Econometria

Prof. Valerio Potì

*Esercizi Empirici II**

*Questi esercizi sono tratti e adattati dalla 3° edizione del libro di testo, con soluzioni svolte in maniera più dettagliata e approfondita che nelle soluzioni fornite dal sito web (anche correggendone alcuni errori). Ciò vale anche in confronto alle soluzioni degli esercizi empirici disponibili sul sito della 4° edizione del libro di testo. <u>Si consiglia pertanto l'uso della presente anche a chi utilizzi la 4° edizione del libro di testo</u>.

DOMANDE

ESERCIZIO EMPIRICO 14.7 (14.2 nella 4° Ed.)

a) Si ripetano i calcoli riportati nella tabella 14.3 (14.2 nella 4° ed.) usando le regressioni stimate per il periodo campionario 1932:1-2001.12.

b) Si ripetano i calcoli riportati nella tabella 14.7 (14.6 nella 4° ed.) usando le regressioni stimate per il periodo campionario 1932:1-2001:12.

c) La variabile ln(dividend yield) è molto persistente?

d) Si costruiscano pseudo previsioni fuori campione dei rendimenti in eccesso per il periodo 1981:1-2001:12 usando le regressioni che iniziano nel 1932:1.

e) I risultati (a) e (d) suggeriscono cambiamenti importanti delle conclusioni a cui si era giunti nei riquadri? Si argomenti la risposta.

ESERCIZIO EMPIRICO 15.2.

In questo esercizio analizziamo due serie aggregate di prezzi per gli Stati Uniti: il CPI e il PCED contenuti nel file USMacro_Monthly.xls.

Con il primo (Consumer price index) si intende il prezzo di un paniere di beni la cui composizione viene aggiornata ogni 5/10 anni; con il secondo (Personal Consumption Expenditures Deflator) invece si intende il prezzo di un paniere di beni che cambia ogni mese.

Le serie di dati comprendono gli anni dal 1965 al 2004, per un totale di 480 osservazioni.

Si calcolino preliminarmente i tassi di inflazione, sulla base dei due indici di prezzi al consumo le cui serie temporali sono disponibili nel file di dati, ovvero CPI e PCED. Le si denotino π_t^{CPI} e π_t^{PCED}, rispettivamente.

a) Si calcoli le media campionarie di π_t^{CPI} e π_t^{PCED}. Queste stime puntuali sono coerenti con la presenza di distorsione da sostituzione nel CPI?

b) Si calcoli la media campionaria di Y_t. Si spieghi perché essa è numericamente uguale alla differenza tra le medie calcolate nella (a).

c) Si mostri che la media di Y è uguale alla differenza delle medie dei due tassi di inflazione.

d) Si consideri la regressione “con la sola costante” $Y_t = \beta_0 + u_t$. Si mostri che $\beta_0 = E(Y)$. Si ritiene che u_t sia serialmente correlato? Si spieghi.

e) Si costruisca un intervallo di confidenza al 95% per β_0. Quale valore del parametro standard di troncamento m si è scelto per la stima HAC? Perché?

f) C’è evidenza statisticamente significativa che il tasso di inflazione medio per il CPI sia maggiore di quello per il PCED?

ESERCIZIO EMPIRICO 16.4

Per questo esercizio, si faccia riferimento ai dati nel file USMacro_Monthly.xls. Si calcolino preliminarmente i tassi di inflazione, sulla base dei due indici di prezzi al consumo le cui serie temporali sono disponibili nel file di dati, ovvero CPI e PCED. Le si denotino π_t^{CPI} e π_t^{PCE}, rispettivamente.

a) Si verifichi la presenza di una radice unitaria nell'autoregressione per $Y_t \coloneqq \pi_t^{CPI} - \pi_t^{PCE}$. Si effettui la verifica usando il test ADF che include una costante e 12 differenze prime ritardate di Y_t.

b) Si verifichi la presenza di una radice unitaria nell'autoregressione per π_t^{CPI} e in quella per π_t^{PCE}. Come in (a) si usi il test ADF e, volendo, il test DF-GLS, includendo una costante e 12 differenze prime ritardate.

c) Che cosa dicono i risultati di (a) e di (b) circa la cointegrazione tra questi due tassi di inflazione? In particolare, si denotino come θ_1 e θ_2 i coefficienti di (eventuale) cointegrazione di π_t^{CPI} e π_t^{PCE}, ovvero i coefficienti della combinazione lineare $z_t = \theta_1 \pi_t^{CPI} + \theta_2\, \pi_t^{PCE}$ dei due tassi di inflazione tali per cui z_t sia I(0). Per comodità, si riscriva la ipotizzata relazione di cointegrazione come segue

$$\pi_t^{CPI} = \frac{\theta_2}{\theta_1}\pi_t^{PCE} + \frac{z_t}{\theta_1} = \theta\pi_t^{PCE} + u_t, \qquad \theta := \frac{\theta_2}{\theta_1}, u_t := \frac{z_t}{\theta_1}$$

Qual è il valore del coeffciente di cointegrazione θ suggerito dalle risposte ai punti (a) e (b)?

d) Si supponga di non sapere che il coefficiente di cointegrazione sia $\theta = 1$. Come si potrebbe stimare θ e verificare la cointegrazione secondo la metodologia di Engle e Granger.

e) Opzionale (per approfondire): Si stimi il valore di θ usando la regressione DOLS di π_t^{CPI} su π_t^{PCE} e sei anticipi e ritardi di $\Delta\pi_t^{PCE}$. il valore stimato di θ è vicino a 1?

DOMANDE E RISPOSTE

ESERCIZIO EMPIRICO 14.7.

a) Si ripetano i calcoli riportati nella tabella 14.3 usando le regressioni stimate per il periodo campionario 1932:1-2001.12.

L'esercizio si focalizza sul rendimento mensile di un indice di mercato in eccesso al tasso privo di rischio, che in generale è ciò che si guadagna in termini percentuali acquistando un titolo alla fine del mese precedente e vendendolo alla fine del mese corrente, sottraendo ciò che si sarebbe guadagnato se si fosse realizzato un investimento sicuro. Il rendimento del titolo include il guadagno (o la perdita) in conto capitale per la variazione di prezzo, più i dividendi ricevuti nel mese.

I risultati sono riportati nella ***Tabella 14.a*** a fine esercizio.

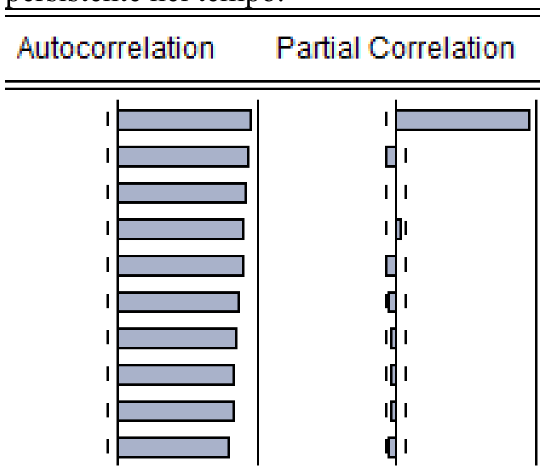
L' $\overline{\mathrm{R}}^2$ dei modelli sono molto vicini a zero e ciò suggerisce che nessuno di questi modelli è utile alla previsione. Il coefficiente dei rendimenti ritardati nel modello AR(1) non è statisticamente significativo (in valore assoluto 1,50 è minore di 1,96). Lo stesso ragionamento può essere fatto osservando le Statistiche F associate ad AR(2) e AR(3): i p-values sono maggiori di 0,10 quindi non si può rifiutare l'ipotesi nulla che tutti i coefficienti dei regressori siano pari a zero.

b) Si ripetano i calcoli riportati nella tabella 14.7 usando le regressioni stimate per il periodo campionario 1932:1-2001:12.

I risultati sono riportati nella ***Tabella 14.b*** a fine esercizio.

c) La variabile ln(dividend yield) è molto persistente?

Dal correlogramma riportato di seguito si evince che la variabile ln(dividend yield) è molto persistente nel tempo.



AC	Prob
0,987	0,000
0,972	0,000
0,957	0,000
0,943	0,000
0,927	0,000
0,910	0,000
0,892	0,000
0,874	0,000
0,855	0,000
0,834	0,000

Figura 1. La colonna denominata AC contiene i coefficienti di autocorrelazione, ottenuti facendo come $\hat{\rho} = Cov(y_t, y_{t-s})/Var(y_{t-s})$, dove s va da 1 a 10. La colonna denominata Prob indica invece i p-value del test dell'ipotesi nulla che la autocorrelazione sia pari a zero contro l'alternativa che sia diversa da zero.

d) Si costruiscano pseudo previsioni fuori campione dei rendimenti in eccesso per il periodo 1981:1-2001:12 usando le regressioni che iniziano nel 1932:1.

Sulla base della regressione AR(1) stimata ottenuta al punto (a), ovvero $exreturn_t = c + exreturn_{t-1}$, sono stati stimati i rendimenti in eccesso previsti in-sample e out of sample.

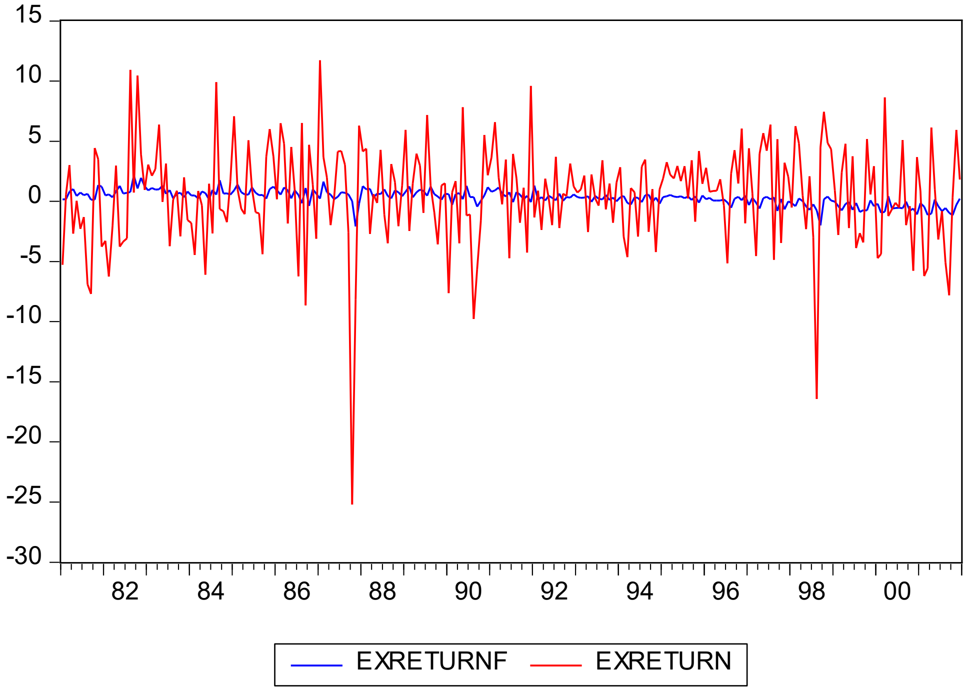


Figura 2. Il grafico mostra le previsioni fuori campione (ExreturnF) con le osservazioni effettive del rendimento in eccesso atteso (Exreturn).

e) I risultati (a) e (d) suggeriscono cambiamenti importanti delle conclusioni a cui si era giunti nei riquadri? Si argomenti la risposta.

Osservando i risultati ottenuti al punto (d) non possiamo che concordare con le conclusioni raggiunte nella discussione della domanda (a): è evidente la difficoltà di ottenere stime attendibili dei rendimenti futuri col modello di regressione utilizzato.

TABELLE ESERCIZIO 14.7.

	(1)	(2)	(3)
	AR(1)	**AR(2)**	**AR(4)**
Exreturn(-1)	0,10	0,10	0,10
	(0,10)	(0,10)	(0,09)
Exreturn(-2)		-0,04	-0,03
		(0,46)	(0,56)
Exreturn(-3)			-0,10
			(0,06)
Exreturn(-4)			0,01
			(0,84)
Intercetta	0,55	0,58	0,63
Wald F-statistic	2,64	1,55	1,55
	(0,10)	(0,21)	(0,19)
SER	5,13	5,13	5,10
R^2	0,010	0,012	0,023
Adjusted R^2	0,0089	0,0095	0,018
n	840	840	840

Tabella 14.a. I rendimenti in eccesso sono misurati in percentuale su base mensile. Le regressioni sono stimate per il periodo 1965:1 – 2004:12 (T=840 osservazioni). La tabella riporta tre diverse regressioni effettuate su campioni con n = 840 osservazioni: il primo campione nella colonna (AR1) riporta l'autoregressione di primo ordine del rendimento in eccesso atteso. Nella colonna (AR2) si riporta l'autoregressione di secondo ordine del rendimento in eccesso atteso; la colonna (AR4) riporta l'autoregressione di quarto ordine del rendimento in eccesso atteso. Per ciascun valore dei regressori viene riportato in parentesi il p-value associato. Si riportano inoltre: la statistica F di Wald per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati ($\boldsymbol{R^2}$ e adjusted $\boldsymbol{R^2}$).

	(1)	(2)	(3)
	ADL(1,1)	**ADL(2,2)**	**ADL(1,1)**
Exreturn(-1)	0,009	0,105	0,106
	(0,53)	(0,40)	(0,08)
Exreturn(-2)		-0,098	
		(0,52)	
Δln(dividend Yield-1)	-0,01	0,0001	
	(0,92)	(0,99)	
Δln(dividend Yield-2)		-0,056	
		(0,67)	
ln(dividend Yield-1)			0,011
			(0,09)
Intercetta	0,56	0,60	4,33
Wald F-statistic	1,34	0,82	2,65
	(0,26)	(0,51)	(0,07)
SER	5,13	5,13	5,11
R^2	0,010	0,012	0,017
Adjusted R squared	0,0077	0,0075	0,014
n	840	840	840

Tabella 14.b. Nella prima colonna abbiamo effettuato una regressione della variabile rendimento in eccesso sul suo ritardo e sul ritardo della variabile $\mathbf{\Delta ln(dividend\ Yield)}$; nella seconda colonna abbiamo considerato due ritardi di entrambe le variabili, mentre nella terza colonna abbiamo considerato solamente un ritardo della variabile rendimento in eccesso e il ritardo primo della variabile $\mathbf{ln(dividend\ Yield)}$. Per ciascun valore dei regressori viene riportato in parentesi il p-value associato. Si riportano inoltre: la statistica F di Wald per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati ($\boldsymbol{R^2}$ e adjusted $\boldsymbol{R^2}$).

ESERCIZIO EMPIRICO 15.2.

In questo esercizio analizziamo due serie aggregate di prezzi per gli Stati Uniti: il CPI e il PCED. Con il primo (Consumer price index) si intende il prezzo di un paniere di beni la cui composizione viene aggiornata ogni 5/10 anni; con il secondo (Personal Consumption Expenditures Deflator) invece si intende il prezzo di un paniere di beni che cambia ogni mese.
Le serie di dati comprendono gli anni dal 1965 al 2004, per un totale di 480 osservazioni.
I tassi di inflazione medio mensile basato sul CPI e sul PCED sono, rispettivamente, i seguenti:

$$\pi_t^{CPI} = 1200 \times \frac{CPI_t}{CPI_{t-1}}$$

$$\pi_t^{PCED} = 1200 \times \frac{PCED_t}{PCED_{t-1}}$$

Gli istogrammi che raffigurano la distribuzione empirica delle serie così ottenute sono riportati, nell'ordine, di seguito:

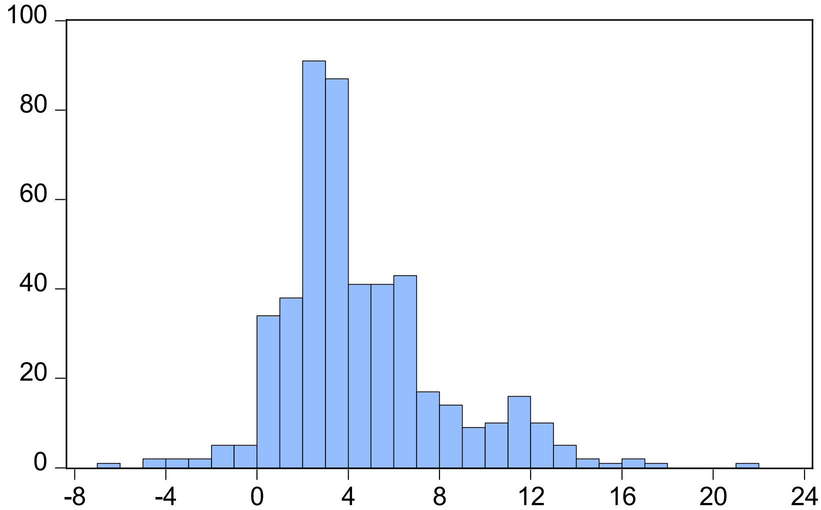


Figura 3. Istogramma del tasso d'inflazione medio mensile basato sul CPI

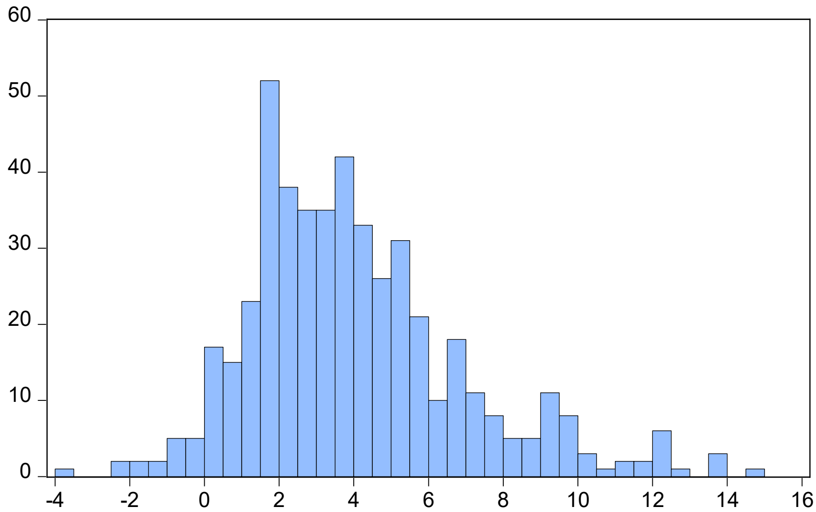


Figura 4. Istogramma del tasso d'inflazione medio mensile basato sul PCE

a) Si calcolino le medie campionarie di π_t^{CPI}e π_t^{PCED}. Queste stime puntuali sono coerenti con la presenza di distorsione da sostituzione nel CPI?
Dato il tasso di inflazione medio mensile basato sul CPI, ne calcoliamo la media campionaria che è 4,53.

Dato il tasso mensile di inflazione dei prezzi basato sul PCED, ne calcoliamo la media campionaria che è 4,03.

La differenza tra queste medie è 4,53 – 4,03 = 0,50 e dunque non è nulla. Questa differenza porta alcuni economisti a ritenere la stima basata su CPI distorta rispetto a quella basata su PCED motivandola come una cosiddetta distorsione da sostituzione nel CPI. Ovviamente, prima di giungere ad una conclusione al riguardo, bisognerà valutare se questa differenza è statisticamente significativa.

b) Si calcoli la media campionaria di Y_t. Si spieghi perché essa è numericamente uguale alla differenza tra le medie calcolate nella (a).

La differenza tra i due tassi di inflazione è

$$Y_t = \pi_t^{CPI} - \pi_t^{PCED}$$

L'istogramma della serie storica di questa variabile (la differenza tra i due tassi) è riportata in Figura 5.

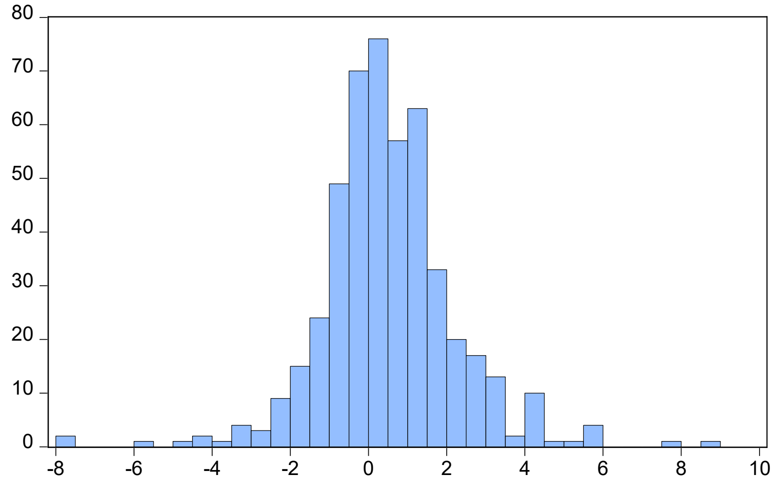


Figura 5. Distribuzione a istogrammi della media campionaria di Y_t come differenza dei due tassi mensili di inflazione.

La media campionaria della differenza dei due tassi di inflazione, ovvero di Y_t, è pari a 0,50. È numericamente uguale alla differenza tra le medie calcolata nel punto (a) perché l'operatore media è lineare.

$$\bar{Y}_t = \overline{\pi_t^{CPI} - \pi_t^{PCED}} = \overline{\pi_t^{CPI}} - \overline{\pi_t^{PCED}} = 4{,}53 - 4{,}03 = 0{,}50$$

Si noti che qui si denota con la barretta la media campionaria delle variabili che appaiono sotto la barretta stessa. Ad esempio, $\bar{Y}_t := \frac{1}{T}\sum_{t=1}^{T} Y_t$, dove T denota l'ampiezza campionaria, piuttosto che $E(Y_t)$.

c) Si mostri che la media di Y_t è uguale alla differenza delle medie dei due tassi di inflazione.

In questa domanda, per media si intende il valore atteso invece della media campionaria. Poiché l'operatore valore atteso è lineare come la media aritmetica (e dunque quella campionaria), la risposta è quella data nel punto (b),

$$E(Y_t) = E(\pi_t^{CPI} - \pi_t^{PCED}) = E(\pi_t^{CPI}) - E(\pi_t^{PCED})$$

d) Si consideri la regressione "con la sola costante" $Y_t = \beta_0 + u_t$. Si mostri che $\beta_0 = E(Y)$. Si ritiene che u_t sia serialmente correlato? Si spieghi.

Consideriamo la regressione con la sola costante $Y_t = \beta_0 + u_t$. Abbiamo che $E(Y_t) = E(\beta_0 + u_t) = \beta_0 + E(u_t)$ poiché $E(u) = 0$.

Infatti, dalla stima della regressione si nota anche come il valore della stima del coefficiente β_0 è 0,50, identico alla media campionaria Y_t.
La presenza di correlazione seriale può essere rilevata guardando il correlogramma e i test dell'ipotesi nulla sulla assenza della stessa, tra cui il test basato sulla statistica Q di Ljung-Box.

e) Si costruisca un intervallo di confidenza al 95% per β_0. [Si ignori, a meno che non particolarmente interessati, la parte restante della domanda, ovvero: Quale valore del parametro standard di troncamento m si è scelto per la stima HAC? Perché?]
Applicando lo stimatore HAC, il valore del coefficiente β_0 non cambia, mentre varia la stima del suo errore standard, che passa da 0,08 a 0,09, e di conseguenza l'intervallo di confidenza è:

$$\beta_0 \pm 1{,}96 \times \sigma(\beta_0) = 0{,}50 \pm 1{,}96 \times 0{,}09 = 0{,}50 \pm 0{,}18$$

Il parametro di troncamento m è scelto per essere grande in grandi campioni ma comunque molto inferiore rispetto a T. Una regola pratica è quella di utilizzare la seguente formula:

$$m = 0{,}75\,\frac{T}{3} = 5{,}56 \approx 6$$

f) C'è evidenza statisticamente significativa che il tasso di inflazione medio per il CPI sia maggiore di quello per il PCED?
Qui per tassi di inflazione medi si intendono i valori attesi di tali tassi, che sono stati stimati tramite le rispettive medie campionarie. Dunque, ci si chiede se la media campionaria di Y_t, ovvero della differenza tra i due tassi, offra evidenza statisticamente significativa contro l'ipotesi nulla che i valori attesi dei due tassi siano uguali. Questa ipotesi nulla, in virtù di quanto detto, può essere espressa come la ipotesi che il valore atteso di Y_t sia zero.
L'intervallo di confidenza appena costruito ci consente di sottoporre a verifica questa ipotesi nulla e di rigettarla al livello di significatività del 5% poiché il valore di zero (ovvero il valore dell'aspettativa di Y_t) è al di fuori dell'intervallo di confidenza al 95%.
Intuitivamente, possiamo dire che l'intervallo di confidenza costruito sopra ci suggerisce che possiamo essere al 95% confidenti che β_0 sia compreso tra 0,32 e 0,68 punti percentuali, e dunque in un intervallo che non include lo zero. Dunque, possiamo rigettare al livello di significatività del 5% (a due code) l'ipotesi nulla che i due tassi siano uguali.

ESERCIZIO EMPIRICO 16.4.

a. Si verifichi la presenza di una radice unitaria nell'auto-regressione per $Y_t = \pi_t^{CPI} - \pi_t^{PCE}$. Si effettui la verifica usando il test ADF che include una costante e 12 differenze prime ritardate di Y_t. [Si ignori la parte rimante della domanda, ovvero: "Si effettui inoltre il test usando la procedura DF-GLS", a meno che non fortemente interessati].

Per verificare la presenza di una radice unitaria possiamo utilizzare il test ADF, utilizzando 12 differenze prime ritardate, ovvero *lags*, e una costante. Tale test, come mostra la ***Tabella 16.a***, rigetta l'ipotesi nulla che la variabile Y_t (cioè $\pi_t^{CPI} - \pi_t^{PCE}$) contenga un trend stocastico, ovvero l'ipotesi nulla che non sia stazionaria.

b. Si verifichi la presenza di una radice unitaria nell'autoregressione per π_t^{CPI} e in quella per π_t^{PCE}. Come in (a) si usino sia il test ADF che (opzionalmente) il test DF-GLS, includendo una costante e 12 differenze prime ritardate.

Implementiamo i test ADF per verificare l'esistenza di un trend stocastico in π_t^{CPI} e π_t^{PCE}.
Nelle righe seconda e terza della ***Tabella 16.a*** sono mostrati i test riguardo alla variabile π_t^{CPI} e alla variabile π_t^{PCE}, rispettivamente, includendo 12 differenze prime ritardate nella specificazione del processo stocastico di queste variabili stimata ai fini del test. In entrambi i casi non possiamo rigettare al livello di significatività almeno del 5% l'ipotesi nulla di non stazionarietà e dunque accettiamo l'alternativa che le variabili contengano trend stocastici (ovvero che abbiano ordine di integrazione 1).

c. Che cosa dicono i risultati di (a) e di (b) circa la cointegrazione tra questi due tassi di inflazione? Qual è il valore del coefficiente di cointegrazione (θ) suggerito dalle risposte ai punti (a) e (b)?

Dalle risposte ai punti (a) e (b), vi è evidenza che:

- Le due variabili π_t^{CPI} e π_t^{PCE} hanno ordine di integrazione 1,
- Una loro combinazione lineare, ovvero $\pi_t^{CPI} - \pi_t^{PCE}$, è stazionaria. Ciò vuol dire che, in $\pi_t^{CPI} = \theta\pi_t^{PCE} + u_t$, $\theta = 1$.

Possiamo concludere che vi sia evidenza che le due variabili siano cointegrate e che il coefficiente di cointegrazione θ sia pari a 1.

d) Si supponga di non sapere che il coefficiente di cointegrazione è $\theta = 1$. Come si potrebbe stimare θ e verificare la (assenza di) cointegrazione secondo la metodologia di Engle e Granger.

La cointegrazione si può verificare con la procedura di Engle-Granger e il relativo test ADF con apposita distribuzione della statistica del test.
Come primo step stimiamo col metodo OLS una regressione che descrive la ipotizzata cointegrazione, ovvero:

$$\pi_t^{CPI} = \alpha + \theta\pi_t^{PCE} + u_t$$

Come mostrato dalla ***Tabella 16.b*** si ottiene un coefficiente di cointegrazione stimato pari a 1,09, un valore molto prossimo a 1, oltre a una intercetta non statisticamente differente da zero.
Il secondo step consiste nell'effettuare un test *t*, simile a un test ADF (Augmented Dickey- Fueller) ma con distribuzione apposita, sui residui di tale regressione stimata. Attraverso l'implementazione di questo test, il cui risultato è riportato nella terza riga della ***Tabella 16.a***, è possibile rigettare l'ipotesi nulla che gli errori della regressione stimata abbiano un trend stocastico e di conseguenza si accetta l'alternativa che le due variabili π_t^{CPI} e π_t^{PCE} siano cointegrate.

TABELLE ESERCIZIO EMPIRICO 16.4

Augmented Dickey Fuller Test

Variable	t-statistic	Prob
Y_t	**-18,57**	0,00
π_t^{CPI}	**-2,46**	0,12
π_t^{PCE}	**-2,80**	0,06
$\hat{u}_t$	**-20,75**	0,00
Livelli di significatività	Valori critici della statistica del test	
5%	-2,87	
1%	-3,44	

Tabella 16.a. Nella tabella si riporta la statistica del test di Dickey – Fuller Aumentato (ADF) con 12 ritardi dell'ipotesi nulla che ci sia una radice unitaria nel processo della varabile indicata nella prima colonna, il relativo valore-p (Prob.) e i valori critici della statistica del test con intercetta.

Variabile dipendente: π_t^{CPI}	
Intercetta	0,16
	(0,24)
π_t^{PCE}	1,09
	(0,00)

Tabella 16.b. Nella tabella viene mostrata la regressione, con intercetta, di π_t^{CPI} su π_t^{PCE}. Si riportano i coefficienti dei regressori e, tra parentesi, i valori-p del test della ipotesi nulla che i coefficienti siano pari a zero.