

Note di econometria II

(Serie temporali)

Valerio Potì

30 maggio 2021

QUESTE DISPENSE SONO ALLO STATO DI BOZZA

L'AUTORE SAREBBE MOLTO GRATO CON IL LETTORE PER QUALUNQUE
SUGGERIMENTO O APPUNTO

(Work-in-progress, comments and feed-back very welcome)

NB:

- Queste note non sono direttamente ed integralmente oggetto di esame, salvo quanto trattato in classe e specificato nel syllabo;
- si distribuiscono agli studenti del corso di Econometria solo come riferimento aggiuntivo per studi ulteriori, se eventualmente interessati, e con la preghiera di offrire, in cambio, suggerimenti per migliorarle.

Indice

2.	Processi stocastici e modelli per serie temporali.....	3
2.1	Processi stocastici	4
2.2	Stazionarietà ed ergodicità	5
2.3	Autoregressioni	10
2.4	Modelli più complessi	14
2.5	Il problema della previsione.....	17
2.6	Criteri di selezione.....	18
2.7	Incertezza e intervalli di confidenza previsionali	19
2.8	Piccola digressione: causalità e previsione	21
2.9	Modelli VAR.....	22
2.10	Variabili non-stazionarie: trend e discontinuità strutturali	25
2.11	Discontinuità strutturali (structural breaks)	26
2.12	Trend deterministici e stocastici	29
2.13	Qual'è il problema se c'è un trend?	33
2.14	Come stabiliamo se c'è un trend stocastico?.....	35
2.15	Cosa fare in presenza di trend stocastici	37

2. Processi stocastici e modelli per serie temporali

2.1 Processi stocastici

Come già accennato, un tipo particolare di dati con cui si lavora in econometria sono le serie temporali. Nel caso delle serie temporali, le variabili casuali prese in esame sono viste in successione temporale e ci si riferisce al loro DGP con il termine **processo stocastico**.

I modelli per le serie temporali sono particolarmente utili quando si cerca di prevedere i valori futuri di una variabile utilizzando le informazioni contenute nei valori passati del processo e in molti casi solo queste. Possono richiedere o meno una comprensione delle forze che influenzano, direttamente o indirettamente, la variabile da modellare.

Come già detto all'inizio, il modello di regressione può essere utilmente impiegato per descrivere le relazioni tra variabili di diversa natura, inclusi i processi stocastici e le serie temporali cui questi danno luogo. Quando le variabili siano processi stocastici, occorre però notare che la dimensione temporale acquista una rilevanza particolare. Ovviamente, la variabile che si intende spiegare e/o prevedere può dipendere in maniera lineare o non lineare da una o più variabili indipendenti e da un termine di errore, ma in questo caso occorrerà tenere presente che le relazioni tra variabili si sviluppano tipicamente in maniera dinamica, ovvero tramite un susseguirsi di relazioni di causa ed effetto in cui la dimensione temporale ha una rilevanza imprescindibile.

La previsione è data dal valore atteso condizionale, ovvero da $y_{t+1|t} \equiv E(y_{t+1}|I_t)$ ove I_t rappresenta l'informazione disponibile allo studioso al momento in cui la previsione viene effettuata, il tempo t in questo caso, e dunque la storia del processo fino a quel momento. Spesso si distingue tra previsione **'in-sample'** e **'out-of-sample'**. Il primo tipo di previsione è $E(y_{t+s}|y_t)$ per $t+s \in [1, 2, \dots, T]$, mentre la previsione out-of-sample è $E(y_{T+s}|y_T)$. Entrambe richiedono la conoscenza dei parametri del modello o una loro stima. Molti autori distinguono in maniera marcata tra errore di previsione out-of-sample, ovvero $y_{T+s} - y_{T+s|T}$, e errore di previsione in-sample, $y_{t+s} - y_{t+s|t}$ e riservano a quest'ultimo la definizione di residuo. Per coerenza con le definizioni adottate nel caso più generale, in queste note si riserverà l'utilizzo del termine "residuo" per riferirsi alla differenza tra valore osservato della variabile dipendente e valore stimato ovvero, nel contesto delle serie temporali, sia a $y_{t+s} - \hat{y}_{t+s|t}$ che a $y_{T+s} - \hat{y}_{T+s|T}$.

Nel seguito, dopo aver discusso alcuni concetti di fondamentale importanza nello studio delle serie storiche, introdurremo processi di complessità e generalità crescente avvertendo sin d'ora lo studente che l'intrusione della dimensione temporale introduce problemi peculiari.

2.2 Stazionarietà ed ergodicità

Come accennato sopra, nel considerare una serie storica, facciamo spesso l'ipotesi che il passato sia una affidabile ed intellegibile guida da cui possiamo imparare per prevedere il futuro.

Ciò equivale ad assumere che un campione sufficientemente ampio di realizzazioni passate del processo sia rappresentativo della intera "popolazione", cioè dell'intero processo (passato, presente e futuro).

Ciò, a pensare bene, richiede due condizioni. La prima è che il processo sia "**stazionario**", ovvero costante nel tempo. In termini formali, un processo **strettamente stazionario** è uno per le cui possibili realizzazioni y_t valga la regola

$$P\{y_{t_1} \leq b_1, \dots, y_{t_n} \leq b_n\} = P\{y_{t_1+m} \leq b_1, \dots, y_{t_n+m} \leq b_n\}$$

cioè la misura di probabilità della successione $\{y_t\}$ è uguale alla misura di probabilità della successione $\{y_t\} \forall m$. Questa è in realtà una definizione molto forte di stazionarietà e in molte circostanze ci possiamo accontentare della cosiddetta "stazionarietà debole". Un processo **debolmente stazionario** o **stazionario in covarianza** invece soddisfa le tre condizioni seguenti:

1. $E(y_t) = \mu \quad t = 1, 2, \dots, \infty$
2. $E[(y_t - \mu)(y_t - \mu)] < \infty$
3. $E[(y_{t_1} - \mu)(y_{t_2} - \mu)] = \gamma_{t_2-t_1} \quad \forall t_1, t_2$

Quindi, in buona sostanza, se il processo è stazionario in covarianza, tutte le varianze (e quindi necessariamente tutti i valori attesi) sono uguali e tutte le autocovarianze dipendono dalla differenza tra le 't'. Un'espressione leggermente più generale per l'autocovarianza di ordine s di un determinato processo è la seguente

$$\gamma_s = Cov(y_t, y_{t-s}) = E[(y_t - E(y_t))(y_{t-s} - E(y_{t-s}))]$$

La successione $\gamma_s, s = 1, 2, \dots$, è nota come funzione di autocovarianza ed è utile a descrivere la dipendenza temporale delle realizzazioni di un dato processo stocastico. Le variabili y_{t-s} sono note come "**ritardi**" di y_t di ordine s . Dunque, è possibile riformulare il punto 3 della definizione di stazionarietà richiedendo che la funzione di autocovarianza, per $s = 1, 2, \dots$, sia costante. Si noti che la grandezza delle autocovarianze dipende dall'unità di misura di y_t . Pertanto, è spesso conveniente usare al loro posto le cosiddette autocorrelazioni, ovvero autocovarianze normalizzate tramite una divisione per la varianza (autocovarianza di ordine 0):

$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

La seconda condizione per imparare dal passato di un processo è che non occorra un campione di ampiezza infinita per inferirne le caratteristiche! Quest'ultima circostanza è l'implicazione principale dell'assunzione di “ergodicità”, ovvero dell'assunzione che le realizzazioni del processo stocastico non siano serialmente “eccessivamente” correlate (o, più in generale, dipendenti). Abbiamo bisogno che la correlazione seriale non sia “eccessiva” per poter interpretare le osservazioni come realmente distinte, e quindi per interpretare un campione numericamente “grande” come “veramente” tale. Dunque, l'assunzione di ergodicità richiede il processo non sia troppo “persistente”. In termini della funzione di autocorrelazione, ciò richiede che, al crescere di s , i valori assunti dalla funzione declinino in maniera sufficientemente rapida da comportare che la serie $\sum_{s=1}^{\infty} \frac{\rho_s}{s}$ converga ad un valore nullo.

Vale la pena sottolineare che tale condizione è legata a quella di stazionarietà debole, in quanto entrambi le proprietà sono necessarie per poter imparare dal passato di un processo, ma non sono la stessa cosa.

In altre parole, se abbiamo una serie storica di T osservazioni su un determinato processo, con T “grande”, l'ergodicità garantisce che, in un certo senso e in una certa misura, possiamo “trattarle” come realizzazioni di un gran numero di variabili aleatorie indipendenti. Quando questa condizione valga, si può dunque essenzialmente trattare una serie temporale e dunque un insieme di dati in sequenza temporale su una data variabile, alla stregua di dati su un campione di individui, ovvero sulle variabili casuali che costituiscono il processo stocastico. Inoltre, se la condizione di stazionarietà è soddisfatta, significa che possiamo trattare tali variabili come identicamente distribuite. Se entrambe le condizioni sono soddisfatte, significa che possiamo trattare tali variabili come *i.i.d.* e applicare la legge dei grandi numeri / teorema del limite centrale per ottenere una stima consistente della media del processo.

Ad esempio, si consideri il tasso di inflazione negli USA. Viene definito come il tasso di crescita percentuale del cosiddetto CPI (Consumer Price Index), ovvero

$$Infl_t = 100 \frac{(CPI_t - CPI_{t-1})}{CPI_{t-1}}$$

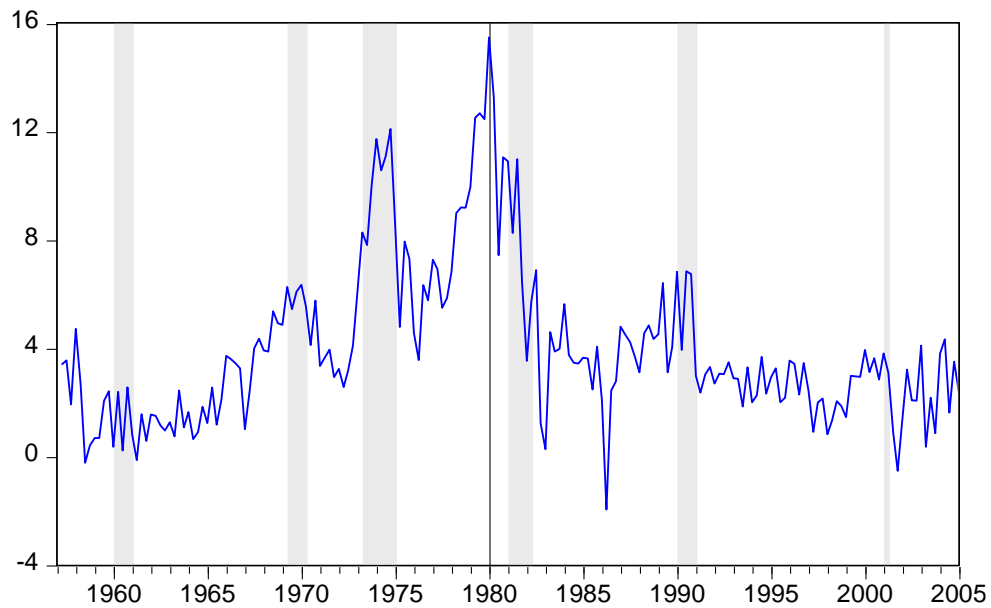
Per il periodo dal primo al secondo semestre del 2004 (cioè da 2004Q1 a 2004Q2), abbiamo

$$\begin{aligned} Infl_{2004Q2} &= 100 \frac{(CPI_{2004Q2} - CPI_{2004Q1})}{CPI_{2004Q1}} = 100 \times \frac{188.6 - 186.57}{186.57} = 1.088 \\ &= 100 \times 1.088\% \end{aligned}$$

ovvero un tasso di inflazione percentuale dell'1.088 percento nel trimestre. Il grafico in Figura 1 riporta la serie temporale di tale tasso in versione annualizzata (ovvero ottenuta moltiplicando $Infl_t$ per 4) dal 1960 al 2005 con le aree in grigio che evidenziano i periodi di recessione economica.

Figura 1

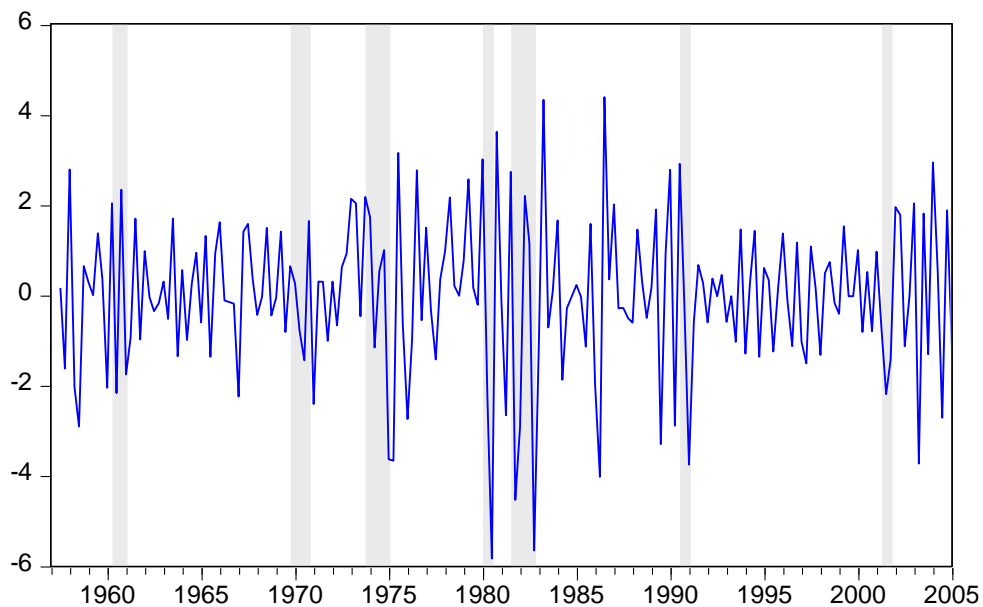
Tasso di inflazione annualizzato negli USA (1960-2004)



Si può notare che questa serie sia piuttosto persistente. Infatti, il grafico è dominato da ampie oscillazioni che si sviluppano su un arco di molti anni, a differenza del grafico della variazione del tasso di inflazione riportato in Figura 2, ovvero di $\Delta Infl_t = Infl_t - Infl_{t-1}$, che non subisce queste oscillazioni di lungo periodo ed è invece più soggetto a frequenti sbalzi di breve periodo.

Figura 2

Variazione del tasso di inflazione negli USA (1960-2004)



Un ovvio stimatore del coefficiente di autocovarianza e di autocorrelazione di ordine s è rappresentato, rispettivamente, dalla autocovarianza campionaria

$$Cov(\widehat{y_b y_{t+s}})$$

e dalla autocorrelazione campionaria

$$\hat{\rho}_s = \frac{Cov(\widehat{y_b y_{t+s}})}{Var(\widehat{y_t})}$$

ove

$$Cov(\widehat{y_b y_{t+s}}) = \frac{1}{T} \sum_{t=j+1}^T (y_t - \bar{y}_{j+1,T})(y_{t-j} - \bar{y}_{1,T})$$

e $\bar{y}_{j+1,T}$ è la media campionaria di y_t calcolata utilizzando le osservazioni da $t = j+1$ a $t = T$ (perchè?).

Come riportato nella tabella qui in basso, il coefficiente di autocorrelazione (di ordine $s = 1$) del tasso di inflazione negli USA è piuttosto alto, il che riflette la persistenza della serie notata prima. Ci sono però nondimeno alcune ‘sorpresa’ in questa serie. Ciò è riflesso in un coefficiente di autocorrelazione alto ma non proprio uguale ad uno e che declina abbastanza velocemente al crescere di s . Possiamo dunque ritenere, almeno in prima analisi, che la serie sia ergodica e che quindi abbia un senso cercare di stimare il valore atteso le altre caratteristiche salienti del processo stocastico che la ha generata.

Autocorrelazione del tasso di inflazione negli USA, 1960-2004

p	Autocorrelazione
1	0.837
2	0.761
3	0.765
4	0.676

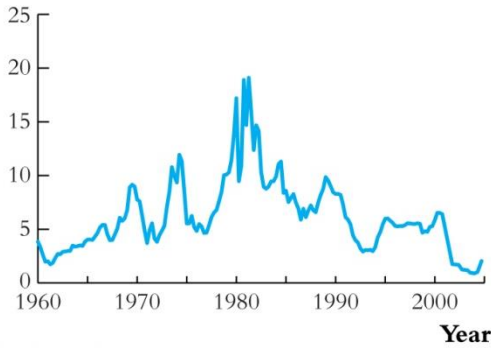
Autocorrelazione della variazione tasso di inflazione negli USA, 1960-2004

p	Autocorrelazione
1	-0.265
2	-0.252
3	0.294
4	-0.077

Se rappresentiamo graficamente la funzione di autocorrelazione riportando in un grafico le ρ_s in funzione delle $s = 1, 2, \dots$, otteniamo il grafico noto come correlogramma. Si noti come la autorelazione della variazione del tasso di inflazione sia considerevolmente meno autocorrelata del tasso stesso.

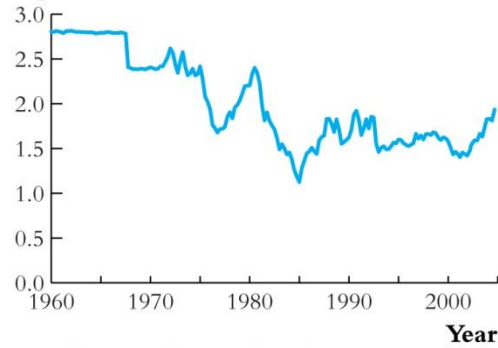
Si considerino ora queste altre serie. Cosa ne pensate? Sono stazionarie ed ergodiche o no?

Percent per Annum



(a) Federal Funds Interest Rate

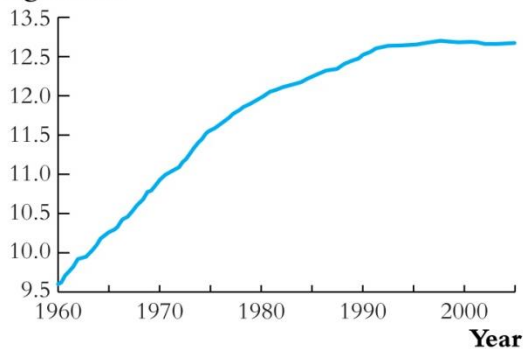
Dollars per Pound



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

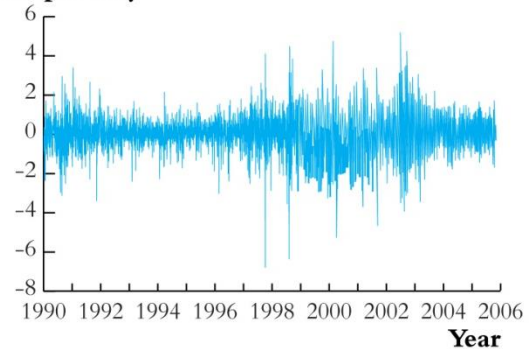
E di queste qui in basso, che ne dite?

Logarithm



(c) Logarithm of GDP in Japan

Percent per Day



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

2.3 Autoregressioni

Tra i processi più comunemente utilizzati per modellizzare serie storiche incontrate di frequente negli studi economici abbiamo le cosiddette **autoregressioni**. Un'autoregressione è un modello in cui la variabile che si vuole prevedere viene regredita sui propri ritardi. In un'autoregressione del primo ordine, cui ci riferisce scesso con l'abbreviazione **AR(1)**, y_t viene regredita su y_{t-1} come mostrato di seguito¹

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

Più in generale, in un'autoregressione di ordine p , ovvero in un modello AR(p), y_t viene regredita su $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ come mostrato di seguito

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + u_t$$

Può essere utile a questo punto far presente che possiamo rappresentare i ritardi che appaiono in queste equazioni come il risultato dell'applicazione di un operatore apposito che si comporta a tutti gli effetti come una variabile e segue le regole ordinarie del calcolo. Tale operatore è il cosiddetto **operatore ritardo di ordine k** il cui effetto è quello di generare un ritardo dello stesso ordine della variabile cui è applicato. Per esempio, se denotiamo con L tale operatore, possiamo rappresentare $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}$, come $L^0 y_t, L y_t, L^2 y_t, \dots, L^p y_t$ e quindi la equazione di cui sopra diventa

$$y_t = \beta_0 L^0 y_t + \beta_1 L y_t + \dots + \beta_p L^p y_t + u_t$$

In ogni caso, si noti che i coefficienti del modello non hanno necessariamente un'interpretazione causale ma perchè il modello sia valido, occorre che fornisca almeno una 'affidabile' rappresentazione "in forma ridotta" delle relazioni causali che determinano il DGP, le quali a loro volta plausibilmente coinvolgono altre variabili non incluse nel modello. Se conoscessimo i coefficienti del modello, potremmo utilizzarlo per effettuare una previsione di y_t . Ad esempio, nel modello AR(1), la previsione un periodò in la è data da

$$E(y_{t+1}|y_t) = E(\beta_0 + \beta_1 y_t + u_t | y_t) = \beta_0 + E(\beta_1 y_t | y_t) = \beta_0 + \beta_1 E(y_t | y_t) = \beta_0 + \beta_1 y_t$$

In circostanze abbastanza frequenti, i modelli di autoregressione possono essere agevolmente stimati col metodo OLS e, a certe condizioni, possiamo svolgere test di significatività dei coefficienti nella maniera usuale. Per esempio, in grandi campioni, questo è il caso se u_t sia *i.i.d.*

¹ Per un processo y_t tale che $y_t = b y_{t-1} + e_t$, with $E_{t-1}(e_t) \forall t$, il valore atteso condizionato è il seguente:
 $E_t(y_{t+j}) = E_t(b y_{t+j-1}) = b E_t(y_{t+j-1}) = b E_t(b y_{t+j-2}) = b^2 E_t(y_{t+j-2}) = b^3 E_t(y_{t+j-3}) = \dots = b^j E_t(y_{t+j-j}) = b^j y_t$. Dunque, per un processo $y_t = a + b y_{t-1} + e_t$, si ha $E_t(y_{t+j}) = E_t(a + b y_{t+j-1}) = a + b E_t(y_{t+j-1}) = a + b E_t(a + b y_{t+j-2}) = a + b a + b^2 E_t(y_{t+j-2}) = a + b a + b^2 a + b^3 E_t(y_{t+j-3}) = \dots = a + b a + b^2 a + \dots + b^{j-1} a + b^j E_t(y_{t+j-j}) \rightarrow \frac{a}{1-b}$ se $|b| < 1$.

e con momenti secondi finiti. In campioni finiti, avremmo bisogno invece dell'assunzione che u_t sia *i.i.d.* e normalmente distribuito. Per esempio, un test dell'ipotesi nulla che $\beta_1 = 0$ rappresenta un test dell'ipotesi che y_{t-1} non sia utile nello spiegare e dunque prevedere y_t . Nel caso del modello AR(1), ciò equivale ad un test che il DGP possa essere descritto da un processo con valore atteso condizionale costante, ovvero eguale a

$$E(y_t|y_{t-1}) = \beta_0 + \beta_1 y_{t-1} = \beta_0 + 0 \times y_{t-1} = \beta_0 \quad \forall t$$

Esempio: modelli AR(1) e AR(4) della variazione del tasso di inflazione

Domanda: possiamo usare i valori assunti in passato dal tasso di inflazione per prevederne il futuro? Bè, tanto per cominciare direi che, per andare sul sicuro, forse per ora è meglio lavorare con la variazione del tasso di inflazione, che appare meno persistente del tasso di inflazione, almeno finchè non avremo imparato a distinguere in maniera rigorosa tra serie generate da processi stazionari ed ergotici e serie che non lo sono.

Quindi, potremmo riformulare la domanda come segue: possiamo usare le variazioni passate del tasso di inflazione per prevedere quelle future? Anche in questa versione, non è una domanda peregrina perchè è molto rilevante ai fini della validità di una teorica economica rispettabilissima quale quella delle aspettative adattive. In ogni caso, una volta prevista la variazione del tasso di inflazione, possiamo sempre usare questa previsione per costruire una previsione del tasso stesso, visto che $Infl_t = \Delta Infl_t + Infl_{t-1}$. Dunque, utilizzando dati dal 1962Q1 a 2004Q4 per stimare un modello AR(1) della variazione del tasso di inflazione otteniamo:

$$\Delta Infl_t = 0.017 - 0.23 \Delta Infl_{t-1} \quad R^2 = 0.05$$

(0.126) (0.07)

Eviews screenshot:

Dependent Variable: DINFL
 Method: Least Squares
 Date: 10/13/11 Time: 18:52
 Sample: 1962Q1 2004Q4
 Included observations: 172

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.017101	0.126876	0.134783	0.8929
DINFL(-1)	-0.238047	0.074693	-3.187025	0.0017
R-squared	0.056379	Mean dependent var		0.017016
Adjusted R-squared	0.050829	S.D. dependent var		1.707934
S.E. of regression	1.663962	Akaike info criterion		3.867840
Sum squared resid	470.6909	Schwarz criterion		3.904439
Log likelihood	-330.6342	Hannan-Quinn criter.		3.882689
F-statistic	10.15713	Durbin-Watson stat		2.165851
Prob(F-statistic)	0.001711			

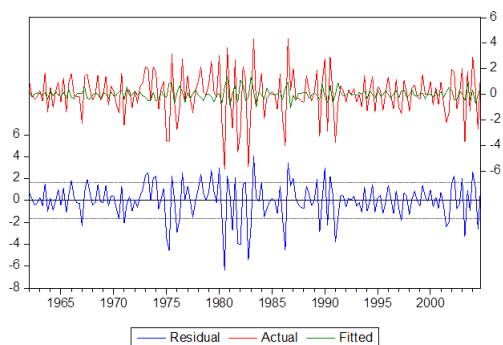
La capacità esplicativa del modello AR(1) nell'esempio appena considerato non è particolarmente eccitante. Al di là dell' R^2 striminzito, anche graficamente non sembra proprio che la serie stimata abbia moltissimo a che fare con quella osservata...come si può vedere nei grafici qui sotto. Nel primo di questi grafici, ovvero in Figura 3, si riportano i valori osservati ("Actual"), ovvero la serie $Infl_t$, e quelli stimati ("Fitted"), ovvero la serie $Infl_{t|t-1}$. Questi ultimi sono stati generati utilizzando le stime puntuali dei parametri ottenute in precedenza, ovvero $\hat{\beta}_0 = 0.017101$ e $\hat{\beta}_1 = -0.238047$, nell'equazione del modello AR(1), ottenendo un valore di $Infl_t$ stimato per ogni $t \in [t+1, T]$ secondo la formula

$$\widehat{Infl}_{t|t-1} = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} = 0.017101 - 0.238047 Infl_{t-1}$$

Il grafico conferma che, come anticipato dal coefficiente R^2 piuttosto basso, il modello AR(1) è in grado di spiegare granchè dei movimenti della variazione 'in sample' del tasso di inflazione USA nel periodo considerato.

Figura 3

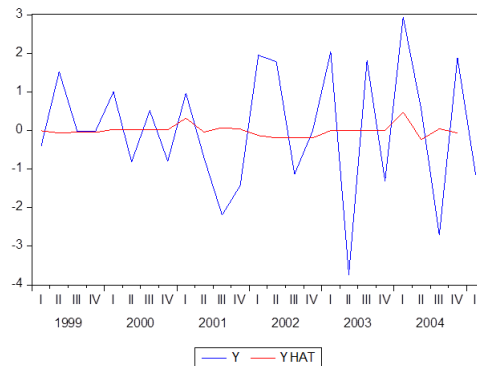
Variazioni del tasso di inflazione stimate vs. osservate - AR(1)



Questa performance piuttosto povera 'in sample' si riflette in una performance deludente 'out-of-sample', come si può vedere nel grafico in Figura 4. In questo grafico (in cui ci si è limitati ad un periodo più ristretto per facilitare la visione degli aspetti essenziali), si riporta come prima la serie osservata $Infl_t$, ma questa viene ora confrontata con la previsione cosiddetta 'pseudo out-of-sample' (*poos*). Quest'ultima è stata generata stimando una sequenza di modelli AR(1) utilizzando un campione di 20 osservazioni consecutive alla volta, e si sono utilizzate le stime puntuali dei coefficienti di questi modelli per generare una previsione 'out of sample', ovvero per generare $Infl_{T_{t-1}+1|T_{t-1}}$ dove T_{t-1} rappresenta il sub-periodo di campionamento di 20 osservazioni che inizia in $t - 1 - 20$ e finisce in $t \leq T - 1$. Il nomignolo 'pseudo', come si vede, è ben meritato perchè, di fatto, vengono prodotte previsioni che si riferiscono ad unità di tempo che fanno parte del periodo di campionamento (il che genera un sacco di confusione anche per economisti di professione).

Figura 4

Variazioni del tasso di inflazione predette (pseudo out-of-sample) vs. osservate - AR(1)



Sorge dunque spontanea la domanda se sia possibile identificare un modello che funzioni un pò meglio. Vediamo come si comporta un AR(4)? Utilizzando come prima dati dal 1960 al 2004 per stimare il modello, abbiamo

$$\Delta Infl_t = 0.008 - 0.30 \Delta Infl_{t-1} - 0.30 \Delta Infl_{t-2} + 0.13 \Delta Infl_{t-3} - 0.01 \Delta Infl_{t-4} \quad R^2 = 0.19$$

(0.12) (0.08) (0.08) (0.08) (0.08) $\bar{R}^2 = 0.17$

Eviews screenshot:

Dependent Variable: Y
 Method: Least Squares
 Date: 10/14/11 Time: 12:21
 Sample: 1960Q1 2004Q1
 Included observations: 177

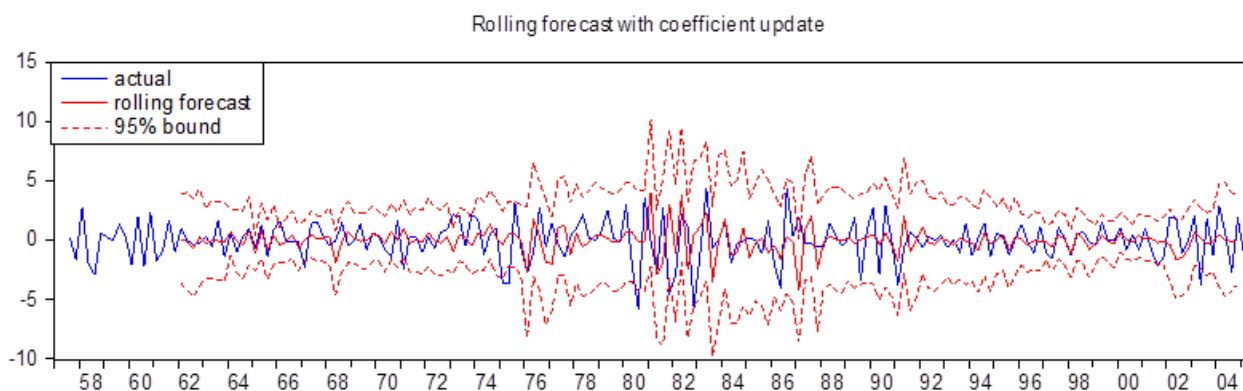
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.008701	0.116850	0.074462	0.9407
Y(-1)	-0.302571	0.077460	-3.906165	0.0001
Y(-2)	-0.298863	0.079639	-3.752736	0.0002
Y(-3)	0.134099	0.079442	1.688009	0.0932
Y(-4)	-0.005692	0.078811	-0.072229	0.9425
R-squared	0.191477	Mean dependent var		0.007867
Adjusted R-squared	0.172674	S.D. dependent var		1.708894
S.E. of regression	1.554366	Akaike info criterion		3.747855
Sum squared resid	415.5615	Schwarz criterion		3.837577
Log likelihood	-326.6852	Hannan-Quinn criter.		3.784243
F-statistic	10.18340	Durbin-Watson stat		1.958363
Prob(F-statistic)	0.000000			

Inoltre, tutti i coefficienti, eccetto l'ultimo, sono statisticamente significativi. Il che suggerisce che almeno i ritardi di ordine 1, 2 e 3 vadano mantenuti. In linea con queste considerazioni, il test F, che è aggiustato per i gradi di libertà e dunque per il numero di parametri del modello, consente di rigettare ad un elevato livello di significatività (p-value pari a 0.000) l'ipotesi che i valori dei

parametri stessi siano tutti nulli. Quindi, ‘in sample’, sembra che il modello AR(4) si comporti molto meglio del modello AR(1). Il che è riflesso anche in una migliore performance out-of-sample, come si può vedere confrontando il grafico in Figura 5 con quello, relativo al modello AR(1) e visto in precedenza, in Figura 4.

Figura 5

Variazioni del tasso di inflazione predette (pseudo out-of-sample) vs. osservate - AR(4)



2.4 Modelli più complessi

Visto che il modello AR(4) sembra migliore del modello AR(1), perchè non proviamo a chiederci se ci siano altre variabili che abbiano una abilità previsiva nei confronti della variazione del tasso di inflazione? Invece di limitarci ad includere nel modello ritardi della sola variabile dipendente, potremmo aggiungere anche ritardi di altre variabili. Modelli di questo tipo che includano p ritardi della variabile dipendente e r ritardi di altre variabili x_t , che fungono da previsori addizionali, sono noti come **autoregressive distributed lag models ADL(p,r)**:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \delta_1' x_{t-1} + \dots \delta_r' x_{t-r} + u_t$$

In questi modelli, i coefficienti di ciascuno degli elementi di δ_r sono noti come **moltiplicatori dinamici**, in quanto rappresentano la reattività della variabile dipendente a variazioni delle variabili indipendenti avvenute r periodi addietro.

Esempio: la relazione dinamica tra inflazione e disoccupazione

Secondo il modello (un po ad hoc ed ‘empirico’ noto come la “curva di Plillips”, il tasso di inflazione tende a salire se il tasso di disoccupazione è al di sotto del tasso di disoccupazione “naturale” o, più formalmente, il tasso di disoccupazione tende ad sperimentare variazioni al rialzo o al ribasso se il tasso di inflazione è al di sotto o al di sopra, rispettivamente, del tasso ‘steady state’ noto come “non-accelerating rate of inflation unemployment rate” (NAIRU,

nell'abbreviazione comunemente usata). Alcune domande sorgono spontanee...è vero tutto ciò empiricamente? E, se sì, possiamo utilizzare il tasso di disoccupazione per migliorare la nostra previsione del tasso di inflazione? Una specificazione abbastanza semplice ma consistente con la curva di Phillips è la seguente

$$Infl_t = \beta_0 + \beta_1 Infl_{t-1} + \dots + \beta_4 Infl_{t-4} + \delta_1 LHUR_{t-1} + \dots + \delta_4 LHUR_{t-4} + u_t$$

dove $LHUR_{t-1}$ denota il tasso di disoccupazione USA alla fine del periodo $t - 1$, e così via. La stima OLS è la seguente:

$$\begin{aligned} \Delta Infl_t = & 1.30 - 0.42 \Delta Infl_{t-1} - 0.37 \Delta Infl_{t-2} + 0.06 \Delta Infl_{t-3} - 0.04 \Delta Infl_{t-4} \\ & (0.44) \quad (0.08) \quad (0.09) \quad (0.08) \quad (0.08) \\ & - 2.64 LHUR_{t-1} + 3.04 LHUR_{t-2} - 0.38 LHUR_{t-3} + 0.25 LHUR_{t-4} \\ & (0.44) \quad (0.86) \quad (0.89) \quad (0.45) \end{aligned}$$

$$R^2 = 0.34$$

$$\bar{R}^2 = 0.18$$

Eviews screenshot:

Dependent Variable: DINFL
 Method: Least Squares
 Date: 10/14/11 Time: 15:21
 Sample: 1962Q1 2004Q4
 Included observations: 172
 White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.304286	0.451605	2.888116	0.0044
DINFL(-1)	-0.419822	0.088696	-4.733251	0.0000
DINFL(-2)	-0.366630	0.094038	-3.898756	0.0001
DINFL(-3)	0.056568	0.084797	0.667099	0.5057
DINFL(-4)	-0.036458	0.083529	-0.436475	0.6631
LHUR(-1)	-2.635568	0.474817	-5.550700	0.0000
LHUR(-2)	3.043088	0.879746	3.459053	0.0007
LHUR(-3)	-0.377371	0.911648	-0.413944	0.6795
LHUR(-4)	-0.248424	0.460506	-0.539459	0.5903
R-squared	0.366278	Mean dependent var		0.017016
Adjusted R-squared	0.335175	S.D. dependent var		1.707934
S.E. of regression	1.392595	Akaike info criterion		3.551122
Sum squared resid	316.1094	Schwarz criterion		3.715817
Log likelihood	-296.3965	Hannan-Quinn criter.		3.617943
F-statistic	11.77630	Durbin-Watson stat		1.976868
Prob(F-statistic)	0.000000			

Come prima, il test F consente di rigettare ad un elevato livello di significatività (p-value pari a 0.000) l'ipotesi che i valori dei parametri stessi siano tutti nulli. Inoltre, la maggior parte dei coefficienti sono statisticamente significativi. Possiamo spingerci più in là e sottoporre a verifica

l'ipotesi che i coefficienti delle variabili x_t , ovvero $\delta_1, \dots, \delta_4$ nel nostro esempio, siano tutti pari a zero. Questo tipo di test sottopone a verifica l'ipotesi che, ai fini della previsione di y_t , altri regressori, quali i ritardi di $LHUR_t$, non abbiano alcun potere 'aggiuntivo'. Sono noti come **Granger-causality** test e, quando l'ipotesi nulla può essere rigettata, si dice che x_t 'Granger-causi' y_t . Bisogna tuttavia prendere un po' con le pinze il termine 'causality', che in realtà sta ad indicare semplicemente una capacità previsiva di y_t da parte di ritardi di x_t , in quanto altre variabili potrebbero causare sia l'una che le altre. Nel nostro esempio, possiamo dire che il tasso di disoccupazione 'Granger-causa' il tasso di inflazione ma non possiamo asserire che ci sia un vero e proprio nesso di causalità diretto.

Il fatto che il tasso di disoccupazione Granger-causi il tasso di inflazione è comunque in sintonia con l'idea centrale della curva di Phillips, possiamo quindi concludere che le nostre stime confermano questa relazione? No, ciò non è possibile se non prima aver verificato che i coefficienti dei ritardi del tasso di disoccupazione siano in accordo con la relazione tra le variabili prevista dalla curva stessa. Lo sono? Il coefficiente del primo ritardo certamente lo è, in quanto negativo e in misura statisticamente significativa, ma quello del coefficiente del ritardo di ordine secondo appare in contrasto con la curva di Phillips in quanto significativamente positivo. I coefficienti degli altri due ritardi sono a loro volta negativi. Dunque come interpretare le nostre stime ai fini dell'inferenza del segno della relazione tra variazione del tasso di inflazione e tasso di disoccupazione? Per rispondere a questa domanda abbiamo bisogno del cosiddetto **moltiplicatore dinamico cumulato**, che traccia l'effetto cumulato nel tempo di uno shock unitario ad una delle variabili, ovvero y_t e x_t . Nel nostro caso il moltiplicatore dinamico cumulato di una variazione del tasso di disoccupazione è $\delta_1 + \delta_2 + \delta_3 + \delta_4 = -2.635568 + 3.043088 - 0.377371 - 0.248424 \cong -0.21$. Ciò vuol dire che, secondo le nostre stime, un incremento dell'un per cento del tasso di disoccupazione comporta una caduta dello 0.21 per cento del tasso di inflazione, quindi in accordo con la 'previsione' della curva dei Phillips.

Medie mobili e modelli ARMA

I processi AR(p) sono a loro volta un caso speciale delle cosiddette medie mobili autoregressive **ARMA(p,q)**, ovvero autoregressive moving averages, di ordine autoregressivo p e con media mobile di ordine q . La loro formulazione generale è la seguente:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \theta_1' u_{t-1} + \dots + \theta_q' u_{t-q} + u_t$$

A sua volta, modelli ARMA(0,q) sono noti come medie mobili MA(q), ovvero moving averages di ordine q . Un risultato fondamentale della teoria dei processi stocastici, noto come il Teorema di Wold, è che qualunque processo che abbia varianza finita può essere rappresentato come un processo MA(q), purchè q sia sufficientemente grande.

2.5 Il problema della previsione

Ma ora dobbiamo porci delle domande ad un tempo moltogenerali e molto ‘pratiche’. Come lo selezioniamo, ‘ex ante’, un buon modello? E che vuol dire ‘buon’ modello? È sempre vero che a una buona performance in-sample corrisponde una buona performance out-of-sample?

In un contesto in cui si lavori con serie temporali e l’obiettivo principale sia quello di effettuare previsioni, il tema dell’individuazione delle relazioni di causa ed effetto diventa di secondaria importanza ed acquista invece preminenza il tema della validità ‘esterna’ del modello, intesa come la sua abilità (o meno) di fornire una rappresentazione ‘in forma ridotta’ ma quanto più possibile affidabile dell’DGP in futuro, anche se prossimo, e non solo di un campione più o meno ampio di osservazioni (passate). Per rappresentazione ‘in forma ridotta’ si intende un modello che rappresenti la distribuzione condizionale delle variabili prescindendo dalle relazioni di causalità che la determinano.

Un modello in forma ridotta, dunque, fornisce indirettamente una rappresentazione del DGP, e dunque dei ‘veri’ nessi di causalità, tramite l’identificazione di correlazioni e interdipendenze utili ai fini della previsione.

La minaccia principale alla validità esterna sono, in buona sostanza, la incostanza e/o inaffidabilità delle correlazioni e interdipendenze individuate. Da questo punto di vista, un problema insidiosissimo è la circostanza che criteri di selezione quali l’ R^2 , anche in presenza dell’aggiustamento classico per tenere conto del numero di regressori, tendono a privilegiare l’impiego a fini previsivi di modelli over-parametrizzati con elevata performance esplicativa in-sample ma limitata performance previsionale out-of-sample. In sostanza, ciò è dovuto al fatto che è difficile discriminare tra la variabilità di della variabile dipendente y_t dovuta alla sua componente di errore, ovvero alla variabilità di u_t , e la variabilità dovuta a fluttuazioni del valore atteso condizionale del processo, ovvero a variazioni in $y_{t|t-1} = E(y_t | I_{t-1})$.

È quest’ultima la componente della variabilità di y_t che un buon modello deve spiegare, non la componente legata ad u_t ma, per costruzione, l’ R^2 di un modello che fa l’opposto può in alcune circostanze essere più elevato persino dell’ R^2 del modello vero, ovvero di quello corrispondente al DGP! Ciò si verifica ogni qual volta, come spesso succede, la varianza di u_t è maggiore della varianza di $y_{t|t-1} = E(y_t | I_{t-1})$.

Come ovviare a questo problema? Ci sono grosso modo due soluzioni complementari:

- si utilizzano, ai fini della selezione del modello previsivo, criteri diversi dall’ R^2 , soprattutto i criteri noti come AIC ed SBC, per confrontare il potere esplicativo ‘in-sample’ di modelli alternativi
- si ricorre a verifiche della validità esterna del modello basate su valutazioni della capacità previsiva di stime pseudo out-of-sample (*poos*) prodotte dal modello stesso

2.6 Criteri di selezione

L'idea dei criteri di selezione è di bilanciare il rischio di selezionare un modello che produce stime distorte, a causa del numero ridotto di variabili/parametri, e quello di selezionare un modello che produce stime troppo volatili, a causa di un numero eccessivo di variabili/parametri (overfitting) che portano il modello a spiegare gli errori casuali piuttosto che la variazione del valore atteso condizionale. I due criteri di selezione più diffusi sono l'Akaike Information Criterion (**AIC**) e lo Bayesian Information Criterion (**BIC**)

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

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

Qui p denota il numero di parametri e SSR , acronimo per sum of squared residuals, è la somma dei residui al quadrato del modello preso in considerazione. I due criteri hanno origini concettuali differenti ma la logica è molto simile. Il primo termine in entrambi i criteri decresce in p e riflette il fatto che il potere esplicativo (ovvero l' R^2) di un modello aumenta all'aumentare del numero di regressori. Il secondo termine invece cresce al crescere di p . È una sorta di penalità per l'utilizzo eccessivo di regressori e parametri e riflette l'aumento della varianza della previsione collegata all'aumento del numero di parametri. Si noti che il modello è tanto 'migliore' quanto più BASSO è il criterio, sia nel caso dell'AIC che in quello dello BIC.

Quale dei due criteri è più appropriato? Dipende...BIC è consistente nel senso che, al crescere dell'ampiezza del campione, BIC seleziona il modello migliore con probabilità che si avvicina asintoticamente ad uno. Ma l'AIC, a causa di una 'penalità' minore, seleziona con probabilità apprezzabilmente più elevata modelli che producono previsioni meno distorte in campioni moderatamente ampi. I software econometrici più comuni (tra cui Eviews) offrono entrambi i criteri per cui il loro utilizzo è molto agevole. Ad esempio, Eviews produce i seguenti AIC, BIC e R^2 per modelli AR(p) con $p = 1, 2, \dots, 6$:

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

Si noti che AIC seleziona un modello meno parsimonioso di BIC ma più parsimonioso di quello selezionato dall' R^2 .

2.7 Incertezza e intervalli di confidenza previsionali

Nel valutare il grado di accuratezza che è possibile aspettarsi da una previsione e costruire intervalli di confidenza intorno al valore predetto occorre una misura di incertezza previsionale. Si consideri la previsione

$$\hat{y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 y_T + \hat{\beta}_2 x_T$$

relativa al modello $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + u_t$. L'errore di previsione è

$$y_{T+1} - \hat{y}_{T+1|T} = u_{T+1} + \left[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) y_T + (\beta_2 - \hat{\beta}_2) x_T \right]$$

Il valore atteso del quadrato di questa differenza è una delle misure di incertezza previsionale più diffuse. È noto come **mean squared forecast error (MSFE)**,

$$MSFE = E(y_{T+1} - \hat{y}_{T+1|T})^2 = \underbrace{\text{var}(u_{T+1})}_{\text{Incertezza 'intrinseca'}} + E \left[\underbrace{(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) y_T + (\beta_2 - \hat{\beta}_2) x_T}_{\text{Incertezza di campionamento dei parametri del modello}} \right]^2$$

e può essere decomposto nella somma della varianza del termine di errore, e dunque all'incertezza di campionamento di u , e in una componente addizionale collegata all'errore di stima e dunque all'incertezza di campionamento dei parametri del modello. In campioni ampi, la parte connessa all'errore di stima risulta tipicamente (molto) minore di $\text{var}(u_T)$ nel qual caso abbiamo

$$MSFE \cong \text{var}(u_T)$$

Al posto del MSFE, si usa spesso la sua radice quadrata, ovvero il **root mean squared forecast error (RMSFE)**

$$RMSFE = \sqrt{E(y_{T+1} - \hat{y}_{T+1|T})^2}$$

Il RMSFE fornisce una misura intuitiva di dispersione della distribuzione dell'errore di previsione, essenzialmente analogo alla deviazione standard di u_t con la differenza che il valore atteso rispetto al quale viene definito l'errore non è dato dalla retta di regressione che caratterizza il DGP ma bensì il valore predetto utilizzando le stime puntuali dei parametri. In quanto tale, può essere

interpretato come una misura dell'ordine di grandezza dell'errore di previsione tipico. Può essere stimato in tre modi:

1. si può usare l'approssimazione

$$RMSFE = \sqrt{MSFE} \cong \sqrt{\text{var}(u_{T+1})} = \sigma_u$$

e dunque stimarlo utilizzando l'errore standard della regressione $\hat{\sigma}_u$;

2. si possono usare serie storiche costituite da previsioni effettuate in passato e dai valori assunti della variabile oggetto della previsione, per ciascun periodo $t = t_1, \dots, T$, e costruire lo stimatore seguente

$$RMSFE = \sqrt{\frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} (y_{T+1} - \hat{y}_{T+1|T})^2}$$

ma questo non è un metodo molto pratico perchè richiede che qualcuno abbia materialmente effettuato le previsioni in $t = t_1, \dots, T$;

3. si possono usare serie storiche costituite da previsioni simulate, ovvero POOS, effettuate come lo sarebbero state se il modello che si intende valutare fosse stato in uso in passato, e confrontarle con i valori assunti della variabile oggetto della previsione alle date corrispondenti, per poi costruire lo stimatore di cui al punto precedente.

Una volta ottenuto una stima di RMSFE, un intervallo previsionale al 95% può essere costruito molto semplicemente come segue:

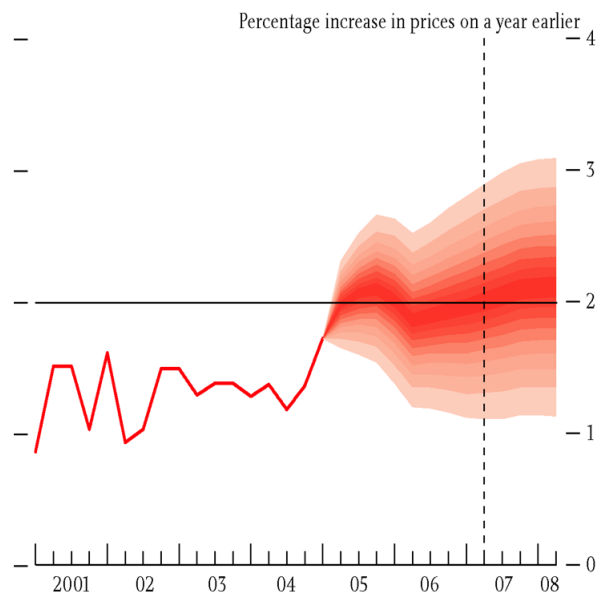
$$\hat{y}_{T|T-1} \pm 1.96 \times RMSFE$$

Le previsioni riportate nel grafico in Figura 5 sono ovviamente POOS giacchè nei lontani anni 60 non ero nemmeno nato e non ho iniziato ad occuparmi di econometria prima degli albori di questo decennio/secolo/millennio. Nel costruire gli intervalli di confidenza riportati nel grafico, si è seguito il metodo di cui al punto 1, ovvero si è sfruttato il fatto che $MSFE \cong \text{var}(u_{T+1})$ e dunque si è semplicemente posto che $RMSFE = \hat{\sigma}_u$. Si tenga però presente che l'intervallo costruito come mostrato sopra è valido solo se u_{T+1} è normalmente distribuito, altrimenti ne rappresenta un'approssimazione più o meno valida.

Con le opportune modifiche, questo impianto concettuale può essere utilizzato per costruire intervalli di previsione per previsioni multi-periodali, ovvero per $\hat{y}_{T+h|T}$ quando $h > 1$. Un esempio molto noto è il grafico seguente, che riporta le previsioni di inflazione pubblicate dalla Bank of England.

Figura 6

Serie storica del tasso di inflazione e previsioni multiperiodali al Novembre 2004 pubblicate dalla Bank of England



2.8 Piccola digressione: causalità e previsione

A questo punto potrebbe venire da chiedersi...ma perchè non buttiamo dentro sul lato destro della regressione anche un valore non ritardato di x_t ? Effettivamente non c'è nulla che lo vieti e potrebbe avere moltissimo senso farlo. Modelli del genere sono noti come **distributed lag models DL(p,r)**:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \delta_0' x_t + \delta_1' x_{t-1} + \dots + \delta_r' x_{t-r} + u_t$$

Solo che nel prendere questa direzione occorre rendersi conto che si abbandona un ambito in cui la preoccupazione principale è quella di formulare previsioni e si torna nell'alveo della inferenza di relazioni di causa ed effetto. Ovvero, a seguito di una scelta del genere, non è possibile più asserire che il problema della validità interna dell'inferenza econometrica sia secondario.

In questo caso, occorre dunque porsi con forza la domanda se le x_t siano esogene. Quando le variabili sul lato destro sono ritardi, e dunque nei modelli ADL(p,r), la condizione di **esogeneità debole** è soddisfatta (quasi) in automatico. Ciò non è invece il caso in presenza di variabili

esplicative non ritardate, e dunque nei modelli DL(p,r), in quando queste potrebbero a loro volta essere determinate dalla variabile dipendente, nel qual caso avremmo un problema di causalità inversa (*reverse causality*), oppure entrambe le variabili potrebbero essere determinate congiuntamente, come per esempio la domanda e l'offerta di un bene a seguito del conseguimento da parte del relativo mercato di uno stato di equilibrio. Se la condizione di esogeneità debole non fosse soddisfatta, ci imatteremmo in grosse difficoltà che, se non adeguatamente affrontate, comporterebbero l'inconsistenza dello stimatore OLS. Ciò, in sostanza, è dovuto al fatto che, come già spiegato nella prima parte di queste note, verrebbe a mancare la condizione di identificazione del nesso di causalità che lega le variabili. Esistono diversi modi di affrontare queste difficoltà, che vanno dall'uso di variabili strumentali alla modellizzazione esplicita delle relazioni tra variabili in un contesto multivariato e multiequazione. Nella sezione seguente accenneremo a tali modelli ma una soluzione alternativa, più semplice ma a volte sensata, è di tenersi semplicemente alla larga da questi problemi evitando di considerare variabili esplicative dalla dubbia esogeneità.

2.9 Modelli VAR

I cosiddetti modelli VAR (vecor autoregression) generalizzano i modelli AR ed ADL ammettendo più di una variabile endogena. Nel caso di un VAR con m variabili endogenee \mathbf{y}_t e variabili endogenee represented by the elements of \mathbf{x}_t , abbiamo

$$\mathbf{y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{y}_{1,t-1} + \dots + \boldsymbol{\beta}_p \mathbf{y}_{t-p} + \boldsymbol{\delta}_0' \mathbf{x}_t + \boldsymbol{\delta}_1' \mathbf{x}_{t-1} + \dots + \boldsymbol{\delta}_r' \mathbf{x}_{t-r} + \mathbf{u}_t$$

Ovvero:

$$\begin{aligned} y_{1,t} &= \beta_0 + \beta_{1,1} y_{1,t-1} + \dots + \beta_{1,p} y_{1,t-p} \\ &\quad + \beta_{2,1} y_{1,t-1} + \dots + \beta_{2,p} y_{2,t-p} \\ &\quad \dots \\ &\quad + \beta_{m,1} y_{1,t-1} + \dots + \beta_{m,p} y_{m,t-p} \\ &\quad + \delta_0' \mathbf{x}_t + \delta_1' \mathbf{x}_{t-1} + \dots + \delta_r' \mathbf{x}_{t-r} + u_{1,t} \\ y_{2,t} &= \beta_0 + \beta_{1,1} y_{1,t-1} + \dots + \beta_{1,p} y_{1,t-p} \\ &\quad + \beta_{2,1} y_{1,t-1} + \dots + \beta_{2,p} y_{2,t-p} \\ &\quad \dots \\ &\quad + \beta_{m,1} y_{1,t-1} + \dots + \beta_{m,p} y_{m,t-p} \end{aligned}$$

$$\begin{aligned}
& +\delta_0'x_t + \delta_1'x_{t-1} + \dots + \delta_r'x_{t-r} + u_{2,t} \\
& \dots \\
y_{m,t} = & \beta_0 + \beta_{1,1}y_{1,t-1} + \dots + \beta_{1,p}y_{1,t-p} \\
& +\beta_{2,1}y_{1,t-1} + \dots + \beta_{2,p}y_{2,t-p} \\
& \dots \\
& +\beta_{m,1}y_{1,t-1} + \dots + \beta_{m,p}y_{m,t-p} \\
& +\delta_0'x_t + \delta_1'x_{t-1} + \dots + \delta_r'x_{t-r} + u_{3,t}
\end{aligned}$$

Come notato più sopra, è essenziale assicurarsi che le variabili esogene, se presenti, siano effettivamente tali, per evitare problemi di causalità inversa che renderebbero lo stimatore OLS inconsistente.

I modelli VAR sono rappresentazioni in forma ridotta di modelli strutturali del DGP. Come tali, per quanto utili, non danno alcuna indicazione diretta sui nessi di causalità tra le variabili. Per inferire tali nessi, possiamo tuttavia utilizzare le stime di un modello VAR in modo indiretto. Far ciò sfrutta le relazioni tra modello in forma ridotta e modello in forma strutturale e richiede quanto segue:

- i. specifichiamo in maniera chiara le relazioni di causalità che ipotizziamo e sulla cui esistenza vogliamo fare inferenza econometrica
- ii. formuliamo in maniera quanto più dettagliata possibile il modello strutturale che implica le relazioni di causalità ipotizzate
- iii. deriviamo il modello in forma ridotta implicato dal modello strutturale, sia per quanto riguarda la forma funzionale del modello che eventuali restrizioni sui suoi parametri
- iv. confrontiamo il modello ridotto derivato dal modello strutturale col modello in forma ridotta stimato
- v. rigettiamo il modello strutturale nel caso in cui tali modelli siano significativamente diversi

Ovvero, limitandoci per semplicità a considerare il caso di un modello VAR(1), ovvero con $p = 1$, il modello in forma ridotta può essere derivato come segue:

$$\begin{aligned}
\underbrace{\mathbf{A}}_{m \times m} \underbrace{\mathbf{y}_t}_{m \times 1} &= \underbrace{\mathbf{c}}_{m \times m} + \underbrace{\mathbf{B}}_{m \times m} \underbrace{\mathbf{y}_{t-1}}_{m \times 1} + \underbrace{\mathbf{e}_t}_{m \times 1} \\
\Rightarrow \mathbf{y}_t &= \underbrace{\mathbf{A}^{-1}\mathbf{c}}_{\beta_0} + \underbrace{\mathbf{A}^{-1}\mathbf{B}}_{\beta_1} \mathbf{y}_{t-1} + \underbrace{\mathbf{A}^{-1}\mathbf{e}_t}_{\mathbf{u}_t} \\
\Rightarrow \mathbf{y}_t &= \beta_0 + \beta_1 \mathbf{y}_{t-1} + \mathbf{u}_t
\end{aligned}$$

In particolare, abbiamo posto $\mathbf{u}_t = \mathbf{A}^{-1}\mathbf{e}_t$. Questo significa che gli errori del modello strutturale sono

$$\mathbf{e}_t = \mathbf{A}\mathbf{u}_t$$

In molti casi, questi errori sono proprio ciò che ci interessa studiare. Questo è il caso, ad esempio, quando tali errori rappresentino degli shock che investono un sistema economico oggetto di studio e che ne determinino la deviazione da una condizione di equilibrio, $\bar{\mathbf{y}}$, come nell'ECM seguente

$$\begin{aligned} \underbrace{\mathbf{A}}_{m \times m} \underbrace{\mathbf{y}_t}_{m \times 1} &= \underbrace{\mathbf{c}}_{m \times m} + \underbrace{\mathbf{B}}_{m \times m} \underbrace{(\mathbf{y}_{t-1} - \bar{\mathbf{y}})}_{m \times 1} + \underbrace{\mathbf{e}_t}_{m \times 1} \\ &= \underbrace{(\mathbf{c} - \bar{\mathbf{y}})}_{m \times m} + \underbrace{\mathbf{B}}_{m \times m} \underbrace{\mathbf{y}_{t-1}}_{m \times 1} + \underbrace{\mathbf{e}_t}_{m \times 1} \end{aligned}$$

Nel modello strutturale, gli errori sono indipendenti gli uni dagli altri. Questa è un'assunzione importante per identificare il modello strutturale (se non fossero indipendenti l'uno dall'altro non sarebbero gli shock 'originari'). La loro matrice varianza-covarianza, $Var(\mathbf{e}_t) := \boldsymbol{\Sigma}_e = \mathbf{D}\mathbf{D}'$, è dunque diagonale e rappresentabile come il prodotto della matrice delle deviazioni standard degli errori strutturali, \mathbf{D} , e della sua trasposta, anch'esse (ovviamente) diagonali. Derivare questi errori a partire da quelli del modello in forma ridotta richiede però la conoscenza di \mathbf{A} . Un modo semplice di ottenere questi errori è di assumere che \mathbf{A} sia rappresentata da una matrice cosiddetta "triangolare". In questo caso, la matrice varianza-covarianza degli errori del modello in forma ridotta è decomponibile, alla maniera di Cholesky, in una matrice triangolare e nella sua trasposta, ovvero

$$\begin{aligned} Var(\mathbf{u}_t) &= Var(\mathbf{A}^{-1}\mathbf{e}_t) = \mathbf{A}^{-1} Var(\mathbf{e}_t)\mathbf{A}^{-1'} = \mathbf{A}^{-1} \boldsymbol{\Sigma}_e \mathbf{A}^{-1'} = \mathbf{A}^{-1} \mathbf{D}\mathbf{D}'\mathbf{A}^{-1'} \\ &= (\mathbf{A}^{-1}\mathbf{D})(\mathbf{A}^{-1}\mathbf{D})' \end{aligned}$$

Poiché è facilissimo ottenere una decomposizione di Cholesky (è una funzione di algebra matriciale standard in qualsiasi software di calcolo, incluso Excel), possiamo facilmente ottenere $\mathbf{A}^{-1}\mathbf{D}$, che è ciò che ci serve per derivare una versione standardizzata (ovvero con varianza unitaria) degli errori strutturali, $\mathbf{D}^{-1}\mathbf{e}_t$, a partire da quelli in forma ridotta,

$$\mathbf{D}^{-1}\mathbf{e}_t = \mathbf{D}^{-1}\mathbf{A}\mathbf{u}_t = (\mathbf{A}^{-1}\mathbf{D})\mathbf{u}_t$$

Questi errori del modello strutturale standardizzati sono trasformazioni lineari degli errori del modello in forma ridotta e dunque esibiscono una dipendenza temporale che è funzione della correlazione seriale di questi ultimi. La sequenza di errori strutturali in funzione del tempo trascorso dopo uno shock (strutturale) iniziale è nota come "impulse response function".

2.10 Variabili non-stazionarie: trend e discontinuità strutturali

Finora abbiamo di fatto assunto di avere a che fare con processi stazionari. Ora invece ci occupiamo di come stabilire se un dato processo sia stazionario o meno e cosa fare in quest'ultimo caso. I due tipi principali di non-stazionarietà sono i seguenti:

- discontinuità strutturali (*structural breaks*)
- presenza di una 'tendenza' (*trend*)

In queste note ci occuperemo prima, abbastanza succintamente, del problema delle rotture strutturali e poi di quello dei trend. Per una discussione più approfondita di entrambi i temi, si rimanda al libro di testo.

2.11 Discontinuità strutturali (structural breaks)

Per structural break si intende **il cambiamento del DGP da un periodo ad un'altro**. In modelli parametrici come quelli di cui ci occupiamo in queste note, ciò si traduce in un cambiamento dei parametri del modello stesso e dunque, con riferimento ad un modello di regressione, dei coefficienti di quest'ultimo.

Dovrebbe essere abbastanza chiaro cosa fare se siamo a conoscenza della presenza di uno structural break ad una certa data, ovvero stimeremo due modelli diversi per i periodi prima e dopo la data in cui il break si manifesta.

Ma come fare ad individuare la presenza di un break?

Un buon punto di partenza, al solito, è la visualizzazione delle serie oggetto di indagine e l'ispezione dei relativi grafici.

Ma ci sono anche metodi più formali per verificare la presenza di discontinuità di questo tipo. Possiamo distinguerne due grandi classi:

- test formali per la presenza di structural breaks
- analisi della performance di previsioni POOS e confronto con la performance in-sample del modello

Alcuni test per la presenza di uno structural break

Si consideri di nuovo l'esempio della curva di Phillips e si consideri il seguente semplice modello ADL(1,1) della relazione tra variazioni del tasso di inflazione e disoccupazione

$$\Delta Infl_t = \beta_0 + \beta_1 \Delta Infl_{t-1} + \gamma_1 LHUR_{t-1} + u_t$$

Si supponga di avere buone ragioni per ritenere che il break sia occorso in data τ . La stabilità dei coefficienti può essere sottoposta a test introducendo delle variabili *dummy*, come mostrato di seguito

$$\begin{aligned} \Delta Infl_t = & \beta_0 + \beta_1 \Delta Infl_{t-1} + \delta_1 LHUR_{t-1} + \\ & + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times \Delta Infl_{t-1}] + \gamma_2 [D_t(\tau) \times LHUR_{t-1}] + u_t \end{aligned}$$

ove $D_t(\tau) = 1$ se $t \geq \tau$ e $D_t(\tau) = 0$ in caso contrario. Dunque, avremmo costanza dei coefficienti su tutto il periodo di campionamento se risultasse $\gamma_0 = \gamma_1 = \gamma_2 = 0$ ma, se almeno uno di questi coefficienti assumesse un valore non nullo, vorrebbe dire che la funzione di regressione, ovvero

$\Delta Infl_t = \beta_0 + \beta_1 \Delta Infl_{t-1} + \gamma_1 LHUR_{t-1} + u_t$, è soggetta ad una discontinuità strutturale in data τ . Uno dei test utilizzati per verificare se questo sia il caso è il **test di Chow** basato su un F -test della ipotesi nulla

$$H_0 : \gamma_0 = \gamma_1 = \gamma_2 = 0$$

contro l'ipotesi alternativa che se almeno uno di questi coefficienti assuma un valore non nullo.

In Eviews questo test è facilissimo da condurre. Nel nostro esempio, una volta stimato il modello ADL(1,1) occorre semplicemente selezionare la 'view' apposita ("Stability diagnostics") e quindi selezionare "Chow breakpoint test". Nella finestra che salta fuori, si possono inserire quante date si vuole come possibili 'break points'. Eviews sottoporrà a verifica l'ipotesi che non ci sia stato uno structural break in alcuna di queste date. Qui di seguito si riporta, a mo di esempio, lo 'screenshot' di Eviews relativo alle date 1980:Q1, 1981:Q1, 1982:Q1 e 1983:Q1.

Eviews screenshot:

Chow Breakpoint Test: 1980Q1 1981Q1 1982Q1 1983Q1
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 1957Q4 2005Q1

F-statistic	2.368340	Prob. F(12,175)	0.0075
Log likelihood ratio	28.59257	Prob. Chi-Square(12)	0.0045

Lo studente è invitato a sperimentare **cosa succede se, invece di effettuare un test per tutte e quattro le date, si sottopone a verifica l'ipotesi nulla per una data alla volta**. Come potrà notarsi, **il risultato sarà molto differente**. Ciò suona un campanello di allarme circa i limiti di test che, come quello di Chow, assumono che la data della discontinuità sia nota. Molto spesso tale 'conoscenza' deriva dalla semplice ispezione del grafico della serie considerata e quindi, di fatto, è il risultato di una specie di stima, pur se rudimentale. La conseguenza è che, quando la selezione della possibile break date sia il risultato di un procedimento induttivo e/o di stima rudimentale e non invece di una deduzione rigorosa da una 'teoria' che si voglia sottoporre a verifica e che faccia parte integrante dell'ipotesi nulla, l'errore di campionamento della statistica F del test dovrebbe tenere conto anche dell'incertezza insita in questa stima della data di accadimento della possibile discontinuità.

Per ovviare almeno in parte a questi problemi, sono stati proposti diversi altri test, come per esempio il Quandt Likelihood Ratio (QLR) test, anche noto come "sup-Wald" test e disponibile in Eview sotto la dicitura "**Quandt-Andrews breakpoint test**" (anche se a mio parere l'algoritmo usato in Eviews ha qualche problema di stabilità numerica). Il test, in buona sostanza, utilizza la statistica F maggiore tra quelle corrispondenti alle varie possibili date (tralasciando un gruppo di date all'inizio e alla fine del periodo in esame per le quali non ci sia un campione di dati sufficientemente ampio per stimare sia il modello pre-break che quello post-break), e la

distribuzione di tale statistica nell'ipotesi nulla che nessuna di queste date sia una data in cui si verifica un break strutturale.

Quelli appena esaminati non sono gli unici tests per verificare la presenza di discontinuità strutturali e, più in generale, per sottoporre a verifica l'ipotesi della stabilità o meno del modello di regressione. Una lista non esaustiva di altri test, sui cui dettagli si rimanda alla guida in linea di Eviews e che si invita lo studente ad esplorare e a sperimentarvici, è disponibile nella view "Stability diagnostics" del software stesso, che è possibile selezionare una volta stimata una determinata equazione.

Previsioni POOS e discontinuità strutturali

Ovviamente, in presenza di una discontinuità strutturale, la performance previsiva out-of-sample risulterà inferiore a quella in sample. Questa semplice idea può essere utilizzata per sfruttare previsioni POOS per valutare se il processo in esame sia soggetto a discontinuità strutturali.

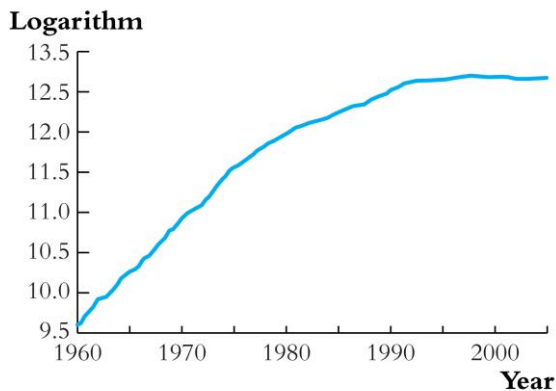
Ad esempio, se volessimo valutare la stabilità dei parametri del modello AR(4) della variazione del tasso di inflazione considerato prima, potremmo dare un'occhiata al grafico in Figura 5 e chiederci se la previsioni POOS siano sempre entro l'intervallo di previsione al 95%. In questo caso, con poche eccezioni, lo sono. Inoltre, le poche occasioni in cui la variazione del tasso di inflazione osservata 'sfora' l'intervallo di previsione al 95% all'insù o all'ingiù appaiono casi sporadici e non fenomeni persistenti, ovvero **lo 'sforamento' in un periodo non sembra precludere allo sforamento in periodi immediatamente successivi** e, in ogni caso, la percentuale di 'sforamenti' è ben inferiore al 5%, ovvero a $1 - 95\%$. Sebbene ciò non costituisca un test formale di stabilità dei parametri, può offrire indicazioni complementari a quelle fornite da uno di questi test, che risultano particolarmente utili in considerazione del fatto che, per ovvi motivi, l'affidabilità di test quali il test di Chow o il test QLR tende a scemare quanto più la discontinuità ipotizzata è vicina alla fine del campione...ma queste sono proprio le date più interessanti dal punto di vista della valutazione della capacità previsiva del modello!

2.12 Trend deterministici e stocastici

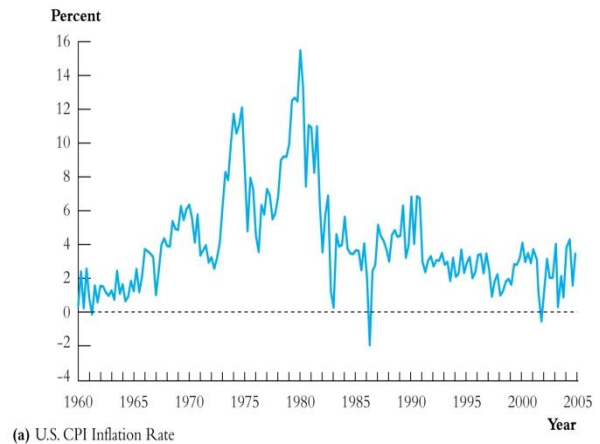
Chiediamoci:

1. Che cos'è un trend?
2. Quali problemi sono causati dalla presenza di un trend e come si possono affrontare/attenuare?
3. Come si rileva la presenza di un trend?

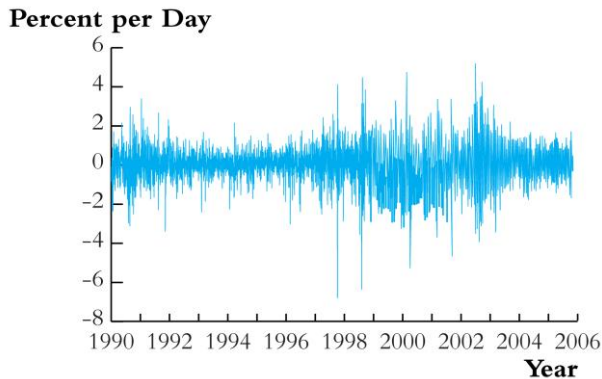
Un trend è un movimento a lungo termine o tendenza nei dati. I trend non sono necessariamente fenomeni lineari, ovvero graficamente non devono necessariamente essere rappresentabili con una linea retta! Vediamo, quale di queste serie ha una tendenza?



(c) Logarithm of GDP in Japan



(a) U.S. CPI Inflation Rate



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

Delle tre serie:

- il (log) PIL del Giappone ha un trend positivo di lungo periodo anche se la tendenza alla crescita decresce nel tempo

- l'inflazione è invece soggetta a lunghi movimenti all'insù o all'ingiù e non è affatto ovvio se esista un livello di lungo periodo cui tenda, prima o poi, a ritornare;
- le variazioni giornaliere percentuale dell'indice azionario del NYSE (New York stock exchange) non sembrano avere alcun trend (ci sono forti sbalzi della volatilità, ma ciò è un fenomeno diverso)

Formalmente, diremmo che il PIL del Giappone è soggetto a un **trend deterministico**, quello del tasso di inflazione a un **trend stocastico**.

Un **trend deterministico** è una tendenza definita da una funzione deterministica, ovvero senza componenti aleatorie, del trascorrere del tempo, ad esempio $E(y_t) = t$, or $E(y_t) = t^2$.

Un **trend stocastico** invece varia nel tempo in maniera aleatoria. Un esempio importante di processo stocastico caratterizzato dalla presenza di un **trend stocastico** è la 'passeggiata casuale', o **random walk**,

$$y_t = y_{t-1} + u_t,$$

dove u_t è serialmente non correlato. Quando y_t è caratterizzato da questo tipo di processo, il valore futuro di y_t è il valore odierno con l'aggiunta di una innovazione di entità imprevedibile. Ciò dà luogo a due proprietà molto importanti:

- $y_{t+h|t} = y_t$, ovvero la migliore stima del valore di y in futuro è il valore di y oggi; ad una prima approssimazione, i (logaritmi dei) prezzi delle azioni seguono un random walk (più precisamente, i rendimenti delle attività finanziarie sono imprevedibili)
- $var(y_{T+h|T} - y_T) = var(u_{T+1} + u_{T+2} + \dots + u_{T+h}) = h\sigma_u^2$, e dunque

$$\begin{aligned} MSFE &= E(y_{T+h} - \hat{y}_{T+h|T})^2 \\ &\cong var(u_{T+1} + u_{T+2} + \dots + u_{T+h}) \\ &= h \times \sigma_u^2 \end{aligned}$$

ovvero la varianza dell'errore previsionale aumenta linearmente al crescere dell'orizzonte previsionale; dunque, più distante nel tempo è la previsione, maggiore è l'incertezza previsionale (tecnicamente, questa è proprio la circostanza che spinge a definire la serie come "non stazionaria")

Un'altro esempio importante di processo stocastico caratterizzato dalla presenza di un **trend stocastico** è la 'passeggiata casuale con tendenza', o **random walk with drift**,

$$y_t = \beta_0 + y_{t-1} + u_t,$$

dove u_t è serialmente non incorrelato. Quando y_t è caratterizzato da questo tipo di processo, che ricorre frequentemente (almeno in una prima approssimazione) in molti fenomeni economici,

segue un random walk ‘intorno’ ad un trend lineare. Ad esempio, la previsione un numero h di periodi più in là è data da

$$y_{T+h/T} = \beta_0 h + y_T$$

Confrontiamo questo processo con un AR(1):

$$y_t = \beta_0 + y_{t-1} + u_t,$$

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t,$$

Dunque il **random walk con drift non è altro che un AR(1) con $\beta_1 = 1$** . Per questo motivo diciamo che il random walk con drift è un AR(1) con radice unitaria. Si veda l’appendice per una derivazione delle proprietà di lungo periodo di un AR(1) con o senza radice unitaria.

Il termine radice unitaria viene dal fatto che, utilizzando l’operatore ritardo, possiamo rappresentare un processo AR(1) senza drift come segue

$$y_t = \beta_1 L y_t + u_t$$

e dunque abbiamo che $E(y_t) = \beta_1 L E(y_t)$, ovvero

$$E(y_t)(1 - \beta_1 L) = 0$$

Questa condizione è soddisfatta quando $E(y_t) = 0$ oppure quando $1 - \beta_1 L = 0$. Nel primo caso il processo è stazionario, in quanto ha valore atteso definito (e costante), mentre nel secondo ha come ‘radice’ $L = \frac{1}{\beta_1}$, che assume il valore $L = \frac{1}{\beta_1} = 1$ quando $\beta_1 = 1$, ovvero se il processo è un random walk. Si noti che, in questo caso, la condizione è vera per qualunque $E(y_t) \in \mathbb{R}$. Ciò vuol dire che, quando $\beta_1 = 1$, il valore atteso del processo è indeterminato, il che è una manifestazione della circostanza che un random walk, come più in generale qualunque processo che contenga radici unitarie, non sia un processo stazionario. La equazione $1 - \beta_1 L = 0$, a sua volta, è la cosiddetta equazione caratteristica del modello AR(1).

Più formalmente abbiamo che

$$E(y_{t+1}) = \beta_1 E(y_t)$$

$$\Rightarrow E(y_{t+1}) = \beta_1 E(\beta_1 y_{t-1} + u_t) = \beta_1^2 E(y_{t-1})$$

$$\Rightarrow E(y_{t+1}) = \beta_1^{h+1} E(y_{t-h})$$

$$\Rightarrow E(y_{t+1}) = \beta_1^{h+1} L^h E(y_t)$$

$$\Rightarrow E(y_{t+1})(1 - \beta_1^{h+1} L^h) = 0$$

E dunque,

$$E(y_{t+1}) \neq 0 \Rightarrow (1 - \beta_1^{h+1}L^h) = 0$$

$$\Rightarrow L^h = \frac{1}{\beta_1^{h+1}}$$

$$\Rightarrow L^1 = \frac{1}{\beta_1^{1+\frac{1}{h}}}$$

$$\Rightarrow L = \frac{1}{\beta_1}$$

$$\Rightarrow L = 1$$

per $\beta_1 = 1$.

Si noti che un processo può contenere più di una radice unitaria. Ad esempio, si lascia come esercizio per il lettore la dimostrazione che l'equazione caratteristica di un AR(2) sia $1 - \beta_1L - \beta_2L^2 = 0$, che può avere una o più soluzioni della forma $L = 1$ e dunque implicare la presenza di una o più radici unitarie.

Processi che contengono radici unitarie sono anche noti come processi integrati di ordine d , dove d rappresenta il numero di radici unitarie e dunque di trend stocastici distinti che caratterizzano il processo stesso, e ci si riferisce loro come processi $I(d)$.

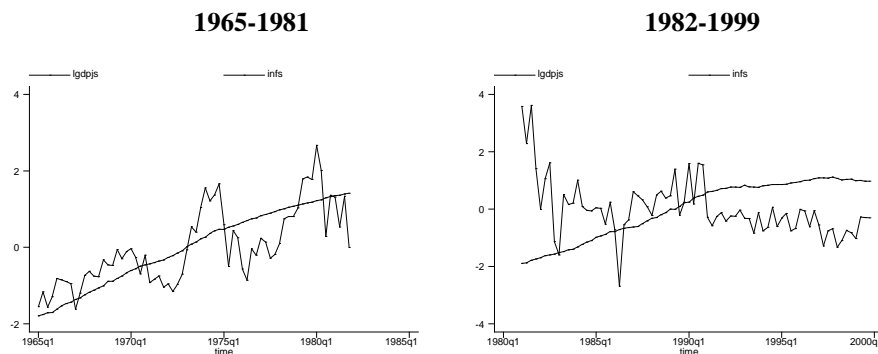
2.13 Qual'è il problema se c'è un trend?

Se c'è un trend deterministico non ci sono problemi particolari. È una variabile esplicativa come le altre, che bisognerebbe includere per non incorrere in problemi collegati all'omissione di variabili rilevanti.

Nel caso di trend stocastici, il problema fondamentale è che **il DGP non è ergodico**, il che ha essenzialmente tre tipi di conseguenze:

1. le stime OLS dei coefficienti dei ritardi in un modello autoregressivo sono distorti, in quanto OLS tende a sottostimarli in presenza di una radice unitaria;
2. le statistiche di alcuni test non hanno la distribuzione standard, anche in campioni ampi;
3. se due variabili x e y hanno entrambe un trend stocastico, possono apparire in relazione l'una con l'altra anche quando non lo sono, come nel caso dell'inflazione americane e PIL giapponese nel periodo 1965-1981 i cui grafici sono mostrati qui sotto unitamente, per confronto, ai grafici relativi al periodo 1982-1999.

Figura 7



Ora studieremo questi problemi e i loro possibili rimedi più in dettaglio.

È importante prima di tutto notare che un dato processo può contenere più di un trend stocastico. Quando i trend stocastici sono distinti, quello di ordine maggiore domina gli altri e dunque una qualunque combinazione lineare di variabili avrà un ordine di integrazione pari al maggiore degli ordini di integrazione di ciascuna variabile. Più formalmente, se $x_{i,t} \sim I(d_i)$, $i = 1, 2, \dots, k$, ovvero se abbiamo k variabili $x_{i,t}$ ciascuna integrata di ordine d_i , e poniamo

$$z_t = \sum_{i=1}^k \alpha_i x_{i,t} \quad \alpha_i \in \mathbb{R}$$

avremo che $z_t \sim I(\max d_i)$. Ora, se risolviamo questa equazione per una qualunque delle variabili, poniamo $x_{1,t}$, abbiamo

$$x_{j,t} = \sum_{i \neq j} \frac{\alpha_i}{\alpha_j} x_{i,t} + \frac{z_t}{\alpha_j} = \sum_{i \neq j} \beta_{i,t} x_{i,t} + \beta_0 + u_t \quad \alpha_i \in \mathbb{R}$$

$\beta_i \qquad \beta_0 + u_t$

il che appare molto simile ad una regressione, ed infatti lo è se scegliamo i coefficienti in maniera tale che $E(u_t | \mathbf{x}_t) = 0$. L'unico problema non da poco è che u_t è una funzione lineare di z_t , essenzialmente proporzionale a quest'ultimo in quanto $u_t = -\beta_0 + \frac{z_t}{\alpha_j}$, e dunque è $I(\max d_i)$ come

z_t . Ciò vuol dire che, in questa regressione, u_t contiene (almeno) un trend stocastico, e dunque non è nemmeno ergodico, il che mina alla radice la possibilità di effettuare inferenze nella maniera cui siamo abituati. Tutti i nostri test, infatti, richiedono stime della varianza e altri momenti secondi degli errori ma come facciamo a calcolare i momenti secondi se il momento primo, ovvero il valore atteso 'di lungo periodo', non ha un valore definito? Come possiamo sperare che il nostro stimatore del valore atteso, ovvero la media campionaria, sia uno stimatore anche solo consistente del valore atteso? E come la mettiamo con la mancanza di ergodicità, che di fatto non ci consente di interpretare un campione, per quanto ampio, come tale? È chiaro che queste difficoltà minano alle fondamenta tutta la teoria dello stimatore OLS costruita nella prima parte del corso.

Se regredissimo una sull'altra due variabili che contengano un trend stocastico, rischieremo di interpretare una temporanea correlazione tra trend stocastici distinti come un rapporto di causa ed effetto, e i test usuali non ci sarebbero di aiuto nel sottoporre a verifica ipotesi circa tale rapporto di causa ed effetto in quanto le loro statistiche hanno una distribuzione non standard, come rilevato al punto precedente.

Si noti tuttavia che **il problema non riguarda solo la regressione di variabili integrate su altre variabili integrate ma anche la regressione di una variabile integrata su variabili non integrate e viceversa**. In ognuno di questi casi, l'ordine di integrazione degli errori sarà una combinazione lineare degli ordini di integrazione della variabile dipendente e dei regressori. Ad esempio, se regrediamo una variabile $I(1)$ su una variabile $I(0)$, gli errori non potranno che essere $I(1)$, con tutte le conseguenze relative alla distribuzione non-standard delle statistiche impiegate nei vari test.

2.14 Come stabiliamo se c'è un trend stocastico?

Un buon punto di partenza è quello di dare un'occhiata a un grafico della serie.

Esistono anche vari test. I più comuni sono il test di **Dickey-Fuller** e il test di **Dickey-Fuller "augmented"**.

Il test di Dickey-Fuller per un AR(1)

Un processo AR(1) è dato da

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

Sottraendo y_{t-1} da entrambi i lati, abbiamo

$$y_t - y_{t-1} = \beta_0 + \underbrace{(\beta_1 - 1)}_{\delta} y_{t-1} + u_t$$

ovvero

$$\Delta y_t = \beta_0 + \delta y_{t-1} + u_t$$

con $\delta = \beta_1 - 1$.

Quindi, nell'ipotesi nulla che sia presente una radice unitaria contro l'ipotesi alternativa che non ci sia, abbiamo

$$H_0: \delta = 0 \text{ (cioè, } \beta_1 = 1) \text{ vs. } H_1: \delta < 0$$

NB: questo test è 'ad una coda', in quanto $\beta_1 > 1$ implicherebbe un processo esplosivo e dunque è una possibilità che escludiamo a priori, e dunque è $\beta_1 < 1$ ad implicare che y_t sia generata da un processo stazionario, il che implica a sua volta $\delta < 0$.

Per condurre il test, occorre calcolare dunque la statistica $t = \frac{\hat{\delta}}{s.e.(\hat{\delta})}$ ed individuarne il livello di significatività. Il problema è che, quando l'ipotesi nulla è valida, questa statistica non ha una distribuzione normale, nemmeno in campioni grandissimi! Segue infatti una distribuzione tutta sua e tale distribuzione cambia ulteriormente in presenza di un trend deterministico. Queste distribuzioni sono state tabulate anni orsono da Dickey e Fuller (e successivamente affinate con l'uso dei moderni computer) e i loro valori critici sono disponibili nel retro di molti libri di testo e memorizzati in praticamente tutti i software econometrici.

Il test di Dickey-Fuller per un AR(p) e altri processi relativamente 'simili'

Per un $AR(p)$ e altri processi relativamente simili, infatti nella maggioranza dei casi, si applica la stessa logica. Si rimanda al libro di testo per dettagli e al manuale del software econometrico utilizzato per istruzioni su come condurre i test.

2.15 Cosa fare in presenza di trend stocastici

Come si è visto, la presenza di trend stocastici può rappresentare un problema molto serio. Si giunge però a conclusioni molto diverse quando almeno un trend stocastico sia in comune, il che consente di individuare alcune soluzioni o, almeno, dei rimedi parziali.

Una prima soluzione: lavorare con le differenze prime

Si consideri in prima istanza un semplice processo AR(1), ovvero

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

Come appena visto, ciò implica

$$y_t - y_{t-1} = \beta_0 + \underbrace{(\beta_1 - 1)}_{\delta} y_{t-1} + u_t$$

ovvero

$$\Delta y_t = \beta_0 + \delta y_{t-1} + u_t$$

Ora, se l'equazione caratteristica del nostro AR(1), ovvero $1 - \beta_1 L = 0$, ha una radice unitaria, cioè se $L = 1$, abbiamo che $\beta_1 = 1$. In questo caso abbiamo che $\delta = 0$ e dunque il modello AR(1) diventa

$$\Delta y_t = \beta_0 + u_t,$$

ovvero un AR(0) delle differenze prime che non contiene radici unitarie.

Nel caso di un processo AR(2), abbiamo

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t$$

Sottraendo e aggiungendo $\beta_2 y_{t-1}$ sul lato destro, otteniamo

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + (\beta_2 y_{t-1} - \beta_2 y_{t-1}) + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) y_{t-1} - \beta_2 (y_{t-1} - y_{t-2}) + u_t \end{aligned}$$

e, sottraendo y_{t-1} da entrambi i lati,

$$y_t - y_{t-1} = \beta_0 + \underbrace{(\beta_1 + \beta_2 - 1)}_{\delta} y_{t-1} - \underbrace{\beta_2}_{\gamma} (y_{t-1} - y_{t-2}) + u_t$$

ovvero

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \gamma \Delta y_{t-1} + u_t$$

Ora, se l'equazione caratteristica del nostro AR(2), ovvero $1 - \beta_1 L - \beta_2 L^2 = 0$, ha una o più radici unitarie, cioè se almeno una soluzione è data da $L = 1$, abbiamo che $\beta_1 + \beta_2 = 1$. Ciò implica che $\delta = 0$ e dunque il modello AR(2) diventa

$$\Delta y_t = \beta_0 + \gamma \Delta y_{t-1} + u_t,$$

ovvero un AR(1) delle differenze prime che non contiene radici unitarie. Dunque otteniamo un risultato simile a quello ottenuto sopra differenziando il processo AR(1).

Questo risultato si applica anche nel caso più generale di un processo AR(p). Infatti, si può dimostrare che un processo AR(p)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + u_t$$

può essere riscritto come un AR($p-1$) delle differenze prime, ovvero

$$\Delta y_t = \beta_0 + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \dots + \gamma_p \Delta y_{t-p} + u_t,$$

e, se nel processo AR(p) originario vi è non più di una radice unitaria, il processo AR($p-1$) risultante non contiene trend stocastici.

Le osservazioni appena svolte suggeriscono una prima soluzione ai problemi collegati alla presenza di un trend stocastico. Ovvero, se ci convinciamo che la variabile dipendente, y_t , oppure le variabili indipendenti, x_t , o entrambe contengono un trend stocastico, la soluzione più immediata è quella di differenziare le serie e lavorare con Δy_t e Δx_t invece di y_t e x . Questo risultato può essere intuitivamente razionalizzato interpretando l'operazione di differenziazione come il sottrarre da una data serie un'altra, ovvero il proprio ritardo, caratterizzata dalla presenza dello stesso trend stocastico, col risultato che nella serie differenziata l'effetto del trend stocastico viene rimosso (in quanto i due trend uguali ma di segno contrario si annullano a vicenda).

Cointegrazione

Quest'ultima osservazione può essere anche interpretata come un caso speciale del fatto che, se due o più variabili $I(d)$ hanno un numero b di trend stocastici in comune, una loro combinazione lineare risulterà $I(d-b)$. Quando ciò si verifica, diciamo che le variabili sono **cointegrate** di ordine (d,b) e sarà sufficiente differenziare solo $d - b$ volte questa combinazione lineare (invece che d volte le variabili che la costituiscono) per ottenere una regressione con errori $I(0)$, che potrà essere impiegata in inferenze econometriche secondo la teoria OLS standard.

Si consideri ad esempio il caso di due variabili $I(1)$, poniamo y_t e x_t . La grandezza

$$z_t = \alpha_x x_t + \alpha_y y_t \quad \alpha_x, \alpha_y \in \mathbb{R}$$

rappresenta appunto una loro combinazione lineare. Ora, se il trend stocastico delle due variabili è in comune, z_t sarà $I(0)$ per una opportuna scelta di α_x e α_y e si avrà

$$y_t = -\frac{\alpha_x}{\alpha_y} x_t + \frac{z_t}{\alpha_y} = \beta_0 + \beta_1 x_t + u_t$$

$\beta_1 \qquad \beta_0 + u_t$

In tale regressione, poichè le variabili sono cointegrate e avendo scelto α_x e α_y in modo opportuno, l'errore della regressione di una sull'altra sarà $I(0)$.

Ciò suggerisce **una soluzione alternativa al problema della presenza di trend stocastici**. Invece di differenziare singolarment le serie convolte (variabile dipendente e regressori, anche in numero maggiore di uno), si può procedere come segue:

1. si regredisca la variabile dipendente sulle variabili indipendenti per stimare i parametri di quella che è nota come la **relazione di lungo periodo (long run relation)**

$$y_t = \beta_0 + \beta' x_t + u_t$$

come dimostrato da Stock (uno degli autori del libro di testo), quando le variabili sono co-integrate lo stimatore OLS è super-consistente e quindi lo possiamo utilizzare per ottenere stime puntuali dei parametri, anche se gli errori standard OLS sono inconsistenti e dunque non possiamo usarli per verificare la significatività dei parametri di questa relazione;

2. si sottoponga a verifica l'ipotesi che u_t sia $I(1)$, utilizzando la distribuzione tabulata da Engle e Granger (i software econometrici, incluso Eviews e Gretl, lo fanno in automatico²) della statistica del test di Dickey-Fueller, che tiene conto sia del fatto che u_t è $I(1)$ nell'ipotesi nulla che del fatto che i parametri del modello di lungo periodo sono stimati;
3. in caso l'ipotesi che u_t sia $I(1)$ venga rigettata, si proceda con l'utilizzare l'errore come una variabile esplicativa addizionale in un cosiddetto modello delle differenze prime a correzione dell'errore (**Error Corection Model, ECM**):

$$\Delta y_t = \beta_0 + \beta' \Delta x_t - \gamma (y_{t-1} - \beta' x_{t-1}) + u_t$$

² Dal menu "Quick", selezionare "Group statistics", quindi "Descriptive statistics" e poi "Common sample". Quindi inserire il nome delle serie nella finestra di dialogo che sarà apparsa e, dal menu "View", selezionare "Cointegration tests" e poi "Single-equation cointegration tests".

È il **Teorema di Rappresentazione di Granger** che stabilisce che, quando le variabili siano cointegrate, è possibile rappresentarne le relazioni di breve e lungo periodo in questo modo.

Per capire cosa rappresenti questo modello, si consideri la relazione di lungo periodo, se ne differenzino entrambe i lati e si effettuino alcune manipolazioni algebriche, come mostrato di seguito:

$$\begin{aligned}
 \Delta y_t &= \boldsymbol{\beta}' \Delta \mathbf{x}_t + \Delta u_t \\
 &= \boldsymbol{\beta}' \Delta \mathbf{x}_t + u_t - u_{t-1} \\
 &= \boldsymbol{\beta}' \Delta \mathbf{x}_t - u_{t-1} + u_t \\
 &= \boldsymbol{\beta}' \Delta \mathbf{x}_t - \underbrace{(y_{t-1} - \beta_0 - \boldsymbol{\beta}' \mathbf{x}_{t-1})}_{u_{t-1}} + u_t \\
 &= \beta_0 + \boldsymbol{\beta}' \Delta \mathbf{x}_t - \underbrace{(y_{t-1} - \boldsymbol{\beta}' \mathbf{x}_{t-1})}_{u_{t-1}} + u_t
 \end{aligned}$$

Questo è l'ECM che avremmo se l'aggiustamento verso la relazione di lungo periodo fosse sempre immediata. Per consentire che invece l'aggiustamento sia più lento, che poi è il caso più interessante dal punto di vista dell'analisi economica, aggiungiamo il parametro γ 'al posto giusto', come mostrato di seguito

$$\Delta y_t = \beta_0 + \boldsymbol{\beta}' \Delta \mathbf{x}_t - \gamma (y_{t-1} - \boldsymbol{\beta}' \mathbf{x}_{t-1}) + u_t \quad |\gamma| \leq 1$$

Quando $\gamma=1$, l'aggiustamento verso la relazione di lungo periodo tra y_{t-1} e \mathbf{x}_{t-1} avviene immediatamente, ovvero tutta e subito nel periodo successivo a quello in cui si verifica la deviazione, mentre quando $|\gamma| < 1$ questo aggiustamento si verifica con un certo ritardo, che è tanto maggiore quanto minore è il valore assunto dal parametro.

È importante sottolineare che i parametri dell'ECM possono essere stimati in maniera consistente col metodo OLS.

Questo approccio dunque non richiede la differenziazione delle variabili e questo rappresenta un vantaggio notevole. Infatti, il livello delle variabili potrebbe contenere informazioni preziose ai fini della loro previsione e della comprensione dei fenomeni oggetto di indagine. In particolare, il trend stocastico che determina la non stazionarietà delle variabili potrebbe essere in comune. In questo caso, trascurare il trend stocastico equivarrebbe a trascurare un legame importante e dunque a buttar via informazione.

Infatti, ciò che abbiamo chiamato regressione di lungo periodo rappresenta la relazione di lungo periodo che è in vigore tra le variabili. Dunque rappresenta l'equilibrio cui ritornano i valori relativi nel lungo periodo una volta smaltito l'effetto delle fluttuazioni di breve periodo. Per tale motivo questi modelli sono molto usati in macroeconomia e in finanza. In particolare, si è di recente molto

diffuso il loro utilizzo tra gli operatori finanziari che ne fanno uso nelle loro strategie di “pair trading”.

Si noti che l’ECM appena visto può essere riscritto risolvendo l’equazione che lo definisce per la variazione di una delle variabili in x_t invece che per la variazione di y_{t-1} . Ad esempio nel caso più semplice in cui ci sia solo una variabile in x_t , possiamo scrivere

$$\Delta x_t = \beta_0 + \beta \Delta y_t - \gamma(x_{t-1} - \beta y_{t-1}) + u_t \quad |\gamma| \leq 1$$

Ciò significa che l’ECM può essere interpretato come il modello strutturale che implica un modello VAR in cui le variabili endogenee siano rappresentate da x_t e y_t (in livello, non alle differenze prime), le quali a propria volta possano essere sia I(0) che I(1).

Appendice: Proprietà di lungo periodo dei processi AR(1)

Si consideri il seguente modello AR(1):

$$\begin{aligned}
 y_t &= \alpha + \beta y_{t-1} + u_t \\
 &= \alpha + \beta(\alpha + \beta y_{t-2} + u_{t-1}) + u_t \\
 &= \alpha + \alpha\beta + \beta^2 y_{t-2} + \beta u_{t-1} + u_t \\
 &= \alpha + \alpha\beta + \beta^2 y_{t-2} + u_t + \beta u_{t-1} \\
 &= \alpha + \alpha\beta + \beta^2(\alpha + \beta y_{t-3} + u_{t-2}) + u_t + \beta u_{t-1} \\
 &= \alpha + \alpha\beta + \alpha\beta^2 + \beta^3 y_{t-3} + \beta^2 u_{t-2} + u_t + \beta u_{t-1} \\
 &= \alpha + \alpha\beta + \alpha\beta^2 + \beta^3 y_{t-3} + u_t + \beta u_{t-1} + \beta^2 u_{t-2} \\
 &= \alpha + \alpha\beta + \alpha\beta^2 + \beta^3 y_{t-3} + u_t + \beta u_{t-1} + \beta^2 u_{t-2} \\
 &\quad \dots \\
 &= \sum_{i=0}^n \alpha\beta^i + \beta^n y_{t-n} + \sum_{i=0}^n \beta^i u_{t-i} \\
 &= \begin{cases} \alpha \sum_{i=0}^n \beta^i + \underbrace{\beta^n y_{t-n}}_{\xrightarrow{n \rightarrow \infty} \begin{cases} 0 \text{ se } |\beta| < 1 \\ y_{t-\infty} \text{ se } \beta = 1 \end{cases}} + \sum_{i=0}^n \beta^i u_{t-i} & \text{se } |\beta| < 1 \\ y_{t-n} + \sum_{i=0}^n u_{t-i} & \text{se } \beta = 1 \text{ e } \alpha = 0 \\ y_{t-n} + n(\alpha\beta) + \sum_{i=0}^n u_{t-i} & \text{se } \beta = 1 \text{ e } \alpha \neq 0 \end{cases} \\
 &\xrightarrow{n \rightarrow \infty} \begin{cases} \frac{\alpha}{1-\beta} + \sum_{i=0}^{\infty} \beta^i u_{t-i} & \text{se } |\beta| < 1 \\ y_{t-\infty} + \sum_{i=0}^{\infty} u_{t-i} & \text{se } \beta = 1 \text{ e } \alpha = 0 \\ \infty & \text{se } \beta = 1 \text{ e } \alpha \neq 0 \end{cases}
 \end{aligned}$$

Ora, poniamo $t = t_0 + \infty$ e dunque trattiamo y_t come il valore assunto dal processo nel lungo periodo rispetto ad un momento iniziale t_0 . Abbiamo, quindi,

$$y_{t_0+\infty} = \begin{cases} \frac{\alpha}{1-\beta} + \sum_{i=0}^{\infty} \beta^i u_{t_0+\infty-i} & \text{se } |\beta| < 1 \\ y_{t_0} + \sum_{i=0}^{\infty} u_{t_0+\infty-i} & \text{se } \beta = 1 \text{ e } \alpha = 0 \\ \infty & \text{se } \beta = 1 \text{ e } \alpha \neq 0 \end{cases}$$

$$= \begin{cases} \frac{\alpha}{1-\beta} + \sum_{i=0}^{\infty} \beta^i u_{t_0+i} & \text{se } |\beta| < 1 \\ y_{t_0} + \sum_{i=0}^{\infty} u_{t_0+i} & \text{se } \beta = 1 \text{ e } \alpha = 0 \\ \infty & \text{se } \beta = 1 \text{ e } \alpha \neq 0 \end{cases}$$

Inoltre, abbiamo dunque che

$$E(y_{t_0+\infty}) = \begin{cases} \frac{\alpha}{1-\beta} & \text{se } |\beta| < 1 \\ y_{t_0} & \text{se } \beta = 1 \text{ e } \alpha = 0 \\ \infty & \text{se } \beta = 1 \text{ e } \alpha \neq 0 \end{cases}$$