

PRINCIPI DI ECONOMETRIA

lezione 18

AA 2015-2016

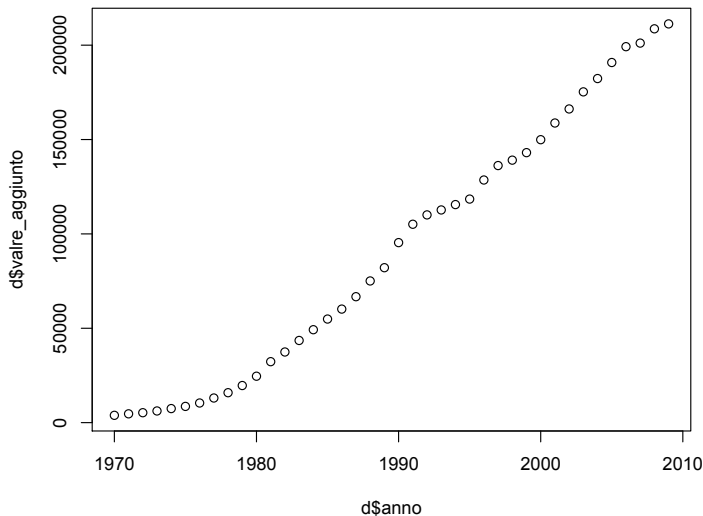
Paolo Brunori

- spesso come economisti siamo interessati a prevedere quale sarà il valore di una certa variabile nel futuro
- quando osserviamo una variabile nel tempo possiamo cercare di imparare qualcosa sul suo andamento e provare a predire come si comporterà in futuro
- il prezzo del petrolio, l'andamento dei prezzi, le precipitazioni piovose, il tasso di fecondità sono solo alcuni esempi
- quando ci concentriamo su una sola variabile e la sua variazione nel tempo ci stiamo occupando di serie temporali

VALORE AGGIUNTO COMPLESSIVO IN ITALIA

PRINCIPI DI
ECONOMETRIA

LEZIONE 18



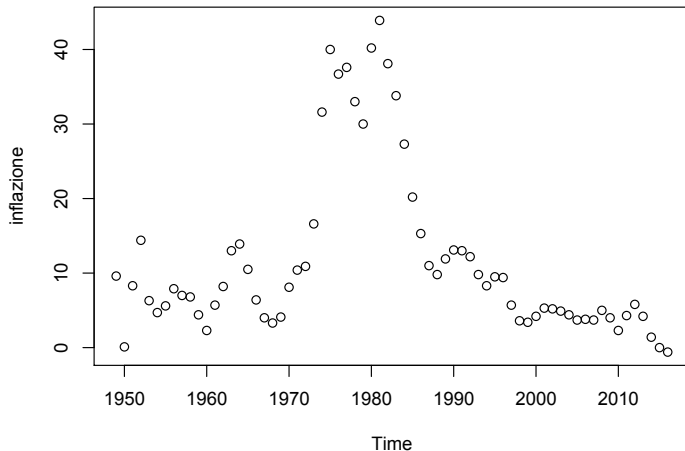
L'INFLAZIONE IN ITALIA

anno	inflazione	
1949	9.6	
1950	0.1	
1951	8.3	
1952	14.4	
1953	6.3	
1954	4.7	
1955	5.6	
1956	7.9	
1957	7	
1958	6.8	
1959	4.4	
1960	2.3	
1961	5.7	

L'INFLAZIONE IN ITALIA

PRINCIPI DI
ECONOMETRIA

LEZIONE 18



- nell'analisi di serie storiche si usa una terminologia specifica per descrivere l'andamento di una variabile di interesse Y_t nel tempo
- Y_{t-1} , il valore precedente a Y_t nella serie si chiama 'primo valore ritardato' di Y
- Y_{t-j} , si chiama 'j-esimo valore ritardato' di Y
- $\Delta Y_t = Y_t - Y_{t-1}$ si chiama differenza prima della variabile Y

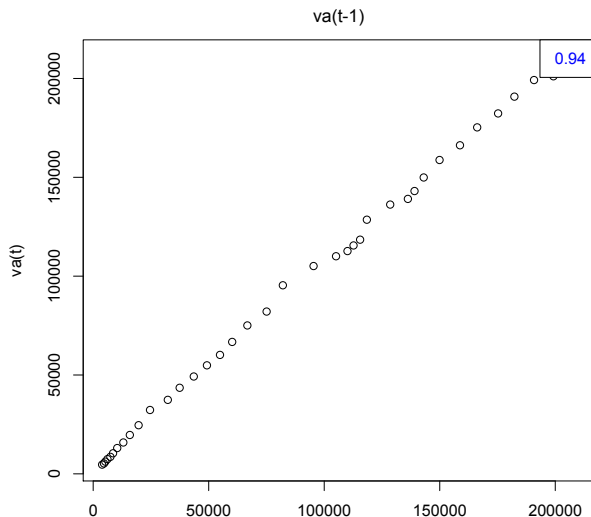
Autocorrelazione o correlazione seriale

- nelle serie temporali la correlazione fra Y_t e Y_{t+1} tende ad essere forte
- d'altra parte è normale che i cambiamenti avvengano lentamente
- la correlazione fra valore aggiunto in un anno e valore aggiunto nell'anno precedente è ad esempio pari al 0.94 (0.93 per l'inflazione)
- similmente l'autocorrelazione e autocovarianza j -esime sono le statistiche calcolate fra Y_t e Y_{t-j}

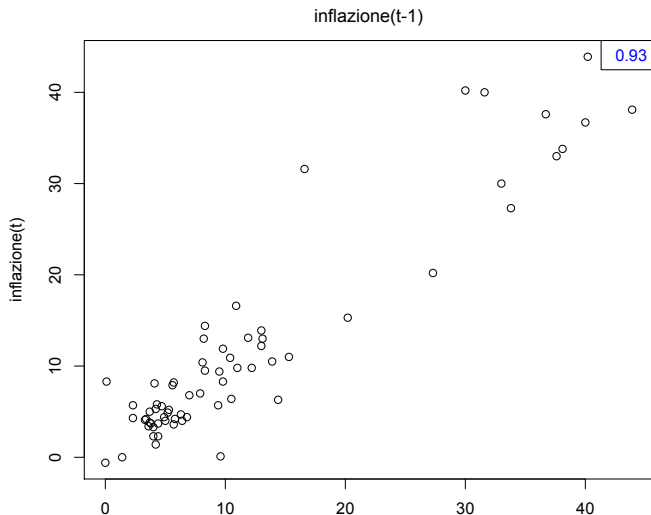
- il modello più semplice che possiamo pensare per spiegare l'andamento di una variabile nel tempo
- il valore è una funzione lineare del primo ritardo e un errore

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

AUTOCORRELAZIONE DEL VALORE AGGIUNTO TOTALE ITALIANO



AUTOCORRELAZIONE DEL TASSO DI INFLAZIONE IN ITALIA



MODELLO AR(1)

- il modello autoregressivo di ordine 1 stimato per spiegare l'inflazione è:

$$Y_t = 0.5161 + 0.9442Y_{t-1} + u_t$$

- $R^2 = 0.8763$, $adj - R^2 = 0.8744$

ERRORE DI PREVISIONE VS. RESIDUO

- l'errore di previsione è la differenza fra il valore che prevediamo per $t + 1$ sulla base dei dati che abbiamo al tempo t
- non è la stessa cosa del residuo che siamo abituati a incontrare nell'analisi di regressione perchè qui stiamo facendo previsioni fuori dal campione per un periodo non ancora osservato
- la differenza è comunque abbastanza labile visto che il regressore che predice Y_{t+1} è Y_t
- una misura di capacità di previsione è la radice quadrata dell'errore di previsione quadratico medio

$$RMSFE = \sqrt{E[Y_{t+1} - \hat{Y}_{t+1|t}]^2}$$

ERRORE DI PREVISIONE VS. RESIDUO

- la RMSFE combina errori che provengono da due fonti:
 u_t e l'incertezza riguardo a β_0 e β_1
- se la componente che deriva da u_t è prevalente (se siamo abbastanza certi riguardo ai valori β_0, β_1)
- allora RMSFE può essere approssimato con $\sqrt{\text{var}(u_t)}$
che non è altro che l'errore standard di regressione (RSE) che siamo abituati a stimare guardando ai residui

- non è detto che guardare al primo ritardo sia il modo migliore di stimare l'andamento nel tempo della variabile dipendente
- in realtà potrebbe essere migliore un modello che guardi a un maggior numero di ritardi

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$$

- per l'inflazione italiana possiamo specificare un modello $AR(2)$

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} u_t$$

	coefficiente	errore standard	t	$p - value$
β_0	1.0620	0.6679	1.590	0.117
β_{t-1}	1.2786	0.1118	11.439	0.0000
β_{t-2}	-0.3635	0.1127	-3.225	0.0020

$$R^2 = 0.9011, adj - R^2 = 0.898, RSE = 3.684$$

COME SI SCEGLIE L'ORDINE p ?

- in teoria si confronta la bontà di adattamento (previsione) ai nostri dati di modelli via via più complessi (p più elevato)
- per definizione però quando p aumento l'errore di previsione diminuisce
- per questo si usano funzioni non lineari in cui la bontà di adattamento del modello è penalizzata da un termine che dipende dall'ordine p utilizzato
- il ragionamento è simile a quello utilizzato per $adj - R^2$ anche se i metodi di correzione non sono esattamente gli stessi
- nel caso dell'inflazione ad esempio uno dei criteri più usati è il criterio di informazioni Bayesiano (BIC)

- Il criterio di informazioni Bayesiano (BIC) si basa sulla minimizzazione del valore:

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

- questo valore da una parte decresce con p perché la somma dei residui è al quadrato (SSR) è monotonicamente decrescente in p
- d'altra parte p fa aumentare il secondo addendo di BIC
- questo aumento di p è tanto più grande quanto maggiore è il rapporto fra il logaritmo e il valore di T (ovvero quanto più limitata nel tempo è la nostra osservazione della serie storica)

- è possibile che ci siano altre variabili che possono essere utili a predire il valore Y_t
- la curva di Phillips ad esempio ci suggerisce che nel breve periodo c'è una relazione negativa fra tasso di disoccupazione e inflazione
- maggiore disoccupazione tende ad essere seguita da minor crescita dei prezzi
- recuperiamo quindi la serie storica della disoccupazione in Italia e vediamo se è possibile stimare un modello che sfrutti sia le informazioni sui ritardi di Y che informazioni sul variare di qualche regressore X
- questi modelli sono chiamati modelli misti autoregressivi

- per l'inflazione italiana possiamo specificare un modello:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 X_{t-1} + u_t$$

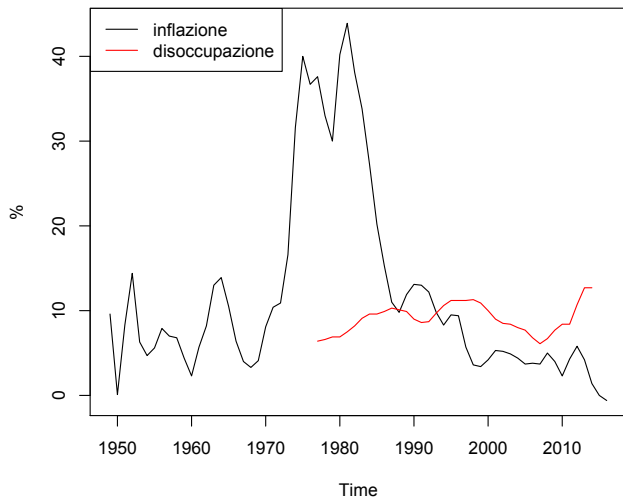
- dove assumiamo che l'inflazione sia dipendente dai suoi primi due ritardi e dalla disoccupazione l'anno precedente

- o potremmo specificare un modello misto ($ADL(p, q)$)
- ADL=Autoregressive Distributed Lag
- in cui selezioniamo la miglior combinazione di ritardi sia della Y che dei regressori X
- nel caso dell'inflazione italiana l'introduzione del regressore 'disoccupazione' ha un costo
- l'Istat ci fornisce la stima solo dal 1977 quindi la serie storica si restringe

INFLAZIONE E DISOCCUPAZIONE IN ITALIA

PRINCIPI DI
ECONOMETRIA

LEZIONE 18



- | | coefficiente | errore standard | t | $p - value$ |
|-------------------|--------------|-----------------|--------|-------------|
| β_0 | 3.8727 | 3.1804 | 1.218 | 0.2323 |
| β_{t-1} | 1.2578 | 0.1635 | 7.692 | 0.0000 |
| β_{t-2} | -0.3294 | 0.1533 | -2.149 | 0.0393 |
| $\beta_{X_{t-1}}$ | -0.8332 | 0.6597 | -1.263 | 0.2157 |

- malgrado il segno del ritardo della disoccupazione sia negativo non è statisticamente significativo

- in questo tipo di modelli si utilizzano osservazioni del passato per predire cosa accadrà nel futuro
- in pratica stiamo assumendo che le relazioni che osserviamo nel tempo siano valide anche nel futuro
- solo sulla base di questa assunzione possiamo considerare le previsioni affidabili

- formalmente una serie storica è detta stazionaria se, per qualsiasi valore di T , le relazioni che legano $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$, non dipendono da s
- ovvero se la sua distribuzione di probabilità non cambia nel corso del tempo
- se abbiamo due serie (come inflazione e disoccupazione) queste si dicono congiuntamente stazionarie se le relazioni che legano $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \dots, X_{s+T}, Y_{s+T})$, non dipendono da s
- ovvero la distribuzione di probabilità congiunta è invariante nel tempo
- ovviamente la relazione dipende da T ma la sua dipendenza da T è costante nel tempo

ASSUNZIONI PER REGRESSIONI TEMPORALI CON PREDITTORI MULTIPLI

- la stazionarietà di regressori e variabile dipendente è una delle assunzioni necessarie a ottenere una stima non distorta dei coefficienti di regressione
- un'altra assunzione diversa da quelle che siamo normalmente abituati a ammettere quando stimiamo regressioni riguarda la relazione esistente fra i valori e i valori ritardati utilizzati nel modello
- l'effetto dei ritardi Y_{t-j} e X_{t-j} deve svanire quando il valore di j diventa molto grande
- in caso contrario sarebbe impossibile stimare il modello con un numero finito di ritardi osservabili

VIOLAZIONE DELL'ASSUNZIONE DI STAZIONARIETÀ

- ci sono molti casi in cui la stazionarietà è violata
- il caso più frequente è la presenza di un trend nella serie storica
- il trend può essere deterministico: βt oppure aleatorio
- quando un trend è aleatorio non sappiamo cosa attenderci ma sappiamo che la relazione fra i valori dipende dal punto della serie (il momento) in cui la osserviamo
- l'esistenza di trend nei dati distorce le stime e deve essere risolta
- generalmente si risolve trasformando la serie ad esempio prendendo invece che Y_t come variabile esplicativa

$$\Delta Y_t = Y_{t+1} - Y_t$$

- un'altro caso di violazione dell'assunzione di stazionarietà sono le rotture strutturali
- per quanto concerne l'inflazione ad esempio potremmo ipotizzare l'esistenza di due periodi distinti: prima e dopo l'unione monetaria
- se vi sono break-strutturali ovviamente ad un certo punto la relazione fra le variabili e i ritardi viene meno per definizione
- se si sospetta che una rottura strutturale possa esserci deve essere testata e nel caso si identifichi occorre correggere il modello
- ad esempio introducendo una variabile dicotomica che individua il prima e il dopo l'intervento della rottura