilon_t^2) = \rho_1 E(\varepsilon_{t-1}\varepsilon_t) + \rho_2 E(\varepsilon_{t-2}\varepsilon_t) + E(u_t\varepsilon_t)$
$E(\varepsilon_t^2) = \rho_1 E(\varepsilon_{t-1}\varepsilon_t) + \rho_2 E(\varepsilon_{t-2}\varepsilon_t) + E(u_t\varepsilon_t)$
$\theta_0 = \rho_1\theta_1 + \rho_2\theta_2 + \sigma^2$
moltiplicando per ε_{t-1} *e prendendo il valore atteso:*
$E(\varepsilon_t\varepsilon_{t-1}) = \rho_1 E(\varepsilon_{t-1}\varepsilon_{t-1}) + \rho_2 E(\varepsilon_{t-2}\varepsilon_{t-1}) + E(u_t\varepsilon_{t-1})$
$E(u_t\varepsilon_{t-1}) = 0$
$\theta_1 = \rho_1\theta_0 + \rho_2\theta_1$
$\theta_1 = \frac{\rho_1}{1-\rho_2}\theta_0$
moltiplicando per ε_{t-2} *e prendendo il valore atteso:*
$E(\varepsilon_t\varepsilon_{t-2}) = \rho_1 E(\varepsilon_{t-1}\varepsilon_{t-2}) + \rho_2 E(\varepsilon_{t-2}\varepsilon_{t-2}) + E(u_t\varepsilon_{t-2})$
$E(u_t\varepsilon_{t-2}) = 0$
$\theta_2 = \rho_1\theta_1 + \rho_2\theta_0$
$\theta_2 = \frac{\rho_1^2}{1-\rho_2}\theta_0 + \rho_2\theta_0$
$\theta_0 = \rho_1(\frac{\rho_1}{1-\rho_2}\theta_0) + \rho_2(\frac{\rho_1^2}{1-\rho_2}\theta_0 + \rho_2\theta_0) + \sigma^2$

$(1 - \rho_1(\frac{\rho_1}{1-\rho_2}) - \rho_2(\frac{\rho_1^2}{1-\rho_2}+\rho_2))\theta_0 = \sigma^2$
$(\frac{1-\rho_2-\rho_1^2-\rho_2\rho_1^2-\rho_2^2+\rho_2^3}{1-\rho_2})\theta_0 = \sigma^2$
$D = 1 - \rho_2 - \rho_1^2 - \rho_2\rho_1^2 - \rho_2^2 + \rho_2^3$
$(\frac{D}{1-\rho_2})\theta_0 = \sigma^2$
$\theta_0 = \frac{1-\rho_2}{D}\sigma^2$ *(varianza)*

$\theta_1 = \frac{\rho_1}{1-\rho_2}\frac{1-\rho_2}{D}\sigma^2$
$\theta_1 = \frac{\rho_1}{D}\sigma^2$ *(covarianza prima ordine)*

$\theta_2 = \frac{\rho_1^2}{1-\rho_2}\theta_0 + \rho_2\theta_0$
$\theta_2 = (\frac{\rho_1^2}{1-\rho_2}+\rho_2)\theta_0$
$\theta_2 = (\frac{\rho_1^2}{1-\rho_2}+\rho_2)\frac{1-\rho_2}{D}\sigma^2$
$\theta_2 = (\frac{\rho_1^2+\rho_2(1-\rho_2)}{D})\sigma^2$ *(covarianza seconda ordine)*
Possiamo fermarci qui perché questo è già sufficiente per calcolare la matrice di varianza-covarianza con $T = 3$.

d) Scrivete la $corr(\varepsilon_t, \varepsilon_{t-1})$.

Semplicemente: $corr(\varepsilon_t, \varepsilon_{t-k}) = \frac{\theta_k}{\theta_0}$
$corr(\varepsilon_t, \varepsilon_{t-1}) = \frac{\theta_1}{\theta_0}$

e) Scrivete la matrice varianza-covarianza di ε_t supponendo che $T = 3$.

Nel caso in cui $T = 3$
$$var(\varepsilon_t) = \begin{pmatrix} var(\varepsilon_1) & cov(\varepsilon_1,\varepsilon_2) & cov(\varepsilon_1,\varepsilon_3) \\ cov(\varepsilon_2,\varepsilon_1) & var(\varepsilon_2) & cov(\varepsilon_2,\varepsilon 3) \\ cov(\varepsilon_3,\varepsilon_1) & cov(\varepsilon_3,\varepsilon_2) & var(\varepsilon_3) \end{pmatrix} = \begin{pmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_1 & \theta_0 & \theta_1 \\ \theta_2 & \theta_1 & \theta_0 \end{pmatrix}$$
Questo ci fa vedere come implementare uno stimatore GLS assumendo un errore autoregressivo di ordine 2 in un campione con 3 osservazioni.

Esercizio 2

Considerate il processo:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

dove ε_t è un processo AR(1):

$$\begin{aligned} \varepsilon_t &= \rho\varepsilon_{t-1} + u_t \text{ con } u_t \sim iidN(0, \sigma_u^2). \\ \text{dove } \rho &= 1 \end{aligned}$$

a) Scrivete la $var(\varepsilon_t)$.

Ricordatevi che nel modelo AR(1), la assunzione $\rho < 1$ è cruciale per garantire che l'autocorrelazione del processo tenda a zero con il passere del tempo. Quindi, in questo caso ci troviamo di fronte a un processo non-stazionario. Quando $\rho = 1$, l'autocorrelazione del processo (detto anche random walk) cresce col tempo. Inoltre, si può notare che non importa quanto ci si allontana nel futuro, la migliore previsione per ε_{t+h} rimane ε_t.

Se $\rho = 1$ la varianza cambia con t. La varianza è una funzione crescente del tempo (t).

$$\varepsilon_t = \varepsilon_{t-1} + u_t = (\varepsilon_{t-2} + u_{t-1}) + u_t = ... = \partial + \sum_{j=1}^{t} u_j \qquad dove\ \partial = \varepsilon_0$$

$$var(\varepsilon_t) = var(\sum_{j=1}^{t} u_j) = t\sigma_u^2$$

$$\lim_{T\to\infty} var(\varepsilon_t) = \infty$$

b) Scrivete la $cov(\varepsilon_t \varepsilon_{t-1})$.

Si noti che: $cov(\varepsilon_t \varepsilon_{t-1}) = E(\varepsilon_t \varepsilon_{t-1}) = E[(\varepsilon_{t-1} + u_t)\,\varepsilon_{t-1}] = E(\varepsilon_{t-1}^2) = (t-1)\sigma_u^2$

Nel caso generale:

$cov(\varepsilon_t \varepsilon_{t-h}) = E(\varepsilon_t \varepsilon_{t-h}) = E[(\varepsilon_{t-1} + u_t)\,\varepsilon_{t-h}] = ... = (t-h)\sigma_u^2$

c) Scrivete la $corr(\varepsilon_t, \varepsilon_{t-1})$.

Dai risultati precedenti, si vede che:

$corr(\varepsilon_t, \varepsilon_{t-h}) = \frac{cov(\varepsilon_t \varepsilon_{t-h})}{\sqrt{var(\varepsilon_t)}\sqrt{var(\varepsilon_{t-h})}} = \frac{(t-h)\sigma_u^2}{\sigma_u^2 \sqrt{t}\sqrt{t-h}} = \frac{\sqrt{t-h}}{\sqrt{t}}$

Nel caso h = 1: $corr(\varepsilon_t, \varepsilon_{t-1}) = \frac{\sqrt{t-1}}{\sqrt{t}}$

Si vede anche che: $corr(\varepsilon_t, \varepsilon_{t+h}) = \frac{cov(\varepsilon_t \varepsilon_{t+h})}{\sqrt{var(\varepsilon_t)}\sqrt{var(\varepsilon_{t+h})}} = \frac{\sqrt{t}}{\sqrt{t+h}}$

Quindi, si noti che in questo caso, al crescere di t, quando h tende all'infinito, l'autocorrelazione tende a zero ma più lentamente di un processo autoregressivo stazionario.

Esercizio 3

Supponiamo di avere il seguente modello:

$$y_t = \beta_0 + \beta_1 C_t + \varepsilon_t \qquad t = 1, ..., T$$

dove $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ con $u_t \sim iid\ N(0, \sigma_u^2)$.

a) Proponete una trasformazione con $t \geqq 2$ in modo tale che l'errore non sia serialmente correlato. Come viene chiamata tale trasformazione?

i) Consideriamo il modello relativo al periodo $t-1$ e moltiplicando per ρ

$$\rho y_{t-1} = \beta_0 \rho + \beta_1 \rho C_{t-1} + \rho \varepsilon_{t-1}$$

ii) sottraiamo $y_t - \rho y_{t-1}$

$$(y_t - \rho y_{t-1}) = \beta_0 (1-\rho) + \beta_1 (C_t - \rho C_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

Quindi possiamo riscrivere il modello nel seguente modo:

$$\widetilde{y_t} = \beta_0 \widetilde{x_{1t}} + \beta_1 \widetilde{x_{2t}} + u_t$$

Questa trasformazione viene chiamata di Cochrane-Orcutt (transformazione dei dati in quasi-differenze).

Questo modello non è proprio identico a GLS (o FGLS), perché abbiamo perso la prima osservazione. Ma è comunque asintoticamente equivalente.

b) Quale tipo di trasformazione considera anche la prima osservazione? Mostrate come modifichereste la trasformazione precedente.

Tale trasformazione si chiama di Prais-Winsten.

Partiamo dal modello:

$$y_1 = \beta_0 + \beta_1 C_1 + \varepsilon_1$$

Moltiplico per $\sqrt{1-\rho^2}$
Dove:

$$\left(\sqrt{1-\rho^2}\right) y_1 = \beta_0 \left(\sqrt{1-\rho^2}\right) + \beta_1 C_1 \left(\sqrt{1-\rho^2}\right) + \varepsilon_1 \left(\sqrt{1-\rho^2}\right)$$

$\left(\sqrt{1-\rho^2}\right) y_1 = \widetilde{y_1}$
$\left(\sqrt{1-\rho^2}\right) = \widetilde{x_{11}}$
$C_1 \left(\sqrt{1-\rho^2}\right) = \widetilde{x_{21}}$
$\varepsilon_1 \left(\sqrt{1-\rho^2}\right) = u_1$
$E(\varepsilon_1 \left(\sqrt{1-\rho^2}\right)) = \sqrt{1-\rho^2} E(\varepsilon_1) = 0$
$var(\varepsilon_1 \left(\sqrt{1-\rho^2}\right)) = \left(1-\rho^2\right) var(\varepsilon_1) = \left(1-\rho^2\right) \frac{\sigma_u^2}{1-\rho^2} = \sigma_u^2$
Quindi, possiamo usare $\widetilde{y_1} = \widetilde{x_{11}} \beta_0 + \beta_1 \widetilde{x_{21}} + u_1$ *in una regressione OLS.*

Asintoticamente non c'è differenza se si esclude la prima osservazione. Ma, visto che la maggiore parte dei campioni non sono grandi, includere la prima osservazione può essere importante.

Esercizio 4

a) Scrivete la statistica del test di Durbin-Watson e spiegate il test. Nel caso di un campione di 50 osservazioni e con 2 regressori oltre alla costante, come potete commentare il valore del test di Durbin-Watson ndei seguenti casi:

DW=1.20
DW=1.40
DW=1.60
DW=1.70

Usate la tabella qui sotto riportata coi valori critici di dL e dU per il test di Durbin-Watson. (Ipotizziamo come ipotesi alternativa la correlazione positiva).

		X variables, excluding the intercept									
Observations		1		2		3		4		5	
N	Prob.	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U
15	0.05	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
	0.01	0.81	1.07	0.7	1.25	0.59	1.46	0.49	1.70	0.39	1.96
20	0.05	1.20	1.71	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
	0.01	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
25	0.05	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
	0.01	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
30	0.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
	0.01	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
40	0.05	1.44	1.54	1.39	1.60	1.34	1.66	1.39	1.72	1.23	1.79
	0.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
50	0.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
	0.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
60	0.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
	0.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
80	0.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
	0.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
100	0.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78
	0.01	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Partendo dal modello:

$$
\begin{aligned}
y_t &= x_t'\beta + \varepsilon_t \\
\varepsilon_t &= \rho\varepsilon_{t-1} + u_t \\
&\mid \rho \mid < 1 \\
u_t &\sim iidN(0, \sigma_u^2).
\end{aligned}
$$

Nel caso in cui $\rho = 0, \varepsilon_t = u_t$
dove $H_0 : \rho = 0$
con ipotesi alternativa:
H_1: $\rho \neq 0$
H_1: $\rho > 0$

H_1: $\rho < 0$

La statistica test utilizzata è:

$$d = \frac{\sum_{t=2}^{T} \left(\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{T} \widehat{\varepsilon^2}_t}$$

dove $\widehat{\varepsilon}_t$ è il residuo della regressione OLS sul modello:

$$y_t = x_t^{'}\beta + \varepsilon_t$$

$$d \simeq 2(1-\widehat{\rho})$$

Se
$\widehat{\rho} = 0$ allora $d \simeq 2$ assenza di correlazione
$\widehat{\rho} = 1$ allora $d \simeq 0$ correlazione positiva
$\widehat{\rho} = -1$ allora $d \simeq 4$ correlazione negativo

In questo caso con 2 variabili oltre alal costante e 50 osservazioni, abbiamo i seguenti valori critici:

al 5% di significatività, 1.46<d<1.63

all'1% di significatività 1.28<d<1.45

Quindi:

DW=1.20: al 5% ed all'1%, rifiutiamo l'ipotesi nulla, abbiamo correlazione positiva

DW=1.40 al 5% rifiutiamo l'ipotesi nulla, ma non all'1% ed in questo caso ricadiamo nella regione di incertezza

DW=1.60 all'1% possiamo dire che c'è assenza di correlazione, ma nel caso del 5% ricadiamo nella regione di incertezza

DW=1.70 al 5% ed all'1%, non rifiutiamo l'ipotesi nulla, non c'è correlazione

b) Ipotizziamo un modello

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

dove

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \phi_3\varepsilon_{t-3} + u_t$$

Che tipo di processo è ε_t? Che test ossiamo usare per verificare il problema di autocorrelazione?

In questo caso il processo ε_t è un processo autoregressivo con 3 ritardi (AR(3)).

In questo caso possiamo usare come test per verificare il problema di autocorrelazione, il test di Breusch-Godfrey, dove $H_0 : \rho = 0$,con ipotesi alternativa $H_1 : \rho \neq 0$.

In questo test, vi diversi passaggi:

1) viene svolta una regressione OLS su $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ *e si ottengono* $\widehat{\beta_0}, \widehat{\beta_1}$ *e* $\widehat{\varepsilon}_t$

2) viene svolta la regressione OLS $\widehat{\varepsilon}_t = \beta_0 + \beta_1 x_t + \alpha_1 \widehat{\varepsilon}_{t-1} + \alpha_2 \widehat{\varepsilon}_{t-2} + \alpha_3 \widehat{\varepsilon}_{t-3} + u_t$

3) viene svolto un TEST di SIGNIFICATIVITA' CONGIUNTA F sugli α, *oppure :*

$$(T-3)xR^2_{aux} \underset{H_0}{\overset{a}{\sim}} \chi^2(3)$$

T-3=numero di osservazioni, dato che abbiamo 3 ritardi

Esercizio 5

a) Consideriamo la seguente regressione, dove le vendite mensili di giocattoli di un dato paese (y) sono regredite sul consumo (c).

Nel seguente output, sono presenti anche due test sui residui della regressione. Commentate.

```
. reg y c

      Source |       SS       df       MS              Number of obs =     524
-------------+------------------------------           F(  1,   522) =   52.88
       Model |  13.8769739     1  13.8769739           Prob > F      =  0.0000
    Residual |  136.988471   522  .262430021           R-squared     =  0.0920
-------------+------------------------------           Adj R-squared =  0.0902
       Total |  150.865445   523  .288461654           Root MSE      =  .51228

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           c |   .4882883   .0671484     7.27   0.000      .356374    .6202027
       _cons |   .0040183    .022384     0.18   0.858    -.0399555    .0479921
------------------------------------------------------------------------------

.  estat dwatson

Durbin-Watson d-statistic(  2,   524) =  1.702273



.  estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         14.561               1                   0.0001
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Il primo test, Durbin-Watson, assume un valore pari a 1.70. In questo caso dobbiamo consultare la tabella coi valori critici, cosniderando 524 come numero delle osservazioni e con 1 il numero dei regressori oltre alla costante. Quindi non possiamo rispondere in merito a tale test. Quindi, possiamo considerare il test di Bruesch-Godaprey, che considera un processo AR (1) .per i residui, tale *test rifiuta l'ipotesi nulla. Quindi potrebbe esserci u n problema di correlazione seriale.*

b) Nel seguente output, il modello è stato trasformato con l'uso della stima di Prais-Winsten. Le stime sono molto diverse rispetto a prima? Cosa ci indica la statistica di Durbin-Watson?

```
. prais y c

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.1488
Iteration 2:  rho = 0.1803
Iteration 3:  rho = 0.1874
Iteration 4:  rho = 0.1891
Iteration 5:  rho = 0.1894
Iteration 6:  rho = 0.1895
Iteration 7:  rho = 0.1895
Iteration 8:  rho = 0.1895
Iteration 9:  rho = 0.1895

Prais-Winsten AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =     524
-------------+------------------------------           F(  1,   522) =   25.73
       Model |  6.56420242     1  6.56420242           Prob > F      =  0.0000
    Residual |  133.146932   522   .25507075           R-squared     =  0.0470
-------------+------------------------------           Adj R-squared =  0.0452
       Total |  139.711134   523    .2671341           Root MSE      =  .50505

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           c |   .3495857    .068912     5.07   0.000     .2142067    .4849647
       _cons |   .0049985   .0272145     0.18   0.854    -.0484649    .0584619
-------------+----------------------------------------------------------------
         rho |   .1895324
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.702273
Durbin-Watson statistic (transformed) 2.007414
```

In questo caso, la stima è minore rispetto a quella ottenuta nel caso di uno stimatore OLS. Notiamo che la procedura si basa sulla stima per iterazione del coefficiente rho. La statistica di DW dopo la trasformazione è aumentata ed è uguale a 2, ciò dimostra che è stato risolto il problema di correlazione segnalato precedentemente.

ECONOMETRIA
TERZO PROBLEM SET

Soluzioni della seconda parte

Esercizio 6

Consideriamo il modello in cui regrediamo la variabile y (che rappresenta il salario di un individuo) sulla variabile x_1 (anni di studio) e x_2^*
(abilità individuale):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^* + u$$

dove la variabile abilità non è osservabile e quindi è misurata con una proxy tale che:

$$x_2^* = \theta_0 + \theta_2 x_2 + v_2$$

a) Cosa rappresentano θ_2 e v_2?

Il parametro θ_2 rappresenta la relazione tra x_2^* e x_2 ; se x_2^* e x_2 sono positivamente correlati, $\theta_2 > 0$.

Se $\theta_2 = 0$, x_2 non è una proxy adatta per x_2^*.

La intercetta θ_0 può essere positiva ($\theta_0 > 0$) o negativa ($\theta_0 < 0$) e semplicimente permette che x_2^* e x_2 siano misurati in scala diversa.

v_2 rappresenta il termine di errore.

b) Scrivete l'equazione che stimereste in questo caso. (Suggerimento: sostituite la seconda equazione nella prima).

$y = \beta_0 + \beta_1 x_1 + \beta_2(\theta_0 + \theta_2 x_2 + v_2) + u$

$y = \beta_0 + \beta_2\theta_0 + \beta_1 x_1 + \beta_2\theta_2 x_2 + \beta_2 v_2 + u$

$y = (\beta_0 + \beta_2\theta_0) + \beta_1 x_1 + (\beta_2\theta_2)x_2 + (\beta_2 v_2 + u)$

Considerate che:

$\pi_0 = \beta_0 + \beta_2\theta_0$

$\pi_2 = \beta_2\theta_2$

$\varepsilon = \beta_2 v_2 + u$

$y = \pi_0 + \beta_1 x_1 + \pi_2 x_2 + \varepsilon$

Adesso è possibile ottenere degli stimatori non distorti di π_0, β_1 e π_2.

c) Quali sono le condizioni che garantiscono la correttezza degli stimatori OLS del modello al punto (b)? Spiegate.

L'assunzione necessaria per garantire che questa soluzione fornisca stimatori consistenti può essere riassunta

in due assunzioni su u e v_2 :

1) $cov(x_1, u) = 0$; e $cov(x_2, u) = 0$

L'errore u deve essere incorrelato con x_1 e x_2^* (assunzione di base del modelo OLS).

In aggiunta, u deve essere incorrelato con x_2.

Quindi x_2^* deve influenzare y diretamente e non attraverso x_2.

x_2 influenza y solo attraverso x_2^*
Questa asunzione richiede che x_2 sia una buona proxy per x_2^*.
2) $cov(v_2, x_1) = 0$; e $cov(x_2, \varepsilon) = 0$
L'errore v_2 deve essere incorrelato con x_1, x_2
d) Assumete adesso che:

$$x_2^* = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + v_2$$

Spiegate se gli stimatori OLS sono distorti in questo caso e calcolate la distorsione.
Si noti che: $y = \beta_0 + \beta_1 x_1 + \beta_2(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + v_2) + u$
$y = \beta_0 + \beta_2\theta_0 + \beta_1 x_1 + \beta_2\theta_1 x_1 + \beta_2\theta_2 x_2 + \beta_2 v_2 + u$
$y = (\beta_0 + \beta_2\theta_0) + (\beta_1 + \beta_2\theta_1)x_1 + (\beta_2\theta_2)x_2 + (\beta_2 v_2 + u)$
$p\lim(\beta_1\_hat) = \beta_1 + \beta_2\theta_1$
$p\lim(\beta_2\_hat) = \beta_2\theta_2$
Quindi, se $\beta_2 > 0$, distorsione asintotica positiva (stimatore inconsistente),

a patto che l'abilità individuale abbia una correlazione (parziale) positiva con l'istruzione ($\theta_2 > 0$).
e) Dimostrate che $Cov(x_2^*, \varepsilon)$ è diversa da zero, dove ε è l'errore del modello stimato al punto (b).
Semplicemente: $Cov(x_2^*, \varepsilon) = E(x_2^*, \varepsilon) = E[(x_2 + v_2)(\beta_2 v_2 + u)] = \beta_2\sigma_{v_2}^2$

Esercizio 7

Nel caso in cui :

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

dove $E(\varepsilon) = 0; cov(x_1, \varepsilon) = E(x_1\varepsilon) \neq 0$
a) Quali sono le condizioni che una variabile strumentale z deve soddisfare?
i) $cov(z, \varepsilon) = E(z\varepsilon) = 0$ CONDIZIONE DI ESOGENEITA'
ii) $dy/dz = \beta_1(dx/dz)$ RESTRIZIONE DI ESCLUSIONE
iii) $cov(z, x_1) \neq 0$ CONDIZIONE DI RILEVANZA
b) Dimostrate che lo stimatore IV (di variabili strumentali) è consistente.

$$\widehat{\beta}_{1IV} = \frac{\widehat{cov}(z, y)}{\widehat{cov}(z, x_1)}$$

$$P\lim\left(\widehat{\beta}_{1IV}\right) = P\lim\left(\frac{\widehat{cov}(z, y)}{\widehat{cov}(z, x_1)}\right) = \frac{cov(z, y)}{cov(z, x_1)}$$

Considerando che:

$$
\begin{aligned}
y - E(y) &= \beta_1 \left[x_1 - E(x_1)\right] + \varepsilon \\
\left[z - E(z)\right]\left[y - E(y)\right] &= \beta_1[z - E(z)][x_1 - E(x_1)] + \varepsilon[z - E(z)] \\
cov(z, y) &= \beta_1 cov(z, x_1) + cov(z, \varepsilon) \\
\beta_1 &= \frac{cov(z, y)}{cov(z, x_1)} - \frac{cov(z, \varepsilon)}{cov(z, x_1)} \\
con\ cov(z, \varepsilon) &= 0 \\
\beta_1 &= \frac{cov(z, y)}{cov(z, x_1)} \Rightarrow P\lim\left(\widehat{\beta}_{1IV}\right) = \beta_1
\end{aligned}
$$

c) Cosa accade se $cov(z, \varepsilon) \neq 0$. Che problema si verifica?

$$
\begin{aligned}
P\lim\left(\widehat{\beta}_{1IV}\right) &= P\lim\left(\frac{\widehat{cov}(z, y)}{\widehat{cov}(z, x_1)}\right) = \frac{cov(z, y)}{cov(z, x_1)} = \\
&= \beta_1 + \frac{cov(z, \varepsilon)}{cov(z, x_1)} = \\
&= \beta_1 + \frac{\frac{cov(z,\varepsilon)}{\sigma_z \sigma_\varepsilon}\sigma_\varepsilon}{\frac{cov(z,x_1)}{\sigma_z \sigma_x}\sigma_x} = \\
&= \beta_1 + \frac{\rho_{z\varepsilon}}{\rho_{zx}} \frac{\sigma_\varepsilon}{\sigma_x}
\end{aligned}
$$

In questo caso si vede che anche se $\rho_{z\varepsilon}$ è piccolo, la distorsione dello stimatore IV può essere molto grande se ρ_{zx} è piccolo e quindi la condizione per cui $cov(z, x_1) \neq 0$ è verificata debolmente. Quindi possiamo definire questo uno strumento debole. Inoltre, possiamo confrontare la distorsione (asintotica) di $\widehat{\beta}_{1IV}$ con quella dello stimatore OLS:

$$
\begin{aligned}
P\lim\left(\widehat{\beta}_1\right) &= \beta_1 + \frac{cov(x_1, \varepsilon)}{var(x_1)} = \\
&= \beta_1 + \frac{cov(x_1, \varepsilon)}{\sigma_\varepsilon \sigma_x} \frac{\sigma_\varepsilon}{\sigma_x} = \\
&= \rho_{x\varepsilon} \frac{\sigma_\varepsilon}{\sigma_x}
\end{aligned}
$$

Affinché la distorsione asintotica di cui soffre lo stimatore IV sia inferiore di quella di cui soffre l'OLS non basta che $|\rho_{zx}| < |\rho_{x\varepsilon}|$, ma deve essere che $|\frac{\rho_{z\varepsilon}}{\rho_{zx}}| < \rho_{x\varepsilon}$ ovvero poichè $|\rho_{zx}| < 1, |\rho_{z\varepsilon}|$ deve essere sufficientemente inferiore a $|\rho_{x\varepsilon}|$.

Esercizio 8

a) Scrivete lo stimatore IV multivariato nel caso di esatta identificazione e di sovraidentificazione. Spiegate perché è importante la condizione d'ordine.

La condizione d'ordine è necessaria per l'identificazione e richiede che il numero delle variabili esogene sia almeno uguale al numero delle variabili endogene incluse. Questa condizione garantisce che vi sia un numero di equazioni uguale al numero di parametri da stimare nel sistema che identifica lo stimatore IV. Se L=K, abbiamo esatta identificazione, se L>K, abbiamo sovraidentificazione.

L=K esatta identificazione

$$\widehat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$

L>K sovraidentificazione

$$\begin{aligned}
\widehat{\beta}_{2SLS} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'Y \\
\text{dove } \widehat{X} &= Z\widehat{\Pi} = Z(Z'Z)^{-1}Z'X = P_z X
\end{aligned}$$

sono i valori fittati della regressione OLS sul modello $X = Z\Pi + U$

b) Dimostrate che nel caso L=K

$$\widehat{\beta}_{IV} = \widehat{\beta}_{2SLS}$$

Per dimostrare questo passaggio, dobbiamo partire considerando il residuo del primo stadio:

$$\begin{aligned}
\widehat{u} &= M_z X \\
M_z &= (I - Z(Z'Z)^{-1}Z') = I - P_z \\
M_z Z &= 0
\end{aligned}$$

Di conseguenza:

$$\begin{aligned}
\widehat{X}'X &= \widehat{X}'(\widehat{X} + \widehat{u}) = \\
&= \widehat{X}'(\widehat{X} + M_z X) = \\
&= \widehat{X}'\widehat{X} + X'Z(Z'Z)^{-1}Z'M_z X = \\
&= \widehat{X}'\widehat{X} = \\
&= \widehat{\Pi}'Z'Z\widehat{\Pi} = \\
&= X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X = \\
&= X'Z(Z'Z)^{-1}Z'X
\end{aligned}$$

$$
\begin{aligned}
\widehat{\beta}_{2SLS} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z')Y \\
\widehat{\beta}_{2SLS} &= \widehat{\beta}_{IV} = (Z'X)^{-1}Z'Y
\end{aligned}
$$

c) Dimostrate la distorsione di $\widehat{\beta}_{2SLS}, \widehat{\beta}_{IV}$

Si noti che:

$$
\begin{aligned}
E(\widehat{\beta}_{IV}|X,Z) &= \\
&= E[(\widehat{X}'\widehat{X})^{-1}\widehat{X}'y|X,Z] = \\
&= E[(\widehat{X}'\widehat{X})^{-1}\widehat{X}'(X\beta+\varepsilon)|X,Z] = \\
&= \beta + (\widehat{X}'\widehat{X})^{-1}\widehat{X}'E(\varepsilon|X,Z)
\end{aligned}
$$

Esercizio 9

Consideriamo il seguente modello stimato in STATA:

$$
lw = \beta_0 + \beta_1 s + \beta_2 tenure + \beta_3 \exp r + \beta_4 iq + \varepsilon
$$

dove *lw*= logaritmo dei salari; *s*=anni di scuola completati dall'individuo; *tenure*=anni di tenure; *expr*= anni di esperienza; *iq*=indicatore del quoziente di intelligenza.

a) Che problema può esserci nell'uso della variabile *iq*? Come è possibile risolverlo? Osservando la tabella delle variabili disponibili, cercate una soluzione la problema.

```
obs:           758                          Wages of Very Young Men, Zvi
                                               Griliches, J.Pol.Ec. 1976
 vars:            27                          31 Oct 2004 14:12
 size:        68,978 (99.3% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
rns             float  %9.0g                  residency in South
rns80           float  %9.0g
mrt             float  %9.0g                  marital status = 1 if married
mrt80           float  %9.0g
smsa            float  %9.0g                  reside metro area = 1 if urban
smsa80          float  %9.0g
med             float  %9.0g                  mother's education, years
iq              float  %9.0g                  iq score
kww             float  %9.0g                  score on knowledge in world of
                                                work test
year            float  %9.0g
age             float  %9.0g
age80           float  %9.0g
s               float  %9.0g                  completed years of schooling
s80             float  %9.0g
expr            float  %9.0g                  experience, years
expr80          float  %9.0g
tenure          float  %9.0g                  tenure, years
tenure80        float  %9.0g
lw              float  %9.0g                  log wage
lw80            float  %9.0g
_Iyear_67       byte   %8.0g                  year==67
_Iyear_68       byte   %8.0g                  year==68
_Iyear_69       byte   %8.0g                  year==69
_Iyear_70       byte   %8.0g                  year==70
_Iyear_71       byte   %8.0g                  year==71
_Iyear_73       byte   %8.0g                  year==73
_est_iv         byte   %8.0g                  esample() from estimates store
-------------------------------------------------------------------------------
```

Nel caso di uso della variabile iq, possono esserci problemi di misurazione e quindi dobbiamo strumentare la variabile che ha problemi di endogeneità con almeno una o più variabili esogene. Dalla tabella, possiamo usare come variabili esogene, oltre a quelle già presenti nella regressione, la variabile kww che è un indicatore di un test di conoscenza e possiamo usare anche la variabile med, che indica gli anni di istruzione della madre.

b) Commentate il seguente output di STATA

```
. ivreg  lw s tenure expr (iq=kww med ), first

First-stage regressions
-----------------------

      Source |       SS       df       MS              Number of obs =     758
-------------+------------------------------           F(  5,   752) =   64.09
       Model |  41951.1677     5  8390.23354           Prob > F      =  0.0000
    Residual |  98448.1581   752  130.915104           R-squared     =  0.2988
-------------+------------------------------           Adj R-squared =  0.2941
       Total |  140399.326   757  185.468066           Root MSE      =  11.442

------------------------------------------------------------------------------
          iq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   2.480376   .2177002    11.39   0.000     2.053004    2.907749
      tenure |   .2372269   .2602616     0.91   0.362    -.2736987    .7481525
        expr |  -.4446049   .2109757    -2.11   0.035    -.8587763   -.0304335
         kww |   .3100834   .0637504     4.86   0.000     .1849335    .4352334
         med |   .4114398   .1626099     2.53   0.012     .0922165    .7306632
       _cons |   55.11403   3.050712    18.07   0.000     49.12511    61.10296
------------------------------------------------------------------------------


Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     758
-------------+------------------------------           F(  4,   753) =   76.06
       Model |  26.0174883     4  6.50437208           Prob > F      =  0.0000
    Residual |  113.268662   753  .150423189           R-squared     =  0.1868
-------------+------------------------------           Adj R-squared =  0.1825
       Total |   139.28615   757  .183997556           Root MSE      =  .38784

------------------------------------------------------------------------------
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0178162   .0059301     3.00   0.003     .0061746    .0294578
           s |   .0520467    .019293     2.70   0.007     .0141722    .0899213
      tenure |   .0297485   .0090933     3.27   0.001     .0118973    .0475997
        expr |   .0442787   .0074261     5.96   0.000     .0297004     .058857
       _cons |   3.007415   .3826309     7.86   0.000     2.256265    3.758565
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s tenure expr kww med
------------------------------------------------------------------------------
```

In questo output abbiamo una regressione a due stadi, nel primo stadio la variabile strumentata, iq, è regredita sui due strumenti scelti e sulle altre variabili esogene della regressione. In questa regressione, tutte le variabili, tranne tenure sembra che siano in grado di spiegare la variabile iq. Inoltre la F statistic ha un valore elevato (oltre a 10, regola del pollice di Watson) e quindi le variabili strumentali non sono strumenti deboli. Nel secondo stadio, i residui della regressione del primo stadio sono usati in sostituzione della variabile iq. Tutti i coefficienti sono significativi.

c) Commentate l'output del test di Sargan. A cosa serve tale test?

```
. overid

Tests of overidentifying restrictions:
Sargan N*R-sq test        0.236  Chi-sq(1)    P-value = 0.6271
Basmann test              0.234  Chi-sq(1)    P-value = 0.6284
```

Il test di Sargan serve per verificare la validità degli strumenti in eccesso (L-K) considerando che K strumenti sono validi. In questo caso, non possiamo rifiutare l'ipotesi nulla, quindi gli strumenti in eccesso sono validi.

d) Commentate l'output del test di Hausman. A cosa serve tale test?

```
. hausman iv .

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       iv           .          Difference          S.E.
-------------+----------------------------------------------------------------
          iq |    .0178162     .0038838        .0139323        .0058275
           s |    .0520467     .0947161       -.0426694        .0180552
      tenure |    .0297485     .0362904       -.0065419        .0045474
        expr |    .0442787     .0390324        .0052463         .003694
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from ivreg
          B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        5.72
                Prob>chi2 =      0.2214

.
```

Il test di Hausman serve per testare l'esogeneità della variabile iq. L'ipotesi nulla è che non ci sia correlazione tra la variabile iq e l'errore. In questo caso, l'ipotesi nulla non viene rifiutata, quindi è possibile utilizzare un semplice modello OLS al posto di una stima a 2 stadi perché entrambi gli stimatori sono consistenti ma OLS è più efficiente.