Introduzione a regressioni temporali e previsioni

- Definizione di serie temporali.

- Domande le cui risposte necessitano dell’analisi di serie temporali:
  – Qual è l’effetto causale su una variabile di interesse Y di un cambiamento di un’altra variabile X nel corso del tempo?
  – Qual è la migliore previsione che si può fare circa il valore di una certa variabile a una data futura?

L’uso dei modelli di regressione per la previsione

- Consideriamo il seguente problema: un genitore che si sposta in un’area metropolitana nuova deve scegliere dove vivere anche in relazione all’offerta scolastica.

- Il problema è di prevedere la prestazione media in un dato distretto sulla base di informazioni correlate con i punteggi del test (dimensioni classi).

- La regressione già vista $Testscore = 698.9 - 2.28 * STR$ anche se presenta uno stimatore della pendenza distorto (perché?) può essere utilizzata dal genitore per scegliere il distretto.

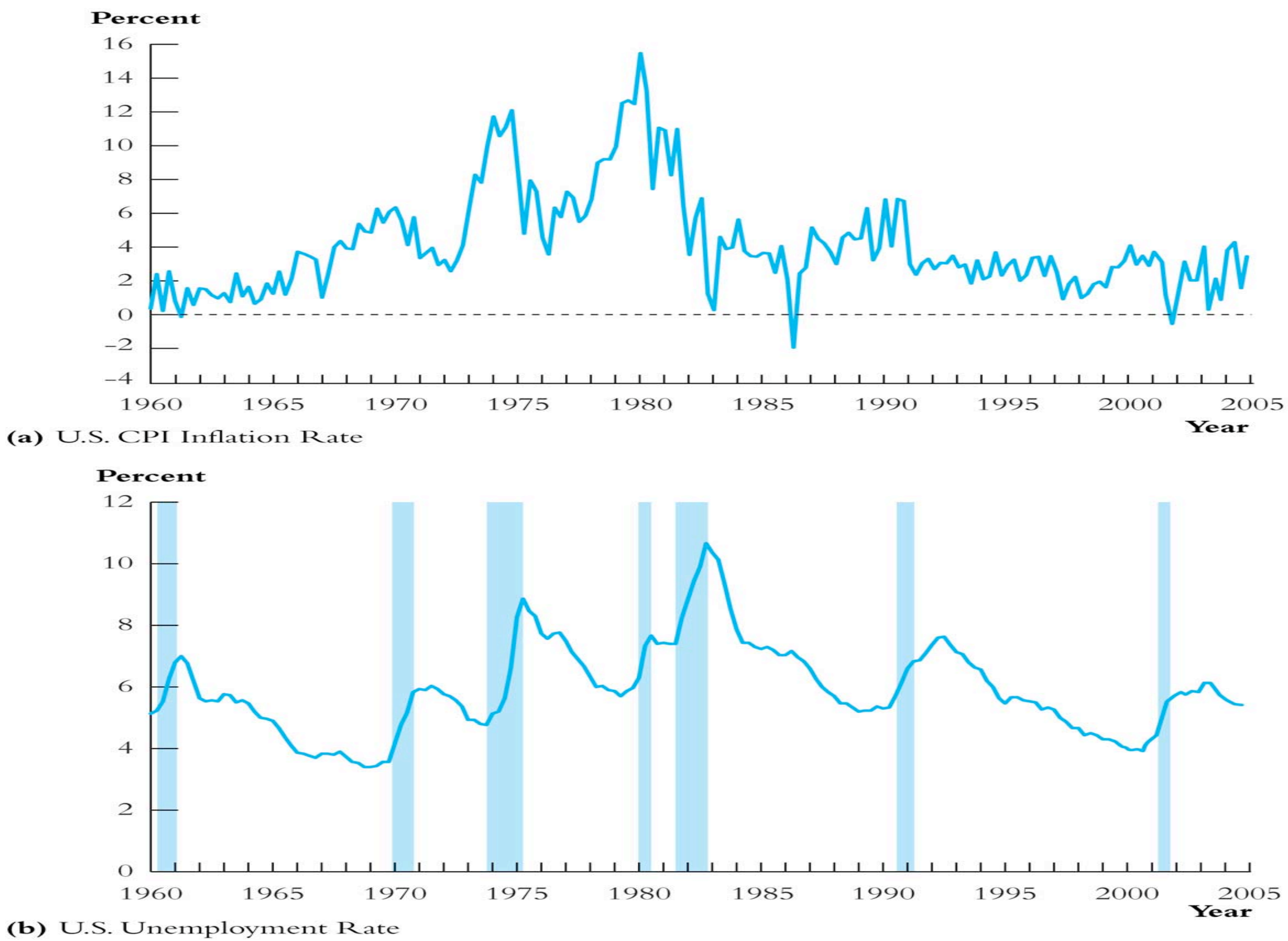
- Al genitore non interessa la bontà della stima dell'effetto causale della grandezza delle classi sui punteggi del test. Ciò di cui ha bisogno è che la regressione spieghi buona parte della variabilità nei punteggi tra i distretti e che sia valida per i distretti che a lui interessano.

- L'obiettivo è quello di utilizzare valori noti di alcune variabili per prevedere il valore di un'altra variabile.

Serie temporali e correlazione seriale

- Un esempio di serie temporali è la relazione tra il tasso d'inflazione e di disoccupazione negli USA.
- Le due figure rappresentano la variazione percentuale annuale del livello dei prezzi negli USA, misurato dall'indice dei prezzi al consumo dal 1960 al 1999, e la frazione della forza lavoro senza un'occupazione.

FIGURE 14.1 Inflation and Unemployment in the United States, 1960–2004



Price inflation in the United States (Figure 14.1a) drifted upward from 1960 until 1980, and then fell sharply during the early 1980s.The unemployment rate in the United States (Figure 14.1b) rises during recessions (the shaded episodes) and falls during expansions.

Terminologia

- Il valore di Y nel periodo precedente è chiamato il suo primo ritardo ed è indicato con Y_{t-1}. In modo analogo Y_{t+1} indica il valore di Y un periodo avanti nel tempo.
- La variazione del valore Y tra il periodo t-1 e il periodo t è $Y_t - Y_{t-1}$. Tale variazione è detta differenza prima della variabile Y_t: $\Delta Y_t = Y_t - Y_{t-1}$.
- La variazione percentuale di una serie temporale Y_t tra i periodi t-1 e t è approssimativamente $100\Delta \ln(Y_t)$, dove l'approssimazione è più accurata quando la variazione percentuale è più piccola.
- Un esempio di ritardi, variazioni e variazioni percentuali è illustrato dalla seguente tabella

TABLE 14.1 Inflation in the United States in 2004 and the First Quarter of 2005

Quarter	U.S. CPI	Rate of Inflation at an Annual Rate (Inf_t)	First Lag (Inf_{t-1})	Change in Inflation (ΔInf_t)
2004:I	186.57	3.8	0.9	2.9
2004:II	188.60	4.4	3.8	0.6
2004:III	189.37	1.6	4.4	−2.8
2004:IV	191.03	3.5	1.6	1.9
2005:I	192.17	2.4	3.5	−1.1

The annualized rate of inflation is the percentage change in the CPI from the previous quarter to the current quarter, times four. The first lag of inflation is its value in the previous quarter, and the change in inflation is the current inflation rate minus its first lag. All entries are rounded to the nearest decimal.

Autocorrelazione

- Nelle serie temporali il valore di Y in un periodo è solitamente correlato con il suo valore nel periodo successivo.
- La correlazione di una serie con i propri valori ritardati è detta autocorrelazione.
- La prima autocorrelazione o primo coefficiente di correlazione è quella tra Y_t e Y_{t-1}.

$$autocovarianza = \operatorname{cov}\left(Y_t, Y_{t-j}\right)$$

$$autocorrelazione = \rho_j = corr\left(Y_t, Y_{t-j}\right) = \frac{\operatorname{cov}\left(Y_t, Y_{t-j}\right)}{\sqrt{\operatorname{var}\left(Y_t\right)\operatorname{var}\left(Y_{t-j}\right)}}$$

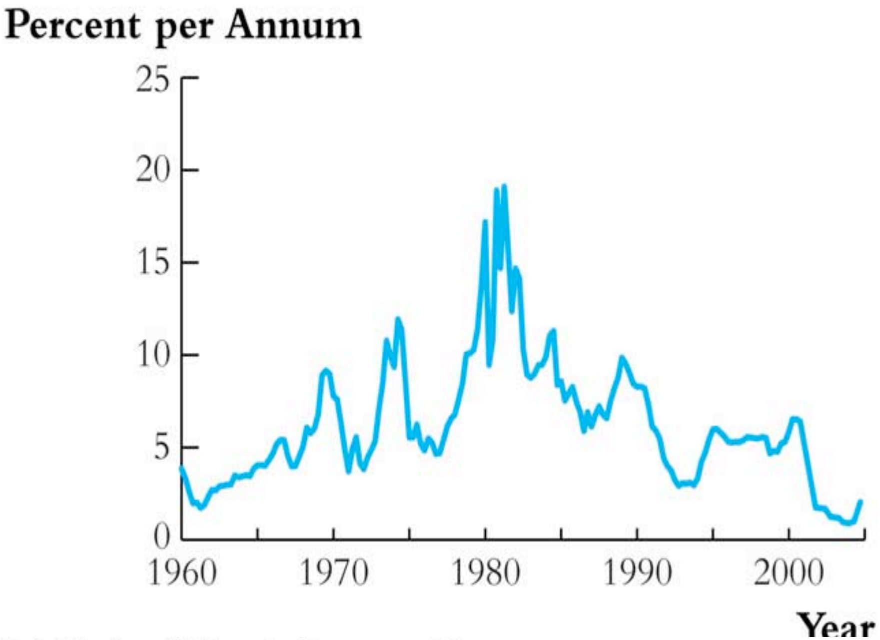
TABLE 14.2 First Four Sample Autocorrelations of the U.S. Inflation Rate and Its Change, 1960:I–2004:IV

	Autocorrelation of:	
Lag	Inflation Rate (Inf_t)	Change of Inflation Rate (ΔInf_t)
1	0.84	−0.26
2	0.76	−0.25
3	0.76	0.29
4	0.67	−0.06

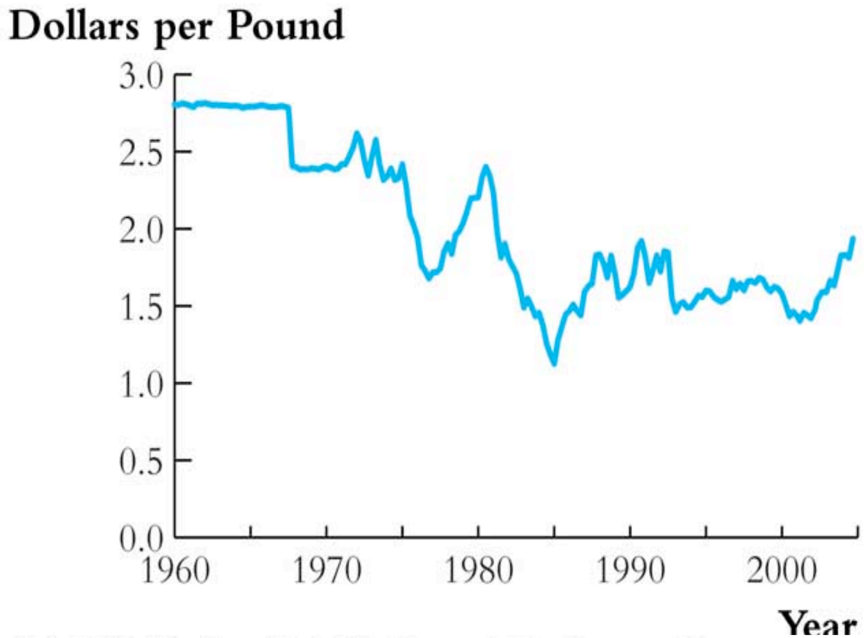
- La forte autocorrelazione positiva nell'inflazione riflette i movimenti di lungo periodo dell'inflazione
- Essa era bassa nel primo trimestre del 1965 e nel secondo; mentre era alta nel primo trimestre de 1981 e nel secondo.
- Diversamente l'autocorrelazione negativa della variazione dell'inflazione significa che in media un aumento dell'inflazione in un trimestre è associato con una diminuzione dell'inflazione nel trimestre successivo.

Esempi

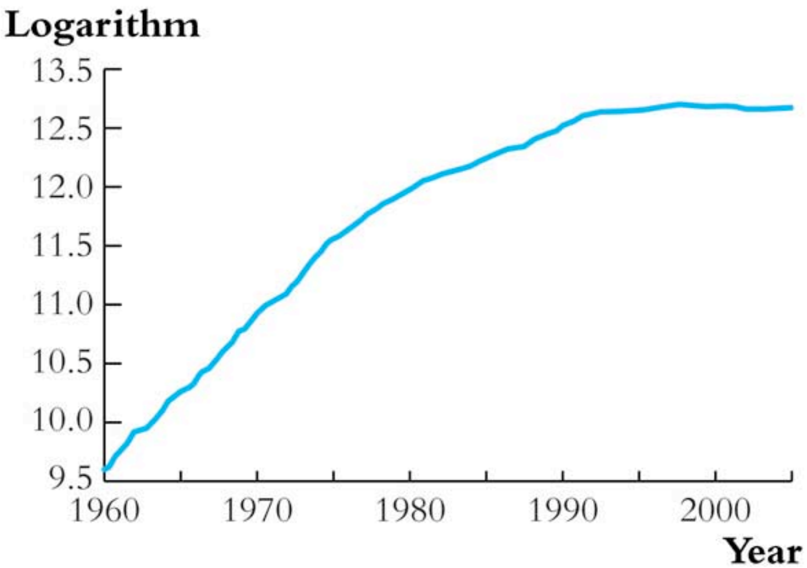
- Altri esempi di serie temporali sono
  - Il tasso sui Federal Funds (tasso d'interesse delle banche USA pagato ad altre banche per prestiti overnight)
  - Il tasso di cambio dollaro-sterlina
  - Il PIL reale del Giappone su base trimestrale
  - Il rendimento giornaliero dell'indice NYSE dei prezzi delle azioni



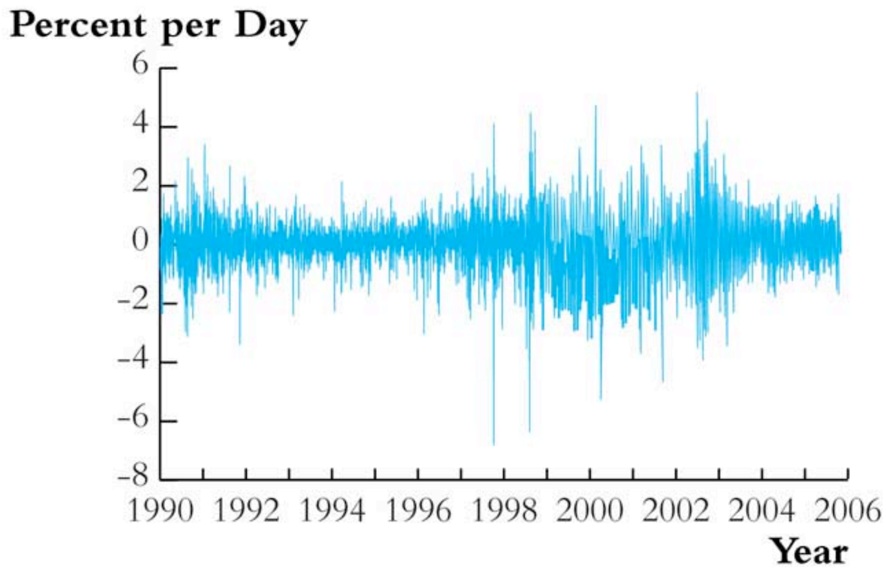
(a) Federal Funds Interest Rate



(b) U.S. Dollar/British Pound Exchange Rate



(c) Logarithm of GDP in Japan



(d) Percentage Changes in Daily Values of the NYSE Composite Stock Index

- Le 4 serie mostrate hanno comportamenti molto diversi.
- La prima richiama l’andamento del tasso d’inflazione.
- La seconda mostra una discontinuità dopo il crollo del sistema di cambio fissi di BW (1972)
- La terza è misurata il ln e mostra una crescita regolare ancorché decrescente
- La quarta mostra l’imprevedibilità della variabile analizzata con una varianza stazionaria.

Autoregressioni

- Un autoregressione è un modello di regressione che mette in relazione una variabile temporale con i suoi valori passati.
- Un modo sistematico per prevedere la variazione dell'inflazione, Δinf_t, usando la variaazione del trimestre precedente, Δinf_{t-1}, è quello di fare la seguente regressione tramite gli OLS:

$$\Delta \hat{I}nf_t = \underset{(0.14)}{0.02} - \underset{(0.106)}{0.211} \Delta Inf_{t-1}$$

- Tale modello è detto autoregressione del primo ordine poiché viene usato un solo ritardo come regressore.

- Più in generale tale modello, detto AR(1), è scritto

$$Y_t=\beta_0+\beta_1 Y_{t-1}+u_t$$

- Usando gli OLS otteniamo la previsione di Y_t basata sul modello AR(1): $\hat{Y}_{t|t-1} = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1}$

- L'errore di previsione è $Y_t - \hat{Y}_{t|t-1}$
- La previsione non è il valore predetto tramite gli OLS e l'errore di previsione non è il residuo degli OLS.
- Le previsioni e gli errori di previsione riguardano le osservazioni "fuori dal campione" mentre i valori predetti e i residui riguardano le osservazioni "dentro il campione"

Radice^2 dell’errore di previsione quadratico medio (RMSFE)

- È una misura dell’entità dell’errore di previsione.

$$RMSFE = \sqrt{E\left[(Y_t - \hat{Y}_{t|t-1})^2\right]}$$

- L’RMSFE considera due fonti di errore:
  - il fatto che i valori di u_t sono ignoti
  - l’errore che scaturisce dalla stima dei coefficienti della regressione

Esempio

- Quale previsione dell'inflazione nel primo trimestre del 2000 (2000:I) si sarebbe fatta nel 1999:IV, basandosi sul modello AR(1)?

$$\Delta\hat{I}nf_t = \underset{(0.14)}{0.02} - \underset{(0.106)}{0.211}\Delta Inf_{t-1}$$

- Usando i valori della tabella precedente, vediamo che il tasso d'inflazione in 1999:IV è ($Inf_{1999:IV}$=3.2%) con un aumento dello 0.4% rispetto al trimestre precedente ($\Delta inf_{1999:IV}$=0.4%).
- Sostituendo i valori nel modello, la previsione della variazione dell'inflazione dal 1999:IV al 2000:I è

$$\Delta I\hat{n}f_{2000:I} = 0.02 - 0.211 \times \Delta\,\mathrm{inf}_{1999:IV} = 0.02 - 0.211 \times 0.4 = -0.06 \cong -0.1$$

- Il tasso d'inflazione predetto è quindi:

$$\hat{Inf}_{t|t-1} = Inf_{t-1} + \Delta\hat{Inf}_{t|t-1}$$

- Il tasso d'inflazione predetto nel 2000:I è

$$\hat{Inf}_{2000:I} = Inf_{1999:IV} + \Delta\hat{Inf}_{2000:I} = 3.2\% - 0.1\% = 3.1\%$$

- Il valore attuale in 2000:I è stato 4.1%, quindi la previsione AR(1) è minore di un punto percentuale. Inoltre l'R^2 corretto è 0.04, quindi la previsione basata su un ritardo non riesce a spiegare la maggior parte della variabilità dell'inflazione nel campione considerato.

Il modello autoregressivo di ordine p

- Con questo modello è possibile tenere in considerazione le informazioni provenienti da un passato più remoto introducendo maggiori ritardi.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t, \text{ dove } E\left(u_t \middle| Y_{t-1}, Y_{t-2}, \ldots\right) = 0$$

- Ad esempio:

$$\Delta I\hat{n}f_t = \underset{(0.12)}{0.02} - \underset{(0.1)}{0.21}\Delta Inf_{t-1} - \underset{(0.09)}{0.32}\Delta Inf_{t-2} + \underset{(0.09)}{0.19}\Delta Inf_{t-3} - \underset{(0.1)}{0.04}\Delta Inf_{t-4}$$

- I coefficienti degli ultimi 3 ritardi sono congiuntamente significativi ($F_{5\%}$=6.43, valore-p<0.001) e l'R^2 corretto è 0.21

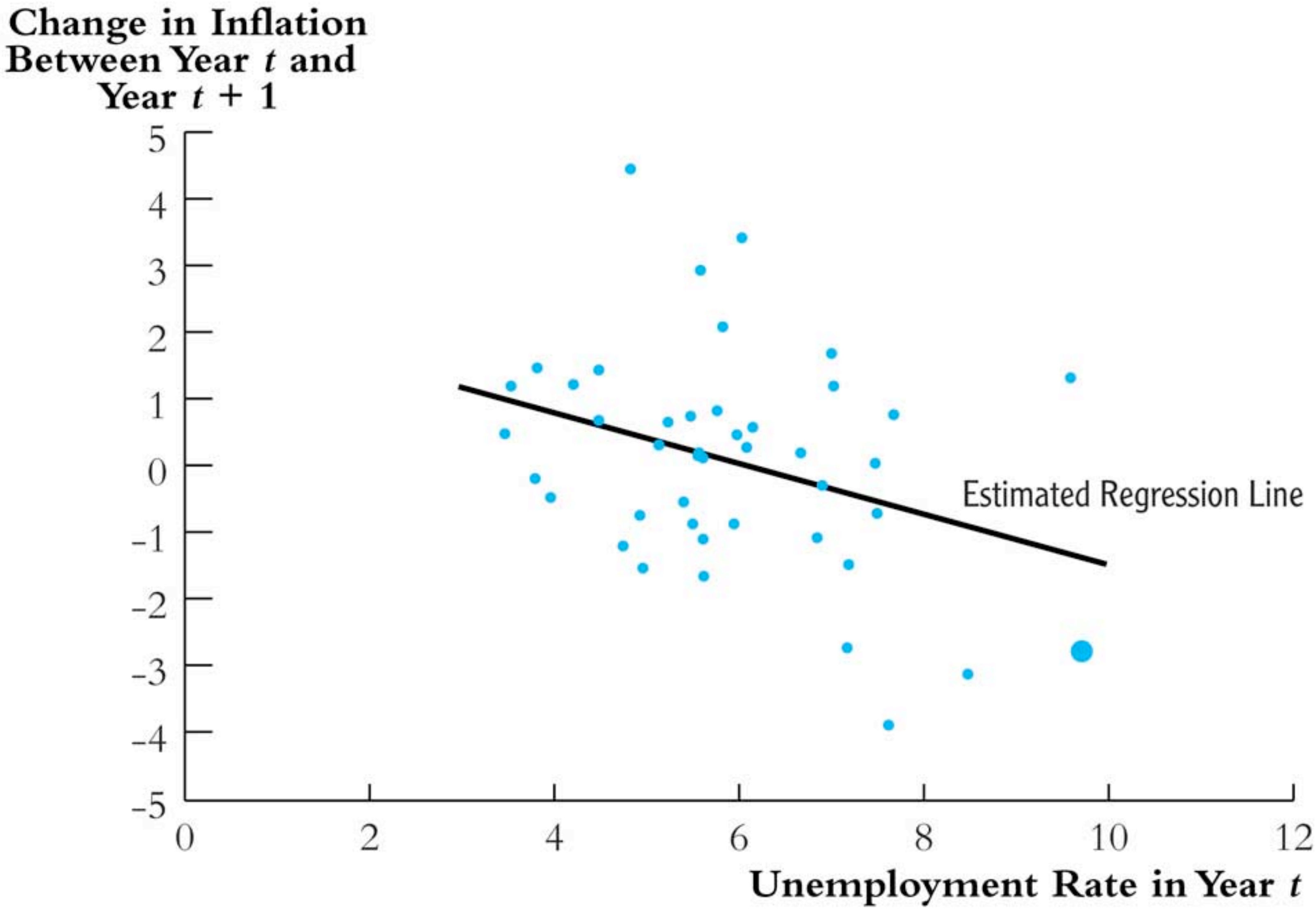
Regressioni temporali con predittori addizionali e modello autoregressivo misto

- È il caso i cui si usano anche altre variabili ad un modello autoregressivo.
- Ad esempio si può usare la disoccupazione per spiegare l'inflazione:

$$\Delta I\hat{n}f_t = \underset{(0.55)}{1.42} - \underset{(0.09)}{0.26}\Delta Inf_{t-1} - \underset{(0.1)}{0.40}\Delta Inf_{t-2} + \underset{(0.08)}{0.11}\Delta Inf_{t-3} - \underset{(0.1)}{0.09}\Delta Inf_{t-4}$$

$$-\underset{(0.1)}{0.23}Unemp_{t-1}, \overline{R}^2 = 0.22$$

- È la statistica t sul coefficiente della disoccupazione significativo?
- Vi è un miglioramento nel modello?


Change in Inflation Between Year t and Year t + 1
5
4
3
2
1
0
–1
–2
–3
–4
–5
0
2
4
6
8
10
12
Estimated Regression Line
Unemployment Rate in Year t

- La previsione dell’inflazione per il 2000:I ottenuta con la nuova regressione è 3.7% con un errore di previsione che è diminuito di 0.4%.
- Possiamo pensare di inserire ance qualche ritardo della disoccupazione nel modello.

$$\Delta \hat{Inf}_t = \underset{(0.47)}{1.32} - \underset{(0.09)}{0.36}\Delta Inf_{t-1} - \underset{(0.1)}{0.34}\Delta Inf_{t-2} + \underset{(0.08)}{0.07}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4}$$
$$- \underset{(0.47)}{2.68}\Delta Unemp_{t-1} + \underset{(0.89)}{3.43}\Delta Unemp_{t-2} - \underset{(0.89)}{1.04}\Delta Unemp_{t-3} + \underset{(0.44)}{0.07}\Delta Unemp_{t-4},$$
$$\overline{R}^2 = 0.35$$

- La statistica F sia per tutti i coefficienti della disoccupazione sia per i soli ritardi 2-4 è significativa. L’errore di previsione dell’inflazione nel 2000:I rimane piccolo, 0.4.

Il modello autoregressivo misto (ADL)

- In generale un modello autoregressivo misto con p ritardi della variabile dipendente e q ritardi di un predittore addizionale è indicato con ADL(p,q).

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \delta_2 X_{t-2} +$$

$$+ \ldots + \delta_q X_{t-q} + u_t, \text{ dove } E\left(u_t \middle| Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots\right) = 0$$

Assunzioni del modello di regressione temporale

- La prima assunzione è che u_t abbia media nulla condizionatamente a tutti i regressori ed ai ritardi addizionali dei regressori oltre a quelli già inseriti in regressione
- La seconda assunzione è formata da 2 parti
  - La prima è che i dati siano estratti da una distribuzione stazionaria (versione temporale di i.d.)
  - La seconda richiede che le variabili aleatorie diventino indipendentemente distribuite al crescere della distanza temporale che le separa (per grandi campioni vi sia sufficiente aleatorietà tra i dati per utilizzare la legge dei grandi numeri ed il teorema limite centrale)

- La terza assunzione è che tutte le variabili abbiano 4 momenti finiti diversi da zero

- La quarta è che i regressori non siano perfettamente collineari.

- Sulla base di queste assunzioni l’inferenza tramite gli OLS sui coefficienti della regressione non subisce variazioni

Test di causalità di Granger

- Un'applicazione della statisica F consiste nel verificare se i ritardi di uno dei regressori inclusi abbia un utile potere predittivo, aggiuntivo rispetto a quelli degli altri regressori del modello

- La causalità nel senso di Granger significa che se X causa Y allora X è un utile predittore per Y date altre variabili nella regressione

- Considerando la stima dell'inflazione, la statistica F su tutti i ritardi della disoccupazione siano nulli è 8.51 ($p<0.001$).

- Se da un lato si può dire che il tasso di disoccupazione causa, nel senso di Granger, variazioni nel tasso d'inflazione. Da un altro lato non si può dire che una variazione del tasso di disoccupazione causerà una conseguente variazione nel tasso d'inflazione.

Lunghezza dei ritardi

- Un primo approccio è quello della statistica F con il problema che molte volte si generino modelli troppo grandi.
- Un secondo approccio è il criterio d'informazione Bayesian (BIC):

$$BIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln(T)}{T}$$

- Il primo termine abbiamo che SSR decresce quando si aggiunge un ritardo. Diversamente, il secondo termine cresce quando viene aggiunto un ritardo.
- Il BIC bilancia queste due forze in maniera tale che il numero di ritardi che minimizza il BIC sia uno stimatore consistente dell'ordine dell'autoregressione.

- Un criterio alternativo è il criterio d'informazione di Akaike (AIC):

$$BIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$$

- Lo stimatore AIC di p non è consistente a causa della sostituzione di ln(T) con 2. Tuttavia tale criterio è molto diffuso nel caso in cui si teme che il BIC porti ad un modello con troppo pochi ritardi.
- Anche nel caso di predittori multipli si utilizzano gli stessi criteri.
- Due considerazioni:
  - Tutti i modelli devono essere stimati sullo stesso campione
  - Nel caso di predittori multipli, il BIC e l'AIC sono molto dispendiosi in quando richiedono il calcolo di molti modelli.

Esempio

# ritardi	BIC	AIC	R^2
0	1.095	1.076	0.000
1	1.067	1.030	0.056
2	**0.955**	0.900	0.181
3	0.957	**0.884**	0.203
4	0.986	0.895	0.204
5	1.016	0.906	0.204
6	1.046	0.918	**0.204**

- Il BIC suggerisce 2 ritardi, l'AIC invece 3.
- Se considerassimo l'R^2, sceglieremmo sempre il modello con il maggio numero di ritardi

Stazionarità

- Una serie temporale Y_t è stazionaria se la sua distribuzione di probabilità non cambia nel corso del tempo, cioè, se la distribuzione congiunta di $(Y_{s+1}, Y_{s+2},\ldots, Y_{s+T})$ non dipende da s.
- Diversamente la serie Y_t è detta non stazionaria.
- Due serie temporali sono dette congiuntamente stazionarie se la distribuzione congiunta di $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2},\ldots,X_{s+T}, Y_{s+T})$ non dipende da s.
- Il futuro sia come il passato in senso probabilistico.

I trend

- Si intende il movimento persistente di lungo periodo di una variabile nel corso del tempo. Si dice che una variabile temporale fluttua attorno al suo trend.
- Vi sono due tipi di trend:
  - Deterministico che è una funzione non aleatoria del tempo
  - Stocastico che è aleatorio e varia nel tempo
- Modelleremo le serie temporali economiche come se avessero un trend stocastico.

- Il più semplice modello di trend stocastico è la random walk (passeggiata aleatoria): $Y_t=Y_{t-1}+u_t$, dove l'errore è i.i.d.
- L'idea di base è che il valore di una serie domani è pari al valore di oggi più un imprevedibile cambiamento. Quindi la migliore previsione del valore di domani è il suo valore di oggi.
- In alcuni casi (ln(Pil) del Giappone) hanno una chiara tendenza verso l'alto. In tali casi la miglior previsione deve anche considerare un fattore aggiuntivo.
- Si ottiene un modello di random walk with drift (passeggiata aleatoria con deriva): $Y_t=\beta_0+ Y_{t-1}+u_{t.}$

- Se Y_t segue una random walk allora non è stazionaria: la varianza della random walk aumenta nel corso del tempo e così facendo cambia la distribuzione di Y_t.
- La random walk è un caso speciale del modello AR(1) in cui β_1=1. Quindi possiamo dire che se Y_t segue un AR(1) con β_1=1 allora Y_t contiene un trend stocastico ed è non stazionario.
- Se un AR(p) ha una radice uguale ad uno, si dice che la serie ha una radice autoregressiva unitaria o una radice unitaria. Se vi è una radice unitaria allora vi è un trend stocastico e la serie non è stazionaria.

Problemi causati dai trend stocastici Coefficienti autoregressivi che sono distorti verso zero

- Se la serie temporale che segue una random walk viene stimata come un modello AR(1) normale non varranno le assunzioni degli OLS per le serie temporali in quanto Y_t non è stazionaria.
- Lo stimatore OLS di β_1 è consistente ma ha una distribuzione non normale anche per grandi campioni.
- Quindi le previsioni basate sul modello AR(1) possono essere peggiori di quelle basata su una random walk

Distribuzioni non normali delle statistiche t

- Se un regressore ha un trend stocastico allora la statistica t degli OLS può avere una distribuzione non normale sotto l'ipotesi nulla anche per grandi campioni.
- Gli intervalli di confidenza non sono validi ed i test di ipotesi non possono essere condotti come in precedenza.

Regressione spuria

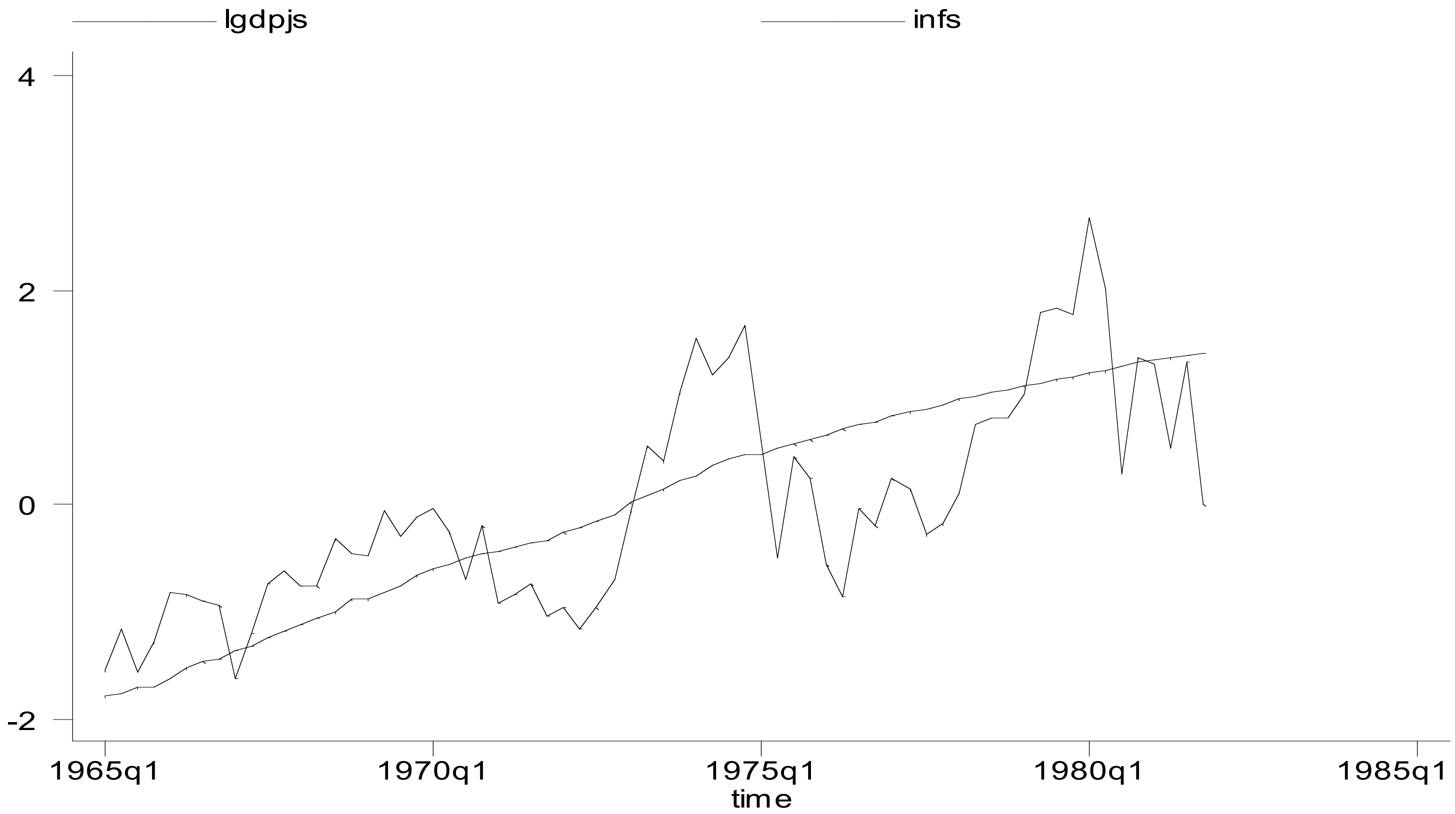
- I trend stocastici possono far sì che due serie temporali sembrino in relazione quando in realtà non lo sono.
- Consideriamo la seguente regressione relativa ai dati da metà degli anni sessanta agli inizi degli anni ottanta:

$$\widehat{U.S.Inflation}_t = \underset{(0.08)}{-2.84} + \underset{(0.02)}{0.18}\,JapanesePIL_t, \bar{R}^2 = 0.56$$

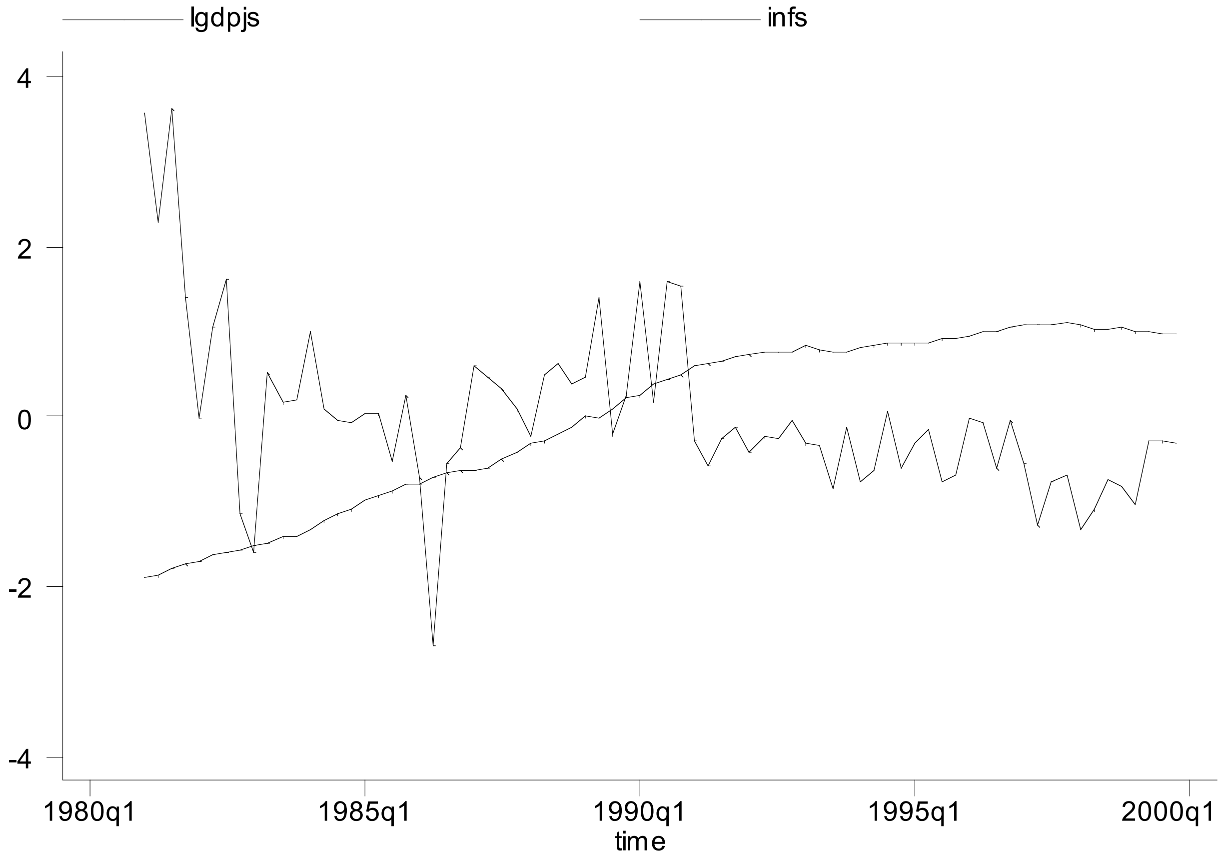
- La stessa regressione calcolata tra il 1982 ed il 1999 diventa:

$$\widehat{U.S.Inflation}_t = \underset{(1.37)}{6.25} - \underset{(0.01)}{0.03}\,JapanesePIL_t, \bar{R}^2 = 0.07$$

Periodo 1965-1981



Periodo 1982-1999



- La fonte dei risultati contrastanti è che entrambe le serie hanno un trend stocastico. I trend si allineano dal 1965 al 1981 ma divergono dal 1982 al 1999.

- Queste regressioni sono spurie.

Verifica della presenza di una radice unitaria

- I metodi informali richiedono di esaminare un grafico temporale dei dati ed il calcolo dei coefficienti di autocorrelazione.
- Se il primo coefficiente di autocorrelazione piccolo combinato con nessun trend apparente dal grafico suggerisce che la serie non abbia trend.
- Tuttavia, vi sono anche le procedure formali.

Il test di Dickey-Fuller per AR(1)

- L'ipotesi che Y_t abbia un trendo si riduce a verificare:

H_0: $\beta_1=1$ contro H_1: $\beta_1<1$ nella $Y_t= \beta_0+ \beta_1 Y_{t-1}+u_t$

- Se $\beta_1=1$, l'AR(1) ha una radice unitaria autoregressiva pari ad uno, quindi l'ipotesi nulla è che l'AR(1) abbia una radice unitaria e l'alternativa è che esso sia stazionario.
- Lo stesso test si può fare stimando una versione modificata della precedente ipotesi nulla, sottraendo all'AR(1) Y_{t-1} da entrambi i lati. Sia $\delta=\beta_1-1$, allora avremo

H_0: $\delta=0$ contro H_1: $\delta<0$ in $\Delta Y_t= \beta_0+ \delta Y_{t-1}+u_t$

- La statistica t degli OLS per la verifica di $\delta=0$ è detta statistica di Dickey-Fuller. Tale statistica è calcolata utilizzando errori "non robusti".

Il test Dickey-Fuller aumentato (ADF)

- Il test ADF per una radice autoregressiva unitaria verifica l'ipotesi nulla H_0: $\delta=0$ contro l'alternativa unilaterale H_0: $\delta<0$ nella regressione

  $$\Delta Y_t=\beta_0+\delta Y_{t-1}+\gamma_1 \Delta Y_{t-1}+ \gamma_2 \Delta Y_{t-2}+\ldots+ \gamma_p \Delta Y_{t-p}+u_t$$

- Sotto l'ipotesi nulla, Y_t ha un trend stocastico diversamente è stazionaria. La statistica ADF è la statistica OLS che verifica $\delta=0$.

- Se l'ipotesi alternativa è che Y_t sia stazionaria attorno a un trend deterministico, allora il trend t deve essere aggiunto come regressore addizionale, e in questo caso la regressione di Dickey-Fuller diventa

$$\Delta Y_t=\beta_0+\alpha t+\delta Y_{t-1}+\gamma_1 \Delta Y_{t-1}+ \gamma_2 \Delta Y_{t-2}+\ldots+ \gamma_p \Delta Y_{t-p}+u_t$$

- I valori critici per il test ADF unilaterale variano a seconda che si includa o meno il trend deterministico.

TABLE 14.5 Large-Sample Critical Values of the Augmented Dickey-Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

Esempio

- Consideriamo il modello ADF

$$\Delta\hat{Inf}_t = \underset{(0.23)}{0.53} - \underset{(0.04)}{0.11} Inf_{t-1} - \underset{(0.08)}{0.14} \Delta Inf_{t-1} - \underset{(0.08)}{0.25} \Delta Inf_{t-2} - \underset{(0.08)}{0.24} \Delta Inf_{t-3} + \underset{(0.08)}{0.01} \Delta Inf_{t-4}$$

- La statistica t dell'ADF è la statistica t che verifica l'ipotesi nulla sul coefficiente di Inf_{t-1}, t=-2.60
- Confrontando la tabella si vede che -2.60<-2.86 e quindi non rifiuta l'ipotesi nulla. Quindi l'inflazione contiene un trend stocastico.
- Se la lunghezza dei ritardi fosse stata decisa tramite l'AIC avremmo avuto 3 ritardi e non quattro.
- Anche in questo caso l'ipotesi nulla non può essere rifiutata al 5%. Tuttavia, considerando il livello del 10% i test rifiutano l'ipotesi nulla

Le rotture strutturali (structural breaks)

- Il secondo tipo di non stazionarietà nasce quando la funzione di regressione cambia all'interno del campione.
- Tali cambiamenti possono comparire a causa di un netto cambiamento nei coefficienti della regressione ad una data precisa o da una graduale evoluzione dei coefficienti nel corso di un periodo di tempo più lungo.
- I problemi causati sono che le stime OLS per l'intero campione stimeranno una relazione valida "in media" combinando gli effetti dei differenti periodi.
- A seconda della posizione e dell'ampiezza della rottura, la funzione di regressione "media" può essere molto diversa dalla vera funzione di regressione alla fine del campione.

Verifica di ipotesi

- Si verifica la presenza di cambiamenti netti nei coefficienti di regressione. Tale metodo dipende dalla conoscenza o meno della possibile data di rottura.

- Nel caso di data nota (Bretton Woods, 1972), utilizziamo un modello simile a quelli visti la settimana scorsa
  $Y_t=\beta_0+\beta_1 Y_{t-1}+\delta_1 X_{t-1}+\gamma_0 D_t(\tau)+\gamma_1 [D_t(\tau)xY_{t-1}]+$
  $+\gamma_2 [D_t(\tau)x X_{t-1}]+u_t$

- Se non c'è rottura allora la funzione di regressione è la stessa in entrambe le porzioni del campione e quindi i termini che contengono la variabile binaria $D_t(\tau)$ non entrano nella regressione. Cioè H_0: $\gamma_0=\gamma_1=\gamma_2=0$.
- Per verificare tale ipotesi si usa la statistica F per effettuare il Chow test.

- Nel caso di data ignota o conosciuta solo entro un intervallo si può utilizzare una versione modificata del Chow test chiamata statistica del rapporto delle verosimiglianze (QLR).
- Esso si basa sul valore più elevato della statistica F all'interno dell'intervallo considerato.
  - Può essere usato per verificare una rottura in tutti o solo alcuni dei coefficienti
  - Per grandi campioni, la distribuzione della statistica QLR sotto l'ipotesi nulla dipende dal numero di restrizioni da verificare, q, e dai punti estremi τ_0 e τ_1 come frazione di T
  - Solitamente si sceglie un troncamento al 15% con τ_0=0.15T e τ_1=0.85T.

La curva di Philips

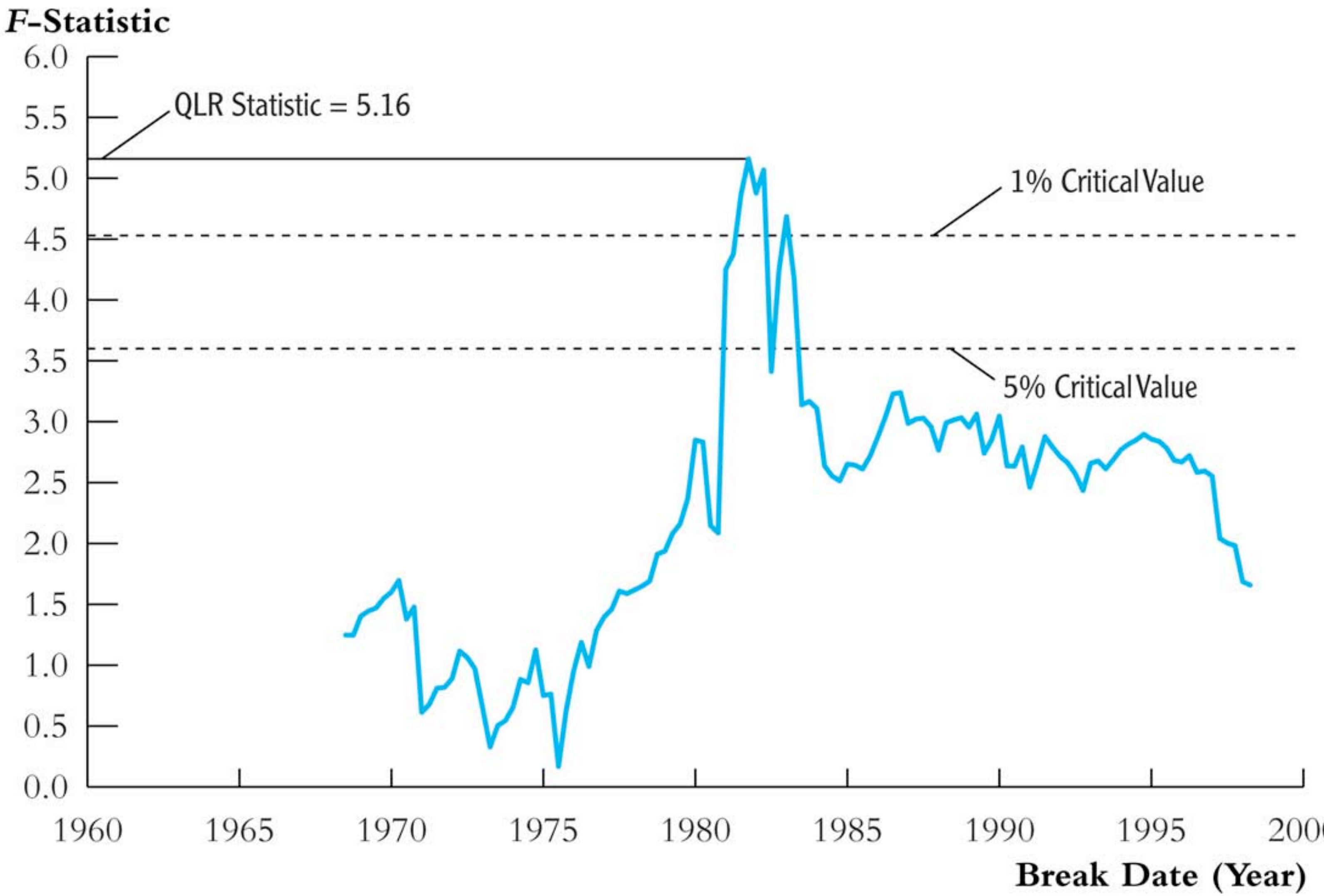
- Considerando il modello ADL(4,4)

$$\Delta \hat{Inf}_t = \underset{(0.47)}{1.32} - \underset{(0.09)}{0.36}\Delta Inf_{t-1} - \underset{(0.1)}{0.34}\Delta Inf_{t-2} + \underset{(0.08)}{0.07}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4}$$

$$- \underset{(0.47)}{2.68}\Delta Unemp_{t-1} + \underset{(0.89)}{3.43}\Delta Unemp_{t-2} - \underset{(0.89)}{1.04}\Delta Unemp_{t-3} + \underset{(0.44)}{0.07}\Delta Unemp_{t-4},$$

$$\bar{R}^2 = 0.35$$

- Possiamo applicare il test QLR per vedere se tale curva è stata stabile nel periodo 1962-1999.
- La statistica F per la verifica della presenza di una rottura nel 1980:I è 2.26. Ogni statistica F verifica per 5 restrizioni e quindi q=5. La più grande è 3.53 (1982:II), questa è la statistica QLR.
- Confrontando tale statistica con valori critici per q=5, l'ipotesi nulla (coeff. Stabili) è rifiutata al 10% ma non al 5%.



F-Statistic
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
QLR Statistic = 5.16
1% Critical Value
5% Critical Value
1960
1965
1970
1975
1980
1985
1990
1995
2000
Break Date (Year)

Pseudo previsioni fuori campione

- Il test ultimo di un modello di previsione è la sua prestazione fuori campione.

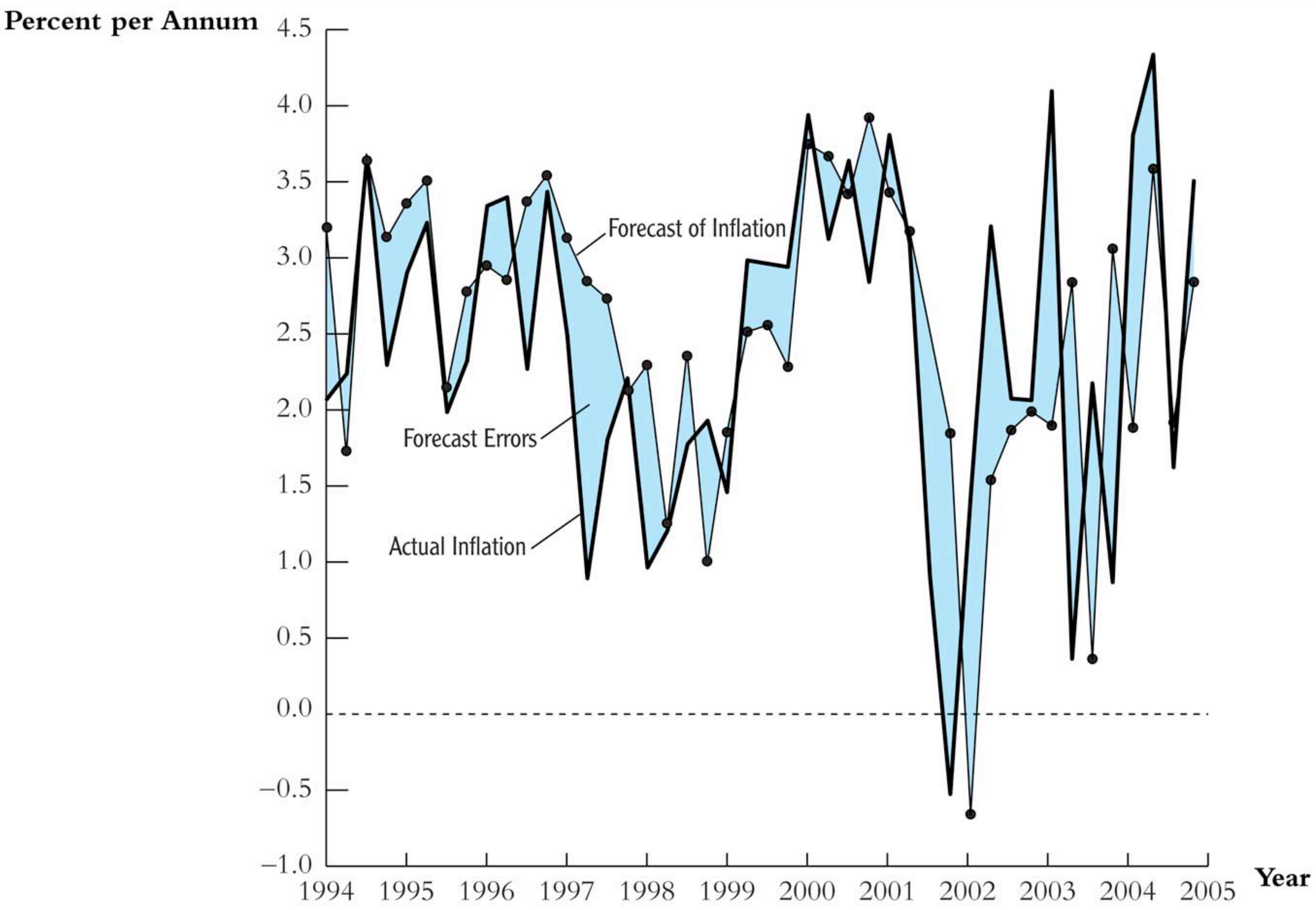
- La pseudo previsione fuori campione è un metodo per simulare la prestazione in tempo reale del modello di previsione.

1. Si scelga un numero di osservazioni P per cui si genereranno pseudo previsioni fuori campione. Ad es. P=10%. Sia s=T-P
2. Si stimi la regressione di previsione usando l'insieme troncato di dati per t=1,…,s.
3. Si calcoli la previsione per il primo periodo oltre il campione trocato, s+1, $\tilde{Y}_{s+1|s}$
4. Si calcoli l'errore di previsione $\tilde{u}_{s+1} = Y_{s+1} - \tilde{Y}_{s+1|s}$
5. Si ripetano le fasi 2-4 per le restanti date da s=T-P a s=T-1.

- Usando date multiple vicine alla fine del campione si ottengono una serie di pseudo previsioni e quindi di errori di pseudo previsione.
- Sono pseudo in quanto si utilizza il proprio modello per simulare delle previsioni in tempo reale (conosciamo i valori futuri della serie).
- Tale metodo dà a chi effettua previsioni un senso di quanto bene il modello stia prevedendo la fine del campione.

La curva di Philips

- Sono state calcolate le pseudo previsioni fuori campione per il periodo 1994:I-1999:IV utilizzando una curva di Philips a 4 ritardi.
- La previsione per 1994:I è stata calcolata con la funzione di regressione utilizzando i dati fino al 1993:IV e poi calcolando la previsione del 1994:I con i coefficienti stimati ed i dati fino al 1993:IV. Ripetendo tale operazione per i 24 trimestri otteniamo 24 pseudo previsioni.
- Le pseudo previsioni in figura seguono abbastanza bene l'inflazione effettiva ma sono mediamente più alte. Tale distorsione può essere stata causata da un declino del tasso naturale di disoccupazione che potrebbe manifestarsi come uno spostamento dell'intercetta della curva di Philips.



Percent per Annum
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
Forecast of Inflation
Forecast Errors
Actual Inflation
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
Year