

# MODELLI ECONOMETRICI PER SERIE STORICHE

Maria Elena Bontempi (e.bontempi@economia.unife.it)

Roberto Golinelli (golinell@spbo.unibo.it)

this version 9 gennaio 2007

1.	Alcuni concetti di base.....	1
2.	Come trasferire una banca dati di serie storiche da Excel a Stata.....	3
3.	La modellazione Box-Jenkins univariata: il modello autoregressivo (AR).....	5
3.1.	(A <sub>1</sub> ) Analisi esplorativa: grafici e correlogramma.....	6
3.2.	(A <sub>2</sub> ) I momenti secondi: il correlogramma.....	10
3.3.	(A <sub>3</sub> ) La verifica <i>statistica</i> della stazionarietà con Dickey-Fuller.....	14
3.3.1.	Un approfondimento analitico del test ADF.....	17
3.4.	(B) Scelta dell'ordine del modello AR(p) di partenza.....	19
3.5.	(C) Analisi dei test diagnostici del modello AR(p).....	20
3.6.	(D) Riduzione dell'ordine del modello AR(p).....	23
3.7.	Le previsione univariata e non condizionale con il modello AR.....	26
3.8.	Riepilogo dei temi trattati e un esempio finale.....	27
4.	Elementi di analisi multivariata: il modello uniequazionale ARDL(p, q).....	31
5.	Rischi di regressioni spurie e rimedi forniti dalla specificazione dinamica.....	43

## 1. Alcuni concetti di base

Una serie storica è una successione di dati ordinati nel tempo:  $Y_t$  con  $t=1,2, \dots, T$ . A livello intuitivo, dal punto di vista statistico, la serie storica è un'unica realizzazione di un processo stocastico, ossia di una serie infinita di variabili casuali caratterizzata da una particolare funzione di densità di probabilità congiunta,  $f(\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots)$ .

La fonte di informazione aggiuntiva di una serie storica, rispetto a quella nei dati cross-section, è che, a priori, le realizzazioni di  $Y_t$  non sono tra loro indipendenti. La teoria economica razionalizza tale dipendenza temporale delle realizzazioni empiriche: i comportamenti degli individui, i meccanismi di trasmissione di prezzi e delle politiche tendono a non esaurirsi in un solo periodo, ma manifestano i loro effetti in un arco temporale più lungo. Quindi, ad esempio, il tasso di inflazione di questo mese tende ad essere correlato con le realizzazioni del mese scorso e del prossimo mese. A seconda dell'entità del legame tra il dato rilevato in  $t$  e quello rilevato in  $t-1$  o in epoche più distanti si definisce la memoria del processo stocastico o la sua persistenza.

A seconda della frequenza di campionamento, le serie storiche possono essere: annuali, trimestrali, mensili, settimanali, giornaliere, ecc... La frequenza di campionamento deve essere tale da rilevare i movimenti della variabile misurata "abbastanza spesso" per informarci del fenomeno che ci interessa studiare. Se ad esempio, il fenomeno di interesse è la dinamica dello stock di capitale, una frequenza di campionamento giornaliera è inutile e non informativa, in quanto i movimenti dello stock di capitale che risultano informativi per i nostri modelli del capitale si sviluppano nel corso degli anni, mentre le fluttuazioni giornaliere sono impercettibili (o peggio ancora erratiche) per ciò che riguarda le dinamiche di interesse. Al contrario, i fenomeni di tipo finanziario, caratterizzati da intense contrattazioni e movimenti giornalieri, sono nascosti (cancellati) in modo irrimediabile se le quotazioni azionarie sono rilevate con frequenza, ad esempio, annuale: la dinamica di interesse ha una frequenza di realizzazione ben più alta dell'anno solare.

La frequenza di campionamento è un concetto cui si associa anche il concetto di intervallo di tempo preso in considerazione (il periodo campionario, o *time span*). L'informazione contenuta in un campione di dati non può essere misurata dal solo numero di osservazioni incluse nella serie storica della variabile di interesse (di solito lo si indica con T). Se, ad esempio, si vuole stimare la dinamica strutturale della funzione di produzione di un paese, T=100 mesi (poco più di 8 anni) sono poco informativi per capire un fenomeno che si sviluppa su archi temporali più ampi, mentre lo sarebbe T=100 trimestri (25 anni). Al contrario, un campione di 100 mesi può essere sufficiente per analizzare un fenomeno di breve periodo quale la dinamica dei tassi di interesse o delle quotazioni azionarie.

In breve, frequenza di campionamento e periodo campionario a nostra disposizione sono informativi per il modello empirico a seconda del fenomeno teorico (di breve o di lungo periodo) che si intende studiare. Spesso la disponibilità di dati è un vincolo al nostro set informativo e, in questi casi, il ricercatore deve fare “di necessità virtù”, ma in ogni caso è sempre bene tenere presente tale problematica, almeno nelle fasi di interpretazione e commento dei risultati.

La tabella contenuta nella cartella “inclasse” del foglio Excel “datiTS.xls” riporta i dati della serie storica trimestrale  $P_t$  per il periodo campionario 1970q1 – 2001q4 (livello dell'indice dei prezzi al consumo in USA, base 1995=1).

Con l'ausilio del foglio Excel, definiamo il concetto di ritardo (lag) di primo ordine,  $P_{t-1}$ ; potremmo anche definire il ritardo di secondo ordine  $P_{t-2}$  e, in generale, di ordine k  $P_{t-k}$ .

A parità di periodo considerato, ad esempio nel terzo trimestre del 1973,  $P_t$  e  $P_{t-1}$  assumono valori diversi; pertanto, così come può essere ragionevole studiare il legame fra Y e X mediante il modello lineare  $Y_i = a + b X_i + \varepsilon_i$  (in cui l'i-esima realizzazione di Y è ipotizzata dipendere dalla i-esima osservazione di X), nel caso di serie storiche un modello proponibile è il modello autoregressivo (del primo ordine), in simboli AR(1):  $Y_t = a + b Y_{t-1} + \varepsilon_t$  (in cui la realizzazione al tempo t di  $Y_t$  dipende dalla realizzazione  $Y_{t-1}$ , cioè da ciò che è accaduto nel periodo precedente). Ritorneremo su questo punto più avanti.

La variazione di P da t a t-1 è detta differenza prima di P:  $\Delta_1 P_t = \Delta P_t = P_t - P_{t-1}$  (in generale la differenza di ordine k è:  $\Delta_k P_t = P_t - P_{t-k}$ ). L'andamento temporale della variabile prezzo  $P_t$  è spesso misurato dal suo tasso di variazione: il tasso di inflazione ottenuto dal rapporto incrementale:  $\Delta P_t / P_{t-1}$  (la variazione dei prezzi in t rispetto al loro livello in t-1). Nel caso di dati con frequenza di campionamento trimestrale il tasso di variazione dei prezzi in t rispetto ai prezzi in t-1 è il tasso di inflazione trimestrale ( $\text{infl}_t$ ), nel 1973q3 era circa il 2.1%. Se si desidera un indicatore annuale di inflazione (ad esempio per confrontarlo con i livelli del tasso di interesse),  $\text{infl}_t$  può essere annualizzato moltiplicandolo per quattro ( $\text{inflA}_t = 4 \times \text{infl}_t$ ): nel 1973q3 era circa l'8.3%. Un'altra misura di inflazione annuale è quella annua tendenziale,  $\text{inflAT}_t$ , ottenuta calcolando il tasso di variazione di  $P_t$  rispetto allo stesso trimestre dell'anno prima ( $P_{t-4}$ ): nel 1973q3 era circa il 6.5%.

In pratica,  $\text{inflA}_t$  e  $\text{inflAT}_t$  sono due misure alternative (e con pari dignità) di inflazione annua, mentre la teoria economica parla semplicemente di “tasso di inflazione”: alla stessa variabile teorica spesso si contrappongono diverse misure di inflazione. Se ci chiedessero qual è stato il tasso di inflazione su base annua nel terzo trimestre del 1973, risponderemmo l'8.3% o il 6.5%? Evidentemente, la scelta della specifica misura di inflazione non è irrilevante per la risposta che si fornisce!

Nelle serie storiche spesso si utilizza la trasformazione logaritmica dei livelli delle variabili di interesse. Le proprietà interessanti di tale trasformazione sono molte:

- (a) la trasformazione logaritmica è monotona;
- (b) stabilizza la varianza della serie di partenza;
- (c) la differenza prima del logaritmo di una serie ne approssima il tasso di variazione (per piccoli tassi di variazione):

$$\Delta \log(P_t) = \log(P_t) - \log(P_{t-1}) = \log\left(\frac{P_t}{P_{t-1}}\right) = \log\left(\frac{P_t + P_{t-1} - P_{t-1}}{P_{t-1}}\right) = \log\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) \cong \frac{P_t - P_{t-1}}{P_{t-1}} = \text{infl}_t$$

Esercitatevi cercando di riprodurre concetti e calcoli appena introdotti completando le informazioni contenute nella cartella “acasa” del file “datiTS.xls”.

## 2. Come trasferire una banca dati di serie storiche da Excel a Stata

Il modo migliore per inserire in Stata serie storiche trimestrali è partire da una banca dati Excel in cui compare la variabile "anno" che indica l'anno di riferimento e la variabile "trimestre" che indica il trimestre di riferimento.

Ad esempio, supponiamo di volere trasferire in Stata le serie storiche per le variabili P (livello dell'indice dei prezzi al consumo in USA, base 1983=1) e UR (tasso di disoccupazione in USA) contenute nella cartella “acasa” del file “datiTS.xls”. Si evidenziano e si copiano le prime 4 colonne della cartella "acasa" e poi le si incolla nell'editor di Stata, ottenendo:

```
list
      anno  trimes~e      ur      p
1.    1959      1    .0583333    28.993
2.    1959      2      .051    29.043
3.    1959      3    .0526667    29.193
4.    1959      4      .056    29.37
5.    1960      1    .0513333    29.397
6.    1960      2    .0523333    29.573
7.    1960      3    .0553333    29.59
8.    1960      4    .0626667    29.78
9.    1961      1      .068    29.84
10.   1961      2      .07    29.83
ecc..
```

A questo punto, bisogna generare una variabile `tempo` che informi Stata della progressione delle osservazioni nel tempo

```
g tempo = yq(anno, trimestre)
```

NOTA: y=year, q=quarter, m=month, ...

```
list
      anno  trimes~e      ur      p      tempo
1.    1959      1    .0583333    28.993      -4
2.    1959      2      .051    29.043      -3
3.    1959      3    .0526667    29.193      -2
4.    1959      4      .056    29.37       -1
5.    1960      1    .0513333    29.397       0
6.    1960      2    .0523333    29.573       1
7.    1960      3    .0553333    29.59       2
8.    1960      4    .0626667    29.78       3
9.    1961      1      .068    29.84       4
10.   1961      2      .07    29.83       5
ecc...
```

La variabile `tempo` è una *singola* variabile che numera progressivamente le osservazioni delle due serie storiche `ur` e `p`, imponendo *convenzionalmente* (e *arbitrariamente*) il valore "0" in corrispondenza della prima osservazione del 1960.

L'ultimo passaggio è quello di informare Stata che la banca dati in uso è relativa al periodo temporale individuato dalla variabile `tempo` e che quest'ultima deve essere visualizzata in un formato più comprensibile:

```

tsset tempo, quarterly
    time variable:  tempo, 1959q1 to 2000q4

list anno trimestre ur p tempo

      anno  trimestre      ur      p      tempo
1.   1959           1  .0583333  28.993  1959q1
2.   1959           2   .051    29.043  1959q2
3.   1959           3  .0526667  29.193  1959q3
4.   1959           4   .056    29.37   1959q4
5.   1960           1  .0513333  29.397  1960q1
6.   1960           2  .0523333  29.573  1960q2
7.   1960           3  .0553333  29.59   1960q3
8.   1960           4  .0626667  29.78   1960q4
9.   1961           1   .068    29.84   1961q1
10.  1961           2   .07     29.83   1961q2
ecc...

```

Salvando le informazioni ottenute sinora nella banca dati Stata "USquarter.dta", l'indicazione del `tsset` viene memorizzata anche per sessioni successive.

Provate a riprodurre le operazioni sopra nel caso della banca dati annuale, memorizzata nella cartella "annuale", e nel caso della banca dati mensile (in "mensile") dello stesso file datiTS.xls

In queste note capiamo come trattare con Stata serie storiche di dati mensili-trimestrali- annuali; per altri casi (ad esempio, dati giornalieri) è bene consultare il manuale di Stata.

Dopo avere memorizzato in Stata tutte le informazioni precedenti, i concetti ottenuti in Excel mediante elaborazioni del contenuto di ogni cella diventano facilmente ottenibili in Stata mediante l'uso del comando `g(enerate)` [ ad esempio, il calcolo della trasformazione logaritmica di  $P$  è: `g lp=log(p)` ] e di particolari **operatori time series**:

**l.** = operatore ritardo di ordine 1: `l.nome-della-serie` (ad esempio `l.Y` equivale a  $Y_{t-1}$ )  
in generale **lk.** misura il ritardo (lag) di ordine  $k$ .

**d.** = operatore differenza prima: `d.nome-della-serie` (ad esempio `d.Y` equivale a  $\Delta Y_t$ )

Se usati insieme, l'operatore differenza prima precede l'operatore ritardo.

Ad esempio:

```

list tempo p l.p lp l.lp d.lp

      tempo      p      L.p      lp      L.lp      D.lp
1.  1959q1  28.993      .    3.367054      .      .
2.  1959q2  29.043  28.993  3.368778  3.367054  .0017231
3.  1959q3  29.193  29.043  3.373929  3.368778  .0051515
4.  1959q4   29.37  29.193  3.379974  3.373929  .0060449
5.  1960q1  29.397  29.37   3.380893  3.379974  .0009186

```

6.	1960q2	29.573	29.397	3.386862	3.380893	.0059693
7.	1960q3	29.59	29.573	3.387436	3.386862	.0005746
8.	1960q4	29.78	29.59	3.393837	3.387436	.0064006
9.	1961q1	29.84	29.78	3.39585	3.393837	.0020127
10.	1961q2	29.83	29.84	3.395514	3.39585	-.0003352
11.	1961q3	29.947	29.83	3.399429	3.395514	.0039146
12.	1961q4	29.99	29.947	3.400864	3.399429	.0014348

ecc...

(confronta i risultati sopra riportati con quelli ottenuti dallo svolgimento dell'esercizio sulla cartella "acasa" del file datiTS.xls)

### 3. La modellazione Box-Jenkins univariata: il modello autoregressivo (AR)

L'analisi statistica delle serie storiche viene spesso definita "metodologia Box-Jenkins" dal nome dei due studiosi che, con la sistemazione metodologica di un vasto insieme di analisi statistiche nel loro volume del 1970, ne hanno favorito lo sviluppo e la diffusione.

Visto lo scopo introduttivo del corso, ci concentreremo sulla modellazione delle serie storiche stazionarie mediante processi autoregressivi (AR). Più in generale l'approccio Box-Jenkins è definito ARMA dalle iniziali di AR, *autoregressive* e MA, *moving average* e la sua utilità empirica è legata ad almeno tre fattori.

L'analisi statistica delle serie storiche può essere, anzitutto, vista come una metodologia capace di modellazione senza teoria economica. Quando il ricercatore è interessato ad ottenere una previsione di breve periodo di una variabile economica non facilmente modellabile dal punto di vista teorico o caratterizzata da penuria di informazioni sulle sue determinanti, l'approccio Box-Jenkins rappresenta sicuramente la più comoda (e spesso la più affidabile) via da intraprendere.

Un secondo punto, strettamente collegato al precedente, è relativo all'utilizzazione delle previsioni "non condizionali" (non dipendenti cioè da ipotesi esterne di scenario e/o di politica) fornite dalla modellistica ARMA come *benchmark* per le previsioni dei modelli che incorporano teoria economica (modelli strutturali). Se la teoria economica impiegata dai modelli strutturali è vera, o più modestamente utile, questo più ampio insieme informativo dovrebbe consentire previsioni più accurate rispetto all'impiego della semplice modellistica ARMA. Se ciò non si verifica (i modelli ARMA prevedono meglio) evidentemente siamo in presenza di modelli econometrici strutturali caratterizzati da problemi di specificazione.

Un ultimo fattore d'interesse dell'approccio Box-Jenkins è l'enfasi che viene dedicata allo studio della struttura temporale dei dati e, quindi, la sua filosofia di modellazione. Per tale motivo, i modelli statistici di serie storiche sono interpretabili come analisi preliminare dei dati alla modellazione strutturale: per dirla con Box e Jenkins, in questo ambito analitico il ricercatore passa "dall'amore per uno specifico modello teorico all'amore per i dati". Si pensi ad esempio all'attenzione dedicata al tema della stazionarietà delle variabili, visto come condizione della loro modellazione. Importanti sviluppi, in ambito di modellazione strutturale, discendono da tale tema, sviluppi collocati nell'analisi di integrazione e di cointegrazione. Questi ultimi argomenti, di carattere più avanzato, sono solitamente affrontati dai corsi di econometria a livello di laurea specialistica o di dottorato di ricerca.

La modellazione AR della variabile di interesse è sintetizzata dalle seguenti fasi:

(A) Analisi univariata delle serie storiche. Costituiscono tecniche esplorative per serie storiche: l'analisi grafica; lo studio della "memoria" del processo stocastico tramite strumenti quali il

correlogramma; l'analisi della stazionarietà della serie storica (tramite il test di Dickey-Fuller, 1979) e la scelta della trasformazione che la rende stazionaria. Nel caso in cui si scelga di renderla stazionaria per differenze successive, la differenziazione necessaria affinché il test di Dickey-Fuller rifiuti l'ipotesi nulla di non stazionarietà fornisce il punto di partenza per l'identificazione del modello generale AR(p).

(B) Scelta dell'ordine (p) del modello AR. Il modello AR(p) è:  $y_t = \alpha + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t$ . Analogamente a quanto vedremo nel caso dell'effettuazione del test di Dickey-Fuller, una regola pratica per la scelta di p è quella di iniziare da un ordine collegato alla periodicità dei dati.

(C) Analisi dei test diagnostici del modello AR(p). Si salvano i residui e si verifica se sono white noise (normalmente e indipendentemente distribuiti nel tempo con media nulla e varianza costante), oppure se sono autocorrelati. Alcune tecniche adottate sono quelle di cui al punto (A). Nel caso in cui emerga che i residui sono autocorrelati, occorre ri-specificare e ri-stimare il modello AR(p) aumentandone l'ordine, ossia aggiungendo ulteriori ritardi. In questa sede è opportuno effettuare anche i test di eteroschedasticità-linearità-normalità.

(D) Solo dopo che i test di scorretta specificazione sui residui - passo (C) - non segnalano evidenti problematiche, si procede alla eliminazione dei regressori cui corrispondono stime dei parametri non significativamente diverse da zero ed al successivo controllo diagnostico dei residui del modello vincolato.

La fase di modellazione AR termina con un modello il cui correlogramma dei residui è "vuoto" (ossia siamo in presenza di un white noise) ed i cui parametri stimati (tranne la costante, da lasciare comunque) sono significativamente diversi da zero.

### 3.1. (A<sub>1</sub>) Analisi esplorativa: grafici e correlogramma.

La caratteristica dipendenza temporale dei dati delle serie storiche implica, da un lato, la disponibilità di tecniche esplorative *ad hoc* che permettono analisi più efficaci, dall'altro, una serie di *caveat* per ciò che riguarda la stima dei momenti primi e secondi.

La formula per la stima del valore atteso o media o momento primo,  $E(Y_t) = \mu$ , di una serie storica è:

$$\hat{E}(Y_t) = \hat{\mu} = \frac{\sum_{t=1}^T y_t}{T}.$$

I momenti secondi o le autocovarianze di ordine  $k=0, 1, 2, \dots$ ,  $COV(Y_t, Y_{t-k}) = \gamma_k$ , sono stimati con:<sup>1</sup>

$$\hat{COV}(Y_t, Y_{t-k}) = \hat{\gamma}_k = \frac{\sum_{t=k+1}^T (y_t - \hat{\mu})(y_{t-k} - \hat{\mu})}{T - k - 1},$$

da cui si costruisce la stima del coefficiente di correlazione lineare di ordine  $k$ :

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}.$$

Which is the equivalent in the time-series context of the sampling assumption i.i.d. for cross-section?

<sup>1</sup> In generale,  $k \geq 0$ ; in particolare, quando  $k=0$ , l'autocovarianza di ordine zero è la  $VAR(Y_t) = \gamma_0$ .

<sup>2</sup> Ricorda che il coefficiente di correlazione lineare è un indice (numero puro,  $-1 < \rho_k < 1$ ). Per definizione,  $\hat{\rho}_0 = 1$ .

Consider the **identical distribution**. This assumption can be analogous to stationarity. We have strict stationarity when the pdf does not change as time goes by:  $f(y_1, \dots, y_T) = f(y_{1+k}, \dots, y_{T+k})$ ,  $\forall k \geq 1$ . In general, however, this condition is too restrictive. Hence, we use the weak stationarity or stationarity in covariance assumption. It means that the moments defining the pdf (mean, variance and covariance) do not depend on time:

$$E(y_t) = E(y_{t+k}) = \mu;$$

$$VAR(y_t) = VAR(y_{t+k}) = \gamma_0;$$

$$COV(y_t, y_{t+k}) = \gamma_k \text{ or equal to a function of } k.$$

Now consider the **independence**. This assumption is replaced by the ergodicity or weak dependence assumption:  $\rho_k = Corr(y_t, y_{t-k}) \rightarrow 0$  for  $k \rightarrow \infty$ , i.e. when the present is far away from the past (the process has a finite and decreasing memory). If the series is stationary, this correlation does not depend from the starting point  $t$ ; stationary stochastic processes respecting weak dependence are called asymptotically uncorrelated.

Prima della fase di modellazione econometrica, è bene che il ricercatore si renda conto delle principali caratteristiche statistiche delle serie oggetto dell'analisi. L'analisi del grafico dell'andamento delle variabili di interesse nel tempo (*plot*) è un modo preliminare di valutare visivamente le serie storiche.

Supponendo di essere interessati allo studio empirico dei prezzi al consumo per gli USA (banca dati USquarter.dta), ispezioniamone l'andamento mediante il grafico della sua trasformata logaritmica (lp), da cui se ne evince una tendenza alla crescita di lungo periodo (trend). Calcoliamo la stima della media campionaria  $\hat{\mu}$  (vedi la formula precedente):

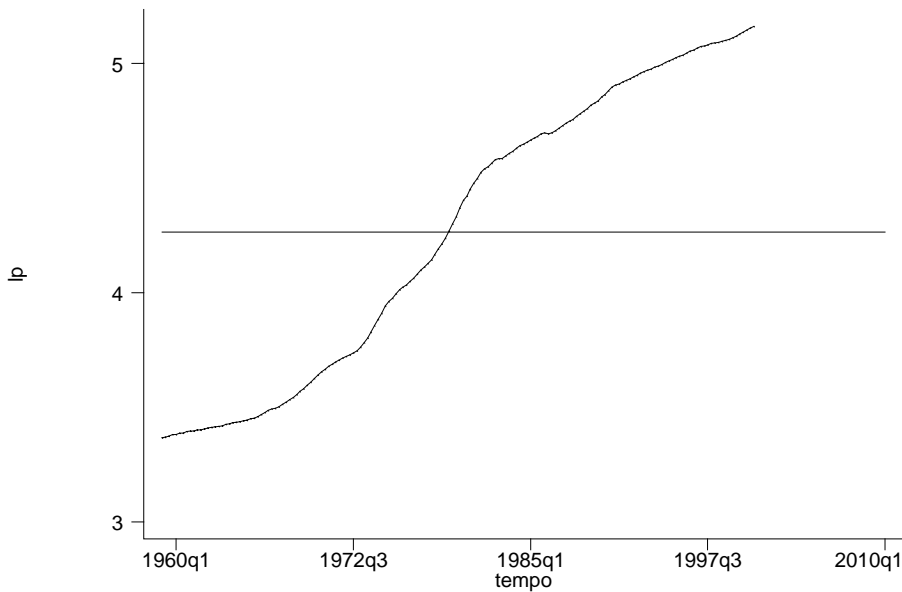
```
summ lp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lp	168	4.263305	.63211	3.367054	5.160589

La retta orizzontale nel plot indica la media campionaria di lp (4.263): evidentemente in questo caso, come in tutti i casi di serie con trend, la media non rappresenta il centro della distribuzione di lp, ma si pone arbitrariamente a metà strada fra i livelli minimi (inizi '60) e massimi (fine '90) del campione.

Chi usa Stata 8 o 9 a questo punto deve dare il comando “version 7” oppure usare “graph7” al posto di “graph”.

```
graph lp tempo, c(1) s(.) xlabel ylabel yline(4.263)
```



Quindi, pare lecito qualificare i prezzi come una variabile il cui modello statistico non è stazionario in media (cioè la media varia nel tempo). Se un processo stocastico per la variabile  $Y_t$  non è stazionario in media, le sue osservazioni sono eterogenee e, quindi, non possiamo “metterle assieme” in un’unica stima della media  $\hat{\mu}$  perché  $E(Y_t) = \mu_t$ , cioè varia anch’essa nel tempo.

Si pensi all’esempio della serie storica  $W_t$  del peso di un individuo dalla nascita fino a 60 anni ( $T=60$ ): le realizzazioni temporali di  $W_t$  non sono estratte dalla stessa distribuzione statistica perché il peso varia strutturalmente con l’età e lo sviluppo dell’individuo. Al contrario, in una cross-section di  $N=60$  individui coetanei la media campionaria può rappresentare il centro della distribuzione della variabile peso e, quindi, ha senso qualificare come “sovrappeso” gli individui il cui peso si trova significativamente al di sopra della media.

Evidentemente, per i fenomeni non stazionari in media, le stime dei momenti secondi (varianza e autocovarianze) non hanno senso perché misurano dispersione e correlazioni rispetto a dati medi non informativi (se non addirittura fuorvianti). Inoltre, si può spesso estendere la proprietà statistica della non stazionarietà anche ai momenti secondi: se da un lato è certo che la distribuzione di  $W_t$  ( $t=1,2, \dots, 60$ ) non è centrata su un’unica media, dall’altro è anche possibile che siano instabili nel tempo (non costanti, non stazionarie) anche la dispersione attorno alla sua media (varianza) e la dipendenza con le sue realizzazioni precedenti (autocovarianze).

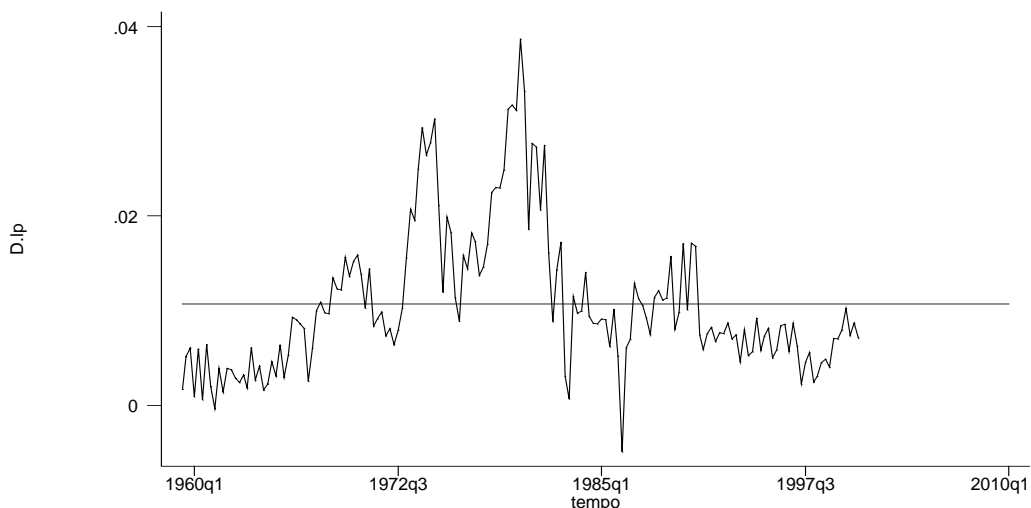
D’altro canto, seppure non stazionaria in media e/o covarianza, non è detto che la variabile di interesse non possa essere resa stazionaria mediante trasformazioni. Una tipica trasformazione che elimina dalla serie il trend è la differenza prima. Quando si studiano fenomeni economici, è sempre bene cercare di interpretare il risultato di una trasformazione statistica: in questo caso la trasformazione equivale al calcolo del tasso di inflazione trimestrale:  $d\ln P_t = \ln(P_t) - \ln(P_{t-1})$ .

```

summ d.ln
-----+-----
Variable | Obs      Mean      Std. Dev.  Min        Max
-----+-----
ln       |
D1       |      167   .0107397   .007669   -.0048857   .0386982

graph d.ln tempo, c(1) s(.) xlabel ylabel yline(0.0107)

```



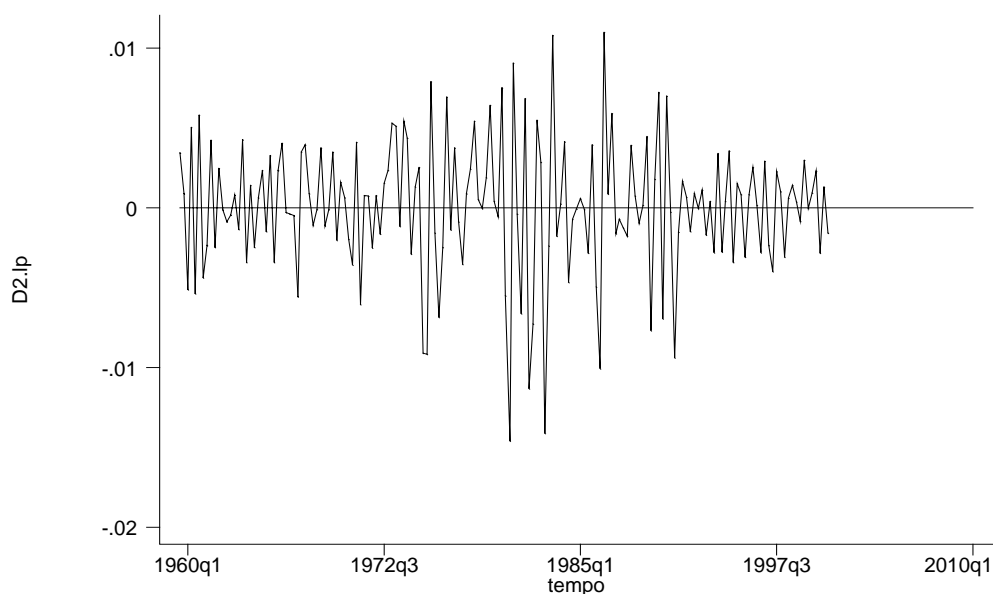
Non c'è trend ma, nonostante ciò, le realizzazioni persistono a lungo sopra/sotto la media campionaria: potrebbe essere che media e varianza di  $d1p_t$  ( $= d.lp$ ) tendano ancora a variare nel tempo. Si ricorda che, oltre ad un trend, anche una forte persistenza al di fuori della media campionaria suggerisce una possibile non stazionarietà della variabile.

Una ulteriore trasformazione in differenze prime ( $\Delta d1p_t = d1p_t - d1p_{t-1} = d2.lp$ ) elimina anche la forte persistenza di  $d1p$ :

```
summ d2.lp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lp					
D2	166	.0000324	.0041961	-.0145984	.0109844

```
graph d2.lp tempo, c(1) s(.) xlabel ylabel yline(0)
```



Il plot riporta ora un'altra variabile economica di interesse collegata con i prezzi al consumo: l'accelerazione/decelerazione del tasso di inflazione.

Spesso una sola differenza prima è sufficiente a rendere stazionaria una variabile reale (PIL, consumo, occupazione, ecc.); per rendere stazionaria una variabile nominale (prezzi, salari, stock di

moneta) potrebbe servire anche la differenza prima della differenza prima, cioè la differenza seconda, come nel nostro caso per  $\log(P_t)$ :  $\Delta^2 \log(P_t) = \Delta \log(P_t) - \Delta \log(P_{t-1}) = \Delta \text{infl}_t$ .<sup>3</sup>

L'accanimento che gli econometrici delle serie storiche dedicano alla valutazione della stazionarietà delle variabili di interesse dei loro modelli dipende dal fatto che, in assenza di tale proprietà statistica, non si possono usare i dati campionari per la stima dei momenti primi e secondi perché è come se i dati fossero eterogenei (non sommabili) in quanto prodotti nel tempo da diversi processi stocastici.

### 3.2. (A<sub>2</sub>) I momenti secondi: il correlogramma

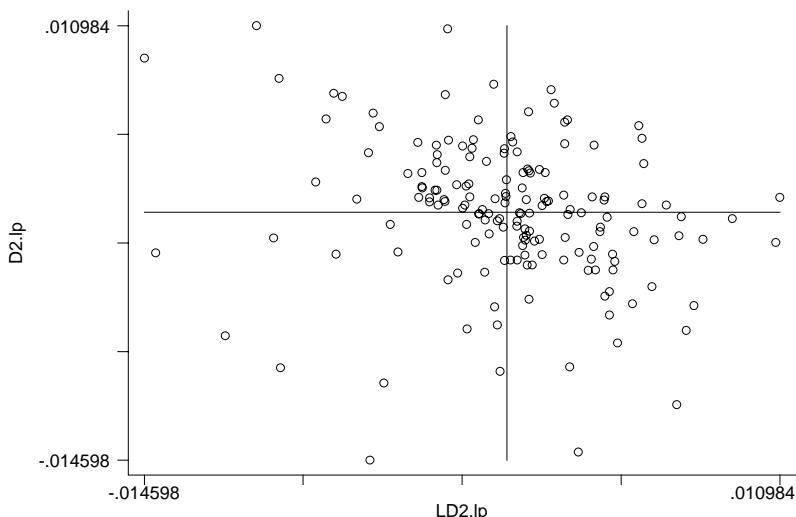
L'autocorrelazione misura la correlazione (il legame) nel tempo di una variabile economica con il proprio passato. Così come i dati cross-section per le variabili  $Y_i$  e  $X_i$  ci permettono di stimare (misurare) la forza del legame fra le variabili  $Y$  e  $X$ ,  $\text{COV}(Y, X)$ , i dati della serie storica  $Y_t$  possono essere usati (se essa è stazionaria) per misurare la forza del legame fra ciò che succede a  $Y$  in  $t$  e quello che è successo  $k$  periodi precedenti:  $\text{COV}(Y_t, Y_{t-k})$ , con  $k=1,2,\dots$ . Se la serie è stazionaria, allora la forza del legame con il proprio passato (memoria del processo stocastico univariato) tende a decrescere in valore assoluto a mano a mano che il tempo passa e, quando  $k$  è elevato, tale memoria tende a zero; ad esempio:  $|\text{COV}(Y_t, Y_{t-1})| > |\text{COV}(Y_t, Y_{t-10})|$ . Il valore assoluto implica che una forte autocorrelazione può anche essere negativa (vicina a  $-1$ ).

Alcuni scatter possono essere utili a chiarire il punto. Per facilitare l'interpretazione della correlazione, ogni grafico riporta anche le rette corrispondenti alle medie campionarie.

```
summ d2.lp ld2.lp
```

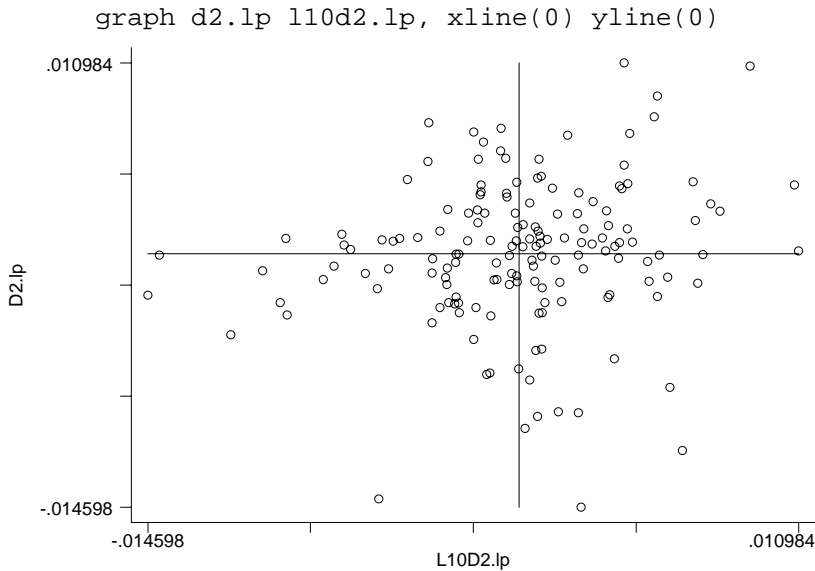
Variable	Obs	Mean	Std. Dev.	Min	Max
-----					
lp					
D2	166	.0000324	.0041961	-.0145984	.0109844
LD2	165	.0000423	.0042069	-.0145984	.0109844

```
graph d2.lp ld2.lp, xline(0) yline(0)
```

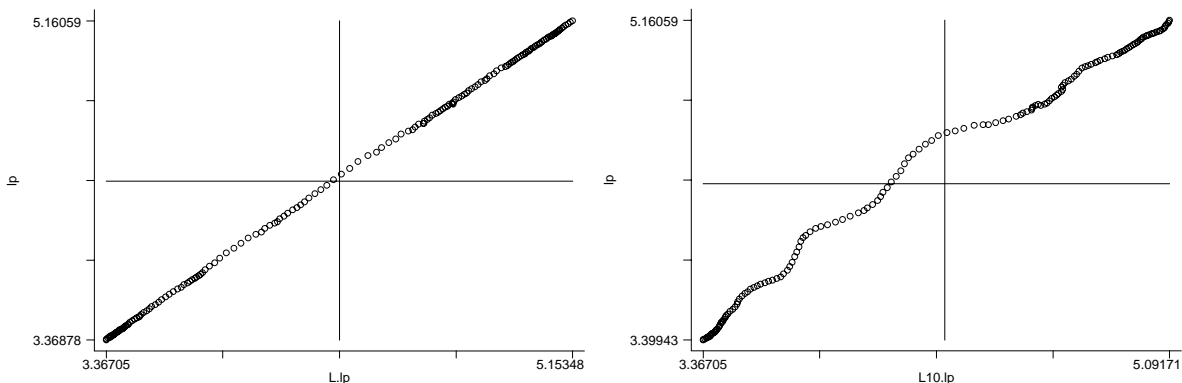


Lo scatter indica l'esistenza di un legame negativo,  $\hat{\rho}_1 < 0$ , fra  $\Delta \text{dlp}_t$  ( $d2.lp$ ) e  $\Delta \text{dlp}_{t-1}$  ( $ld2.lp$ ), mentre il legame sparisce se ci si concentra su  $\Delta \text{dlp}_t$  ( $d2.lp$ ) e  $\Delta \text{dlp}_{t-10}$  ( $l10d2.lp$ ),  $\hat{\rho}_{10} \cong 0$ :

<sup>3</sup> Nota che la differenza seconda NON è la differenza di ordine due:  $\Delta_2 = \log(P_t) - \log(P_{t-2})$ . Quest'ultima in Stata viene costruita come `g diffsecond = lp - l2.lp`.



Il legame, invece, appare sempre molto forte (la memoria non decresce al crescere di  $k$ ) se lo si visualizza per una serie non stazionaria. Ad esempio, la correlazione di  $lp_t$  con  $lp_{t-1}$  [ `graph lp l.lp, xline(4.26) yline(4.26)` ] è riportata qui sotto a sinistra, e quella di  $lp_t$  con  $lp_{t-10}$  [ `graph lp l10.lp, xline(4.26) yline(4.26)` ] è riportata qui sotto a destra.



Data una variabile da ispezionare preliminarmente, un modo “automatico” per quantificarne e presentarne sinteticamente la dispersione dei punti in tutti gli scatter fino ad un certo ordine massimo ( $k_{\max}$ ) è il calcolo del correlogramma della serie storica. Una "regola del pollice" è fissare  $k_{\max} \approx T/4$ , oppure di ispezionare l'autocorrelazione fino ad un ritardo massimo di circa 4-5 anni.

Il correlogramma è un grafico che riporta lungo l'asse verticale le stime  $\hat{\rho}_k$  e lungo l'asse orizzontale i corrispondenti ritardi di ordine  $k$ , fino ad un  $k_{\max}$  in cui l'autocorrelazione è circa zero se la serie è stazionaria.

Come esempio di correlogramma di una serie stazionaria (vedi esiti di DF), studiamo il correlogramma della differenza seconda del log dei prezzi (accelerazione/decelerazione del tasso di inflazione). Stata produce due tipi di correlogramma. Il primo è quello più informativo, ma meno rifinito graficamente:

```
corrgram d2.lp , lags(20)
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	-0.2428	-0.2430	9.9603	0.0016	-			-		
2	-0.2691	-0.3495	22.273	0.0000	--			--		
3	0.3174	0.1775	39.514	0.0000		--			--	
4	-0.0643	-0.0222	40.226	0.0000						-
5	-0.1340	-0.0155	43.335	0.0000	-					
6	0.1386	0.0229	46.682	0.0000		-				
7	0.0825	0.1237	47.875	0.0000						
8	-0.2787	-0.2044	61.581	0.0000	--				--	
9	-0.0115	-0.1381	61.605	0.0000						-
10	0.1521	-0.0493	65.74	0.0000		-				
11	-0.1546	-0.0604	70.04	0.0000	-					
12	-0.1304	-0.1753	73.12	0.0000	-					-
13	0.1744	0.0035	78.665	0.0000		-				
14	-0.0606	-0.0307	79.34	0.0000						
15	-0.0828	0.0490	80.607	0.0000						
16	0.0232	-0.1561	80.707	0.0000						-
17	0.0272	-0.0350	80.845	0.0000						
18	0.0545	0.1089	81.405	0.0000						
19	-0.1151	-0.1332	83.918	0.0000						-
20	0.0520	-0.0818	84.434	0.0000						

Con il comando `corrgram` si riportano nella prima colonna i ritardi  $k$  (LAG) in ordine crescente (con  $k_{\max}=20$  trimestri o 5 anni), nella seconda colonna le corrispondenti stime dei coefficienti di autocorrelazione di ordine  $k$  (AC), nella terza colonna le stime dei coefficienti di autocorrelazione parziale (PAC), nella quarta colonna le statistiche test  $Q$  e, nella quinta colonna, i corrispondenti valori di probabilità. Chiudono gli andamenti grafici (molto stilizzati) rispetto a  $k$  di AC e PAC.

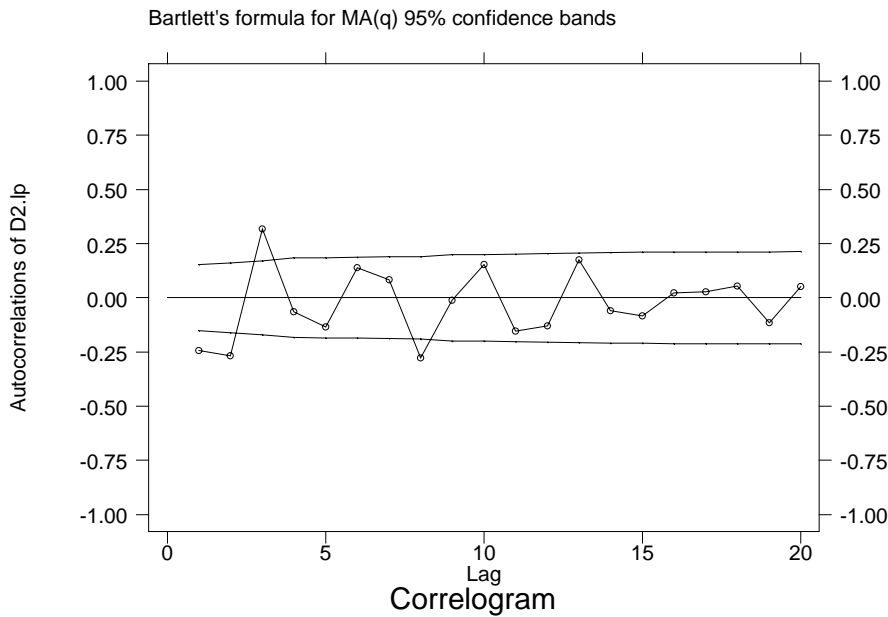
Le stime dei coefficienti di autocorrelazione parziale (PAC) vengono spiegati in seguito, quando si tratterà di scegliere l'ordine dei ritardi del modello autoregressivo (AR).

Un'importante informazione statistica per la previsione e l'analisi univariata della variabile di interesse è scoprire (inferire) dalla struttura del suo correlogramma se nel passato delle sue realizzazioni si trova sufficiente informazione per prevederne il futuro. Se, ad esempio, la serie  $Y_t$  è significativamente correlata con alcune  $Y_{t-k}$  (il suo passato) allora potremo utilizzare queste regolarità statistiche per prevedere le realizzazioni fuori dal campione  $Y_{t+h}$  (lungo l'orizzonte previsivo  $h$ ). Se, invece, tali correlazioni sono tutte non significative, allora la variabile  $Y_t$  non è prevedibile alla luce del suo passato e viene definita un processo white noise.

Il test  $Q$  (dovuto a Box-Pierce) è cumulativo e verifica l'ipotesi nulla che tutti i coefficienti di autocorrelazione fino al corrispondente LAG siano zero e, quindi, sotto  $H_0$ , la variabile  $Y_t$  è white noise. Nel nostro caso, ad esempio, la variazione dell'inflazione presenta, in corrispondenza di  $k = 20$ , una  $Q$  pari a 84.4 e il suo  $p$ -value  $\approx 0$  suggerisce il rifiuto di  $H_0$ : alcuni dei primi 20 coefficienti  $\rho_k$  sono significativamente diversi da zero e, quindi, la variabile può essere prevista a partire da un modello statistico univariato.

Il secondo tipo di rappresentazione del correlogramma è squisitamente “grafico”:

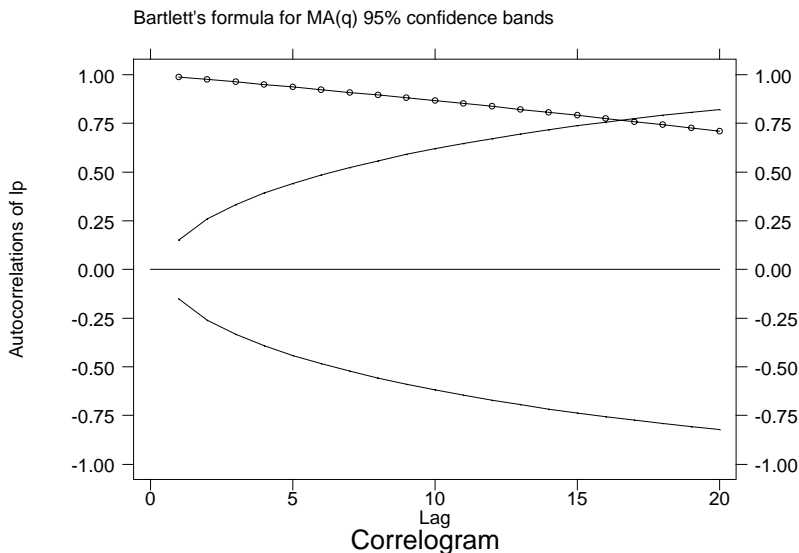
```
ac d2.lp , lags(20)
```



Nella figura si riportano le stime dei coefficienti di autocorrelazione e l'intervallo di confidenza di  $\rho_k = 0$  (per  $k=1,2,\dots,20$ ). Dato che quattro  $\hat{\rho}_k$  "escono" dall'intervallo, i corrispondenti  $\rho_k$  sono significativamente diversi da zero ( $k = 1, 2, 3, 8$ ) e, in questo modo, giustificano il rifiuto di  $H_0$  da parte del test Q (la differenza prima dell'inflazione non è white noise).

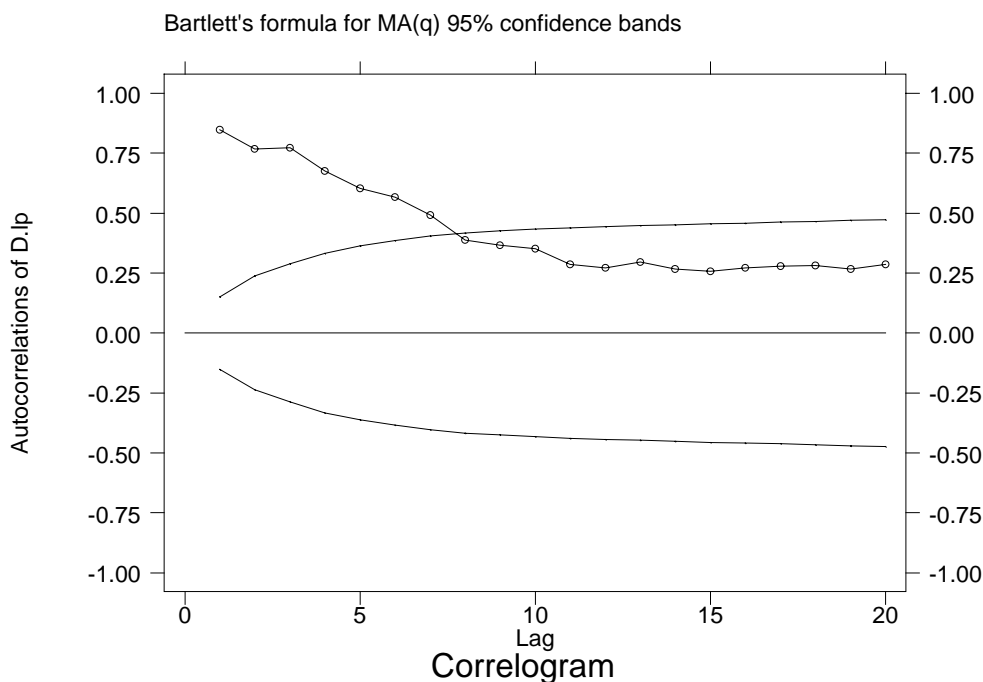
In precedenza si è notato graficamente (scatter) che, quando un processo non è stazionario, le sue autocorrelazioni sono sempre molto forti (decregono molto, molto lentamente con  $k$ ). Il correlogramma di  $lp$  ci fa apprezzare visivamente questo fatto.

`ac lp , lags(20)`



Adesso tocca a voi: come vi aspettate sia il correlogramma del tasso di inflazione? Risposta:

`ac d.lp , lags(20)`



Scende adagio o scende in fretta? Il test DF è più preciso nella valutazione della stazionarietà ...

### 3.3. (A<sub>3</sub>) La verifica statistica della stazionarietà con Dickey-Fuller

Il test di Dickey-Fuller (DF) è una prova di ipotesi che permette di discriminare fra un processo non stazionario sotto l'ipotesi nulla e un processo stazionario sotto l'ipotesi alternativa. Il test DF si articola nei seguenti passi:

- (a) scelta del modello di riferimento per il test (con trend o senza trend) sulla base dell'andamento temporale della variabile da ispezionare. Supponiamo di voler effettuare il test DF della variabile  $\log(P_t)$ ; dal plot riportato sopra emerge che la serie ha un trend e, quindi, si sceglie un modello con trend;
- (b) scelta dell'ordine iniziale di augmentation del test da effettuarsi sulla base della periodicità dei dati; i dati trimestrali suggeriscono di iniziare con una augmentation pari a 5 (4+1, cioè la periodicità + 1).

NOTA: i due test ADF che seguono sono in STATA 9, con correzione dei p-values di MacKinnon da noi suggerita (confronta con le note vecchie).

```
dfuller lp, lags(5) trend regress
```

Augmented Dickey-Fuller test for unit root                      Number of obs    =            162

	Test Statistic	----- 1% Critical Value	----- 5% Critical Value	----- 10% Critical Value
Z(t)	-1.835	-4.019	-3.442	-3.142

MacKinnon approximate p-value for Z(t) = 0.6878

D.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp					
L1.	-.0056673	.0030888	-1.83	0.068	-.0117692    .0004347



		----- Interpolated Dickey-Fuller -----			
Test		1% Critical	5% Critical	10% Critical	
Statistic		Value	Value	Value	
Z(t)		-2.613	-3.490	-2.886	-2.576
* MacKinnon approximate p-value for Z(t) = 0.0903					
D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp					
LD	-.1119155	.0428266	-2.61	0.010	-.196519 - .027312
LD2	-.1770239	.0829425	-2.13	0.034	-.3408758 - .013172
L2D2	-.238573	.0837649	-2.85	0.005	-.4040495 - .0730965
L3D2	.2214303	.0852856	2.60	0.010	.0529496 .3899111
L4D2	.0261928	.0827248	<b>0.32</b>	0.752	-.1372289 .1896146
L5D2	.0227313	.0792953	<b>0.29</b>	0.775	-.1339156 .1793782
_cons	.0012743	.0005525	2.31	0.022	.0001828 .0023657

Data la non significatività delle augmentation 4 e 5, si sceglie una statistica DF augmented di 3. Il test al 5% non rifiuta la non stazionarietà per il tasso di inflazione, anche se, in questo caso, la statistica test è abbastanza prossima all'area di rifiuto di  $H_0$  (-2.613 vs -2.886); in altri termini, il tasso di inflazione è "più vicino ad essere stazionario" di quanto lo sia il livello dei prezzi.

dfuller d.lp, lags(3) regress

Augmented Dickey-Fuller test for unit root                      Number of obs =                      163

		----- Interpolated Dickey-Fuller -----			
Test		1% Critical	5% Critical	10% Critical	
Statistic		Value	Value	Value	
Z(t)		<b>-2.565</b>	-3.489	-2.886	-2.576
* MacKinnon approximate p-value for Z(t) = 0.1005					
D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp					
LD	-.1052704	.0410439	-2.56	0.011	-.1863359 - .0242049
LD2	-.1923163	.0816601	-2.36	0.020	-.3536025 - .0310301
L2D2	-.2348958	.0785479	-2.99	0.003	-.3900351 - .0797565
L3D2	.206795	.0773319	2.67	0.008	.0540573 .3595326
_cons	.0011789	.0005328	2.21	0.028	.0001266 .0022313

Infine, con le differenze seconde dei prezzi (accelerazione/decelerazione dell'inflazione) si ottiene un risultato che rifiuta  $H_0$ : esse sono stazionarie. Si noti che, in questo caso, l'augmentation scelta alla fine è pari a 2 e, dato l'andamento grafico in precedenza, il modello è senza trend.

dfuller d2.lp, lags(2) regress

Augmented Dickey-Fuller test for unit root                      Number of obs =                      163

		----- Interpolated Dickey-Fuller -----			
Test		1% Critical	5% Critical	10% Critical	
Statistic		Value	Value	Value	
Z(t)		<b>-7.889</b>	-3.489	-2.886	-2.576

-----  
 \* MacKinnon approximate p-value for Z(t) = 0.0000  
 -----

D3.lp		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp						
	LD2	-1.373964	.1741604	-7.89	0.000	-1.717931 -1.029998
	LD3	.1088177	.1264894	0.86	0.391	-.1409983 .3586338
	L2D3	-.1775006	.0778139	-2.28	0.024	-.3311828 -.0238185
_cons		.000037	.0002978	0.12	0.901	-.0005511 .0006251

-----

In conclusione, si nota che per i prezzi-inflazione-accelerazione come in molti altri casi, l'esito di un test formale di DF conferma l'impressione di stazionarietà/non stazionarietà già rilevata a livello di rappresentazione grafica delle variabili.

Quando il processo statistico generatore della serie storica  $Y_t$  è stazionario, allora ha un senso stimarne i momenti primo e secondi e rappresentarne il comportamento mediante un modello statistico.

### 3.3.1. Un approfondimento analitico del test ADF

L'effettuazione del test DF richiede la corretta specificazione del modello, ossia la scelta se introdurre o meno il trend deterministico (dall'andamento del grafico) e la scelta dell'ordine dell'augmentation (allo scopo di evitare errori autocorrelati, che, come noto, violerebbero l'assunzione [3] del CLRM e creerebbero problemi di scorretta inferenza).

- Ad esempio, se, dopo essere partiti da un'augmentation pari alla periodicità dei dati +1, la corretta specificazione del modello alla base del test DF ci porta a scrivere il comando finale<sup>4</sup>

```
dfuller y, regress lags(1)
```

il modello che viene stimato è:

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

dove  $\delta = (\beta_1 - 1)$ , con  $\beta_1$  parametro associato alla dipendente ritardata. L'augmentation  $\gamma \Delta y_{t-1}$  con parametro  $\gamma$  significativo indica che se noi scrivessimo erroneamente `dfuller y, regress lags(0)` stimeremmo il modello

$\Delta y_t = b_0 + \delta y_{t-1} + u_t$ , dove  $u_t = \rho u_{t-1} + \varepsilon_t$ , o di ordine più elevato, con  $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$ . Infine, il fatto di utilizzare come dipendente  $\Delta y_t$  non soltanto semplifica l'effettuazione del test di DF (simile ad un test di significatività a zero su un singolo parametro), ma, più importante perché rispetta le ipotesi alla base del CLRM, fa sì che sotto l'ipotesi nulla di non stazionarietà di  $y_t$ ,  $\Delta y_t$  sia stazionaria (se  $y_t \sim I(1)$ , ossia è integrata del primo ordine e basta la differenza prima per renderla stazionaria).

L'ipotesi da sottoporre a verifica con il test DF è  $H_0: \delta = 0$  (ossia  $\beta_1 = 1$ , c'è la radice unitaria o  $y_t$  non è un processo stazionario) contro l'alternativa  $H_1: \delta < 0$  (ossia  $\beta_1 < 1$ , non c'è la radice unitaria o  $y_t$  è un processo stazionario).

Se non rifiutiamo  $H_0$  il modello diventa:

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

---

<sup>4</sup> Le altre augmentation non sono significativamente diverse da zero, valutate con size al 10%.

Il processo  $\Delta y_t$  segue un AR(1) stazionario, mentre  $y_t = \beta_0 + (1+\gamma)y_{t-1} + \gamma y_{t-2} + \varepsilon_t$  è un AR(2) con radice unitaria, non stazionario (notate che la somma dei parametri è 1).

Se rifiutiamo  $H_0$  il modello diventa:

$$y_t = \beta_0 + (\beta_1 + \gamma)y_{t-1} - \gamma y_{t-2} + \varepsilon_t$$

Il processo  $y_t$  segue un AR(2) stazionario.

Le precedenti considerazioni non cambiano nella sostanza per lags più elevati.

- Quando la corretta specificazione del modello alla base del test DF ci porta a scrivere il comando

```
dfuller y, regress lags(0)
```

perché il parametro  $\gamma$  non è significativamente diverso da zero, il modello che viene stimato è:

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

Se non rifiutiamo  $H_0$  il modello diventa:

$$\Delta y_t = \beta_0 + \varepsilon_t$$

Il processo  $\Delta y_t$  è uguale al drift ( $\beta_0$ ) più uno shock casuale white noise ( $\varepsilon_t$ ): non c'è nulla nel passato in grado di spiegare  $\Delta y_t$ . Questo comporta che  $y_t = \beta_0 + y_{t-1} + \varepsilon_t$  è un random walk con drift. Immaginiamo di avere 3 osservazioni temporali:

$$y_1 = \beta_0 + y_0 + \varepsilon_1$$

$$y_2 = \beta_0 + y_1 + \varepsilon_2$$

$$y_3 = \beta_0 + y_2 + \varepsilon_3$$

Procedendo per sostituzione si ottiene:

$$y_3 = \beta_0 + \beta_0 + \beta_0 + y_0 + \varepsilon_3 + \varepsilon_2 + \varepsilon_1$$

In generale, ipotizzando che la condizione iniziale sia nulla ( $y_0=0$ ):

$$y_t = \beta_0 t + \sum_{k=0}^{t-1} \varepsilon_{t-k}, \text{ dove la sommatoria di } \varepsilon \text{ è detta trend stocastico.}$$

Pertanto:

$$E(y_t) = \beta_0 t \text{ e}$$

$$\text{VAR}(y_t) = t\sigma_\varepsilon^2$$

che violano le condizioni di stazionarietà in covarianza.

Se rifiutiamo  $H_0$  il modello diventa:

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

Il processo  $y_t$  segue un AR(1) stazionario. Infatti, immaginiamo, come prima, di avere 3 osservazioni temporali:

$$y_1 = \beta_0 + \beta_1 y_0 + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 y_1 + \varepsilon_2$$

$$y_3 = \beta_0 + \beta_1 y_2 + \varepsilon_3$$

Procedendo per sostituzione si ottiene:

$$y_3 = \beta_0(1+\beta_1+\beta_1^2) + \beta_1^3 y_0 + \varepsilon_3 + \beta_1 \varepsilon_2 + \beta_1^2 \varepsilon_1$$

In generale (si suppone ancora che  $y_0=0$ ):

$$y_t = \beta_0 \sum_{k=0}^{t-1} \beta_1^k + \sum_{k=0}^{t-1} \beta_1^k \varepsilon_{t-k}$$

Pertanto, prendendone il limite per  $t \rightarrow \infty$ :

$E(y_t) = \beta_0 \sum_{k=0}^{\infty} \beta_1^k = \beta_0 \frac{1}{1-\beta_1}$  (NOTA: la condizione di convergenza di una serie geometrica è che

$|\beta_1| < 1$ ) e

$$VAR(y_t) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} (\beta_1^k)^2 = \sigma_\varepsilon^2 \frac{1}{1-\beta_1^2}$$

$$COV(y_t, y_{t-1}) = \sigma_\varepsilon^2 \frac{\beta_1}{1-\beta_1^2}$$

In generale:  $COV(y_t, y_{t-k}) = \sigma_\varepsilon^2 \frac{\beta_1^k}{1-\beta_1^2}$

NOTA che:  $\rho_k = \beta_1^k$

I momenti primi e secondi rispettano le condizioni di stazionarietà in covarianza in quanto non dipendono dal tempo (si veda anche il paragrafo 3.3 più avanti).

➤ Infine, quando la corretta specificazione del modello alla base del test DF ci porta a scrivere il comando

```
dfuller y, regress lags(0) trend
```

perché il parametro  $\gamma$  non è significativamente diverso da zero e il grafico evidenzia la presenza di un trend crescente o decrescente, il modello che viene stimato è:

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

Se non rifiutiamo  $H_0$  il modello diventa:

$$\Delta y_t = \beta_0 + \alpha t + \varepsilon_t$$

per cui  $E(\Delta y_t) = \beta_0 + \alpha t$ . Invece,  $y_t = \beta_0 + \alpha t + y_{t-1} + \varepsilon_t$  è un processo non stazionario, dato dalla somma di un trend deterministico (le prime due sommatorie) e di un trend stocastico (l'ultima sommatoria):

$$y_t = t\beta_0 + \alpha \sum_{k=0}^{t-1} (t-k) + \sum_{k=0}^{t-1} \varepsilon_{t-k} = (\beta_0 + \alpha)t^2 + \alpha \sum_{k=1}^{t-1} (t-k) + \sum_{k=0}^{t-1} \varepsilon_{t-k} .$$

Se rifiutiamo  $H_0$  il modello diventa:

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

Il processo generatore di  $y_t$  segue un AR(1) stazionario intorno ad un trend deterministico:

$$y_t = \beta_0 \sum_{k=0}^{t-1} \beta_1^k + \alpha \sum_{k=0}^{t-1} \beta_1^k (t-k) + \sum_{k=0}^{t-1} \beta_1^k \varepsilon_{t-k}$$

In questo caso occorre inserire il trend (generato con il comando: `g trend=_n`) tra le esplicative dei modelli per  $y_t$ .

### 3.4. (B) Scelta dell'ordine del modello AR(p) di partenza

*Applicazione: un modello AR per i prezzi al consumo negli USA.*

(a) Alla luce dei test di stazionarietà, i log-livelli dei prezzi devono essere differenziati due volte per essere resi stazionari, quindi ci si concentra sulla modellazione AR della variazione del tasso di inflazione. Inoltre, l'analisi del correlogramma per  $\Delta \text{infl}_t$  suggerisce il rifiuto di  $H_0$  del test Q di Box-Pierce: nel passato della serie c'è informazione per prevederne (spiegarne) l'andamento presente e futuro.

(b) Si sceglie  $p=5$  e si stimano i parametri del modello (troppo?) generale AR(5)

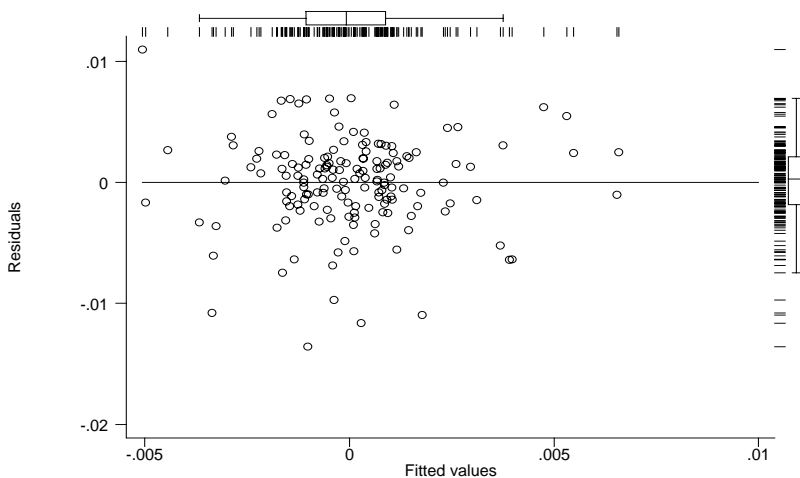
```
reg d2.lp d21.lp d212.lp d213.lp d214.lp d215.lp
```

Source	SS	df	MS	Number of obs = 161		
Model	.000560383	5	.000112077	F( 5, 155) =	7.72	
Residual	.002251216	155	.000014524	Prob > F =	0.0000	
-----				R-squared =	0.1993	
Total	.002811599	160	.000017572	Adj R-squared =	0.1735	
-----				Root MSE =	.00381	
D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----						
lp						
LD2	-.2479529	.0798358	-3.11	0.002	-.4056595	-.0902463
L2D2	-.2972674	.0822008	-3.62	0.000	-.4596457	-.134889
L3D2	.1674624	.084289	1.99	0.049	.000959	.3339657
L4D2	-.0231732	.0820393	-0.28	0.778	-.1852326	.1388863
L5D2	-.0154618	.0793887	-0.19	0.846	-.1722851	.1413615
_cons	.0000533	.0003004	0.18	0.859	-.0005401	.0006468

### 3.5. (C) Analisi dei test diagnostici del modello AR(p)

(c) Test diagnostici sui residui di stima del precedente modello univariato AR(5).

```
rvfplot ,oneway twoway box yline(0) ylabel xlabel
```



```
hettest
```

Cook-Weisberg test for heteroskedasticity using fitted values of D2.lp

```
Ho: Constant variance
chi2(1) = 2.92
Prob > chi2 = 0.0874
```

```
ovtest
```

Ramsey RESET test using powers of the fitted values of D2.lp

```
Ho: model has no omitted variables
F(3, 152) = 0.95
Prob > F = 0.4202
```

Sinora nessun rilevante problema di eteroschedasticità e linearità. Poi salvo i residui:

```
predict resar5,resid
```

(7 missing values generated)

```
sktest resar5
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
resar5	0.001	0.003	15.44	0.0004

Si verifica l'autocorrelazione dei residui (sotto  $H_0$  sono white noise, cioè il correlogramma dei residui è "vuoto"):

```
corrgram resar5 , lags(20)
```

LAG	AC	PAC	Q	Prob>Q	-1 [Autocorrelation]	0	1 -1 [Partial Autocor]	0	1
1	0.0181	0.0182	.05388	0.8164					
2	-0.0222	-0.0230	.13532	0.9346					
3	0.0123	0.0137	.16034	0.9837					
4	-0.0284	-0.0298	.29487	0.9901					
5	0.0686	0.0713	1.0866	0.9553					
6	0.0263	0.0214	1.2038	0.9767					
7	-0.0031	-0.0012	1.2054	0.9908					
8	-0.2317	-0.2366	10.416	0.2371	-			-	
9	-0.0950	-0.0851	11.975	0.2147					
10	-0.0178	-0.0297	12.029	0.2831					
11	-0.1050	-0.1121	13.958	0.2353					
12	-0.1345	-0.1562	17.143	0.1443	-			-	
13	0.0649	0.0974	17.89	0.1618					
14	-0.0483	-0.0388	18.305	0.1932					
15	-0.0804	-0.0821	19.468	0.1933					
16	0.0054	-0.0522	19.473	0.2449					
17	0.0073	-0.0042	19.483	0.3015					
18	0.0798	0.0629	20.651	0.2974					
19	-0.0911	-0.1866	22.185	0.2752				-	
20	0.0551	-0.0110	22.75	0.3013					

Nel complesso, i test Q non rifiutano mai  $H_0$  che i coefficienti di autocorrelazione dei residui siano tutti nulli e, quindi, non c'è nei residui traccia di misspecificazione della dinamica; in altri termini,  $p$  è abbastanza generale da spiegare completamente la struttura dinamica dei dati di  $\Delta lp_t$ .

Passiamo ora ai test di Durbin-Watson sull'autocorrelazione del I ordine e al test di Breusch-Godfrey (di tipo LM) sull'autocorrelazione di ordine  $\geq 1$ .

```
qui reg d2.lp ld2.lp l2d2.lp l3d2.lp l4d2.lp l5d2.lp
```

```
. dwstat
```

```
Durbin-Watson d-statistic( 6, 161) = 1.939939
```

```
. bgodfrey, lags(1)
```

```
Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	3.584	1	0.0583

$H_0$ : no serial correlation

Il test di Godfrey è ottenuto da una regressione ausiliaria in cui occorre utilizzare un accorgimento. Se, infatti, si stimasse direttamente la seguente regressione ausiliaria:

```
. reg resar5 1.resar5 ld2.lp l2d2.lp l3d2.lp l4d2.lp l5d2.lp
```

Source	SS	df	MS	Number of obs = 160			
Model	2.6804e-06	6	4.4674e-07	F( 6, 153)	=	0.03	
Residual	.002201063	153	.000014386	Prob > F	=	0.9999	
Total	.002203744	159	.00001386	R-squared	=	0.0012	
				Adj R-squared	=	-0.0380	
				Root MSE	=	.00379	

resar5	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
resar5						
L1	1.481714	5.11268	0.29	0.772	-8.618846	11.58227
lp						
LD2	-1.46347	5.112177	-0.29	0.775	-11.56304	8.636097
L2D2	-.372497	1.268284	-0.29	0.769	-2.878107	2.133113
L3D2	-.4307946	1.53568	-0.28	0.779	-3.46467	2.60308
L4D2	.2480797	.8835452	0.28	0.779	-1.497444	1.993603
L5D2	-.0317167	.1257797	-0.25	0.801	-.2802059	.2167726
_cons	.0000338	.0004037	0.08	0.933	-.0007637	.0008313

non sarebbe possibile calcolare correttamente la statistica test di Godfrey, in quanto il numero di osservazioni è 160 invece che 161. Si noti come l'osservazione iniziale del periodo di stima sia missing.

```
. list tempo resar5
```

	tempo	resar5
1.	1959q1	.
2.	1959q2	.
3.	1959q3	.
4.	1959q4	.
5.	1960q1	.
6.	1960q2	.
<b>7.</b>	<b>1960q3</b>	.
8.	1960q4	.0068686
9.	1961q1	-.005551
10.	1961q2	-.0008162

Per mantenere parità nel numero di osservazioni tra stima del modello univariato AR(5) e regressione ausiliaria occorre rimpiazzare con 0 il missing dell'osservazione iniziale:

```
. replace resar5=0 if tin(1960q3, 1960q3)
(1 real change made)
```

dove la funzione tin sta per  $\leq$  (mentre twithin si usa per  $<$ ).

```
. list tempo resar5
```

	tempo	resar5
1.	1959q1	.

2.	1959q2	.
3.	1959q3	.
4.	1959q4	.
5.	1960q1	.
-----		
6.	1960q2	.
7.	<b>1960q3</b>	<b>0</b>
8.	1960q4	.0068686
9.	1961q1	-.005551
10.	1961q2	-.0008162

Ri-stimando la regressione ausiliaria:

```
. reg resar5 l.resar5 ld2.lp l2d2.lp l3d2.lp l4d2.lp l5d2.lp
```

Source	SS	df	MS	Number of obs = 161		
Model	.000050117	6	8.3529e-06	F( 6, 154)	=	0.58
Residual	.002201099	154	.000014293	Prob > F	=	0.7424
-----				R-squared	=	0.0223
Total	.002251216	160	.00001407	Adj R-squared	=	-0.0158
-----				Root MSE	=	.00378

resar5	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
resar5						
lp						
L1	1.228308	.6559543	1.87	0.063	-.0675216	2.524138
LD2	-1.210057	.6510426	-1.86	0.065	-2.496184	.07607
L2D2	-.3097864	.1844406	-1.68	0.095	-.6741465	.0545738
L3D2	-.3547468	.2070778	-1.71	0.089	-.7638266	.054333
L4D2	.2044512	.1361775	1.50	0.135	-.0645659	.4734683
L5D2	-.0268785	.0800519	-0.34	0.738	-.18502	.1312629
_cons	.0000203	.0002982	0.07	0.946	-.0005688	.0006094

è ora possibile calcolare la statistica test di Godfrey ed il corrispondente valore di probabilità.

```
. di 161*0.0223
3.5903
```

```
. di chi2tail(1, 161*0.0223)
.05811775
```

Per mettervi alla prova, provate a verificare la presenza di autocorrelazione dei residui del secondo ordine utilizzando la regressione ausiliaria; verificate poi che i vostri calcoli siano giusti con il comando `bgodfrey, lags(2)`.

### 3.6. (D) Riduzione dell'ordine del modello AR(p)

(d) Verifica della significatività dei parametri. Il modello AR(5) sopra evidenzia la non significatività dei parametri corrispondenti ai lags 4 e 5, quindi si stima un modello in cui i corrispondenti parametri sono vincolati a zero. Infine torno a verificare l'autocorrelazione dei residui.

```
reg d2.lp d2l.lp d2l2.lp d2l3.lp
```

Source	SS	df	MS	Number of obs = 163		
Model	.000569078	3	.000189693	F( 3, 159)	=	13.13
Residual	.002297163	159	.000014448	Prob > F	=	0.0000
-----				R-squared	=	0.1985

-----+-----				Adj R-squared = 0.1834	
Total   .002866241 162 .000017693				Root MSE = .0038	
-----+-----					
D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
lp					
LD2	-.2651467	.0778955	-3.40	0.001	-.41899 -.1113034
L2D2	-.2863184	.0772668	-3.71	0.000	-.43892 -.1337168
L3D2	.1775006	.0778139	2.28	0.024	.0238185 .3311828
_cons	.000037	.0002978	0.12	0.901	-.0005511 .0006251
-----+-----					

predict resar3,resid  
(5 missing values generated)

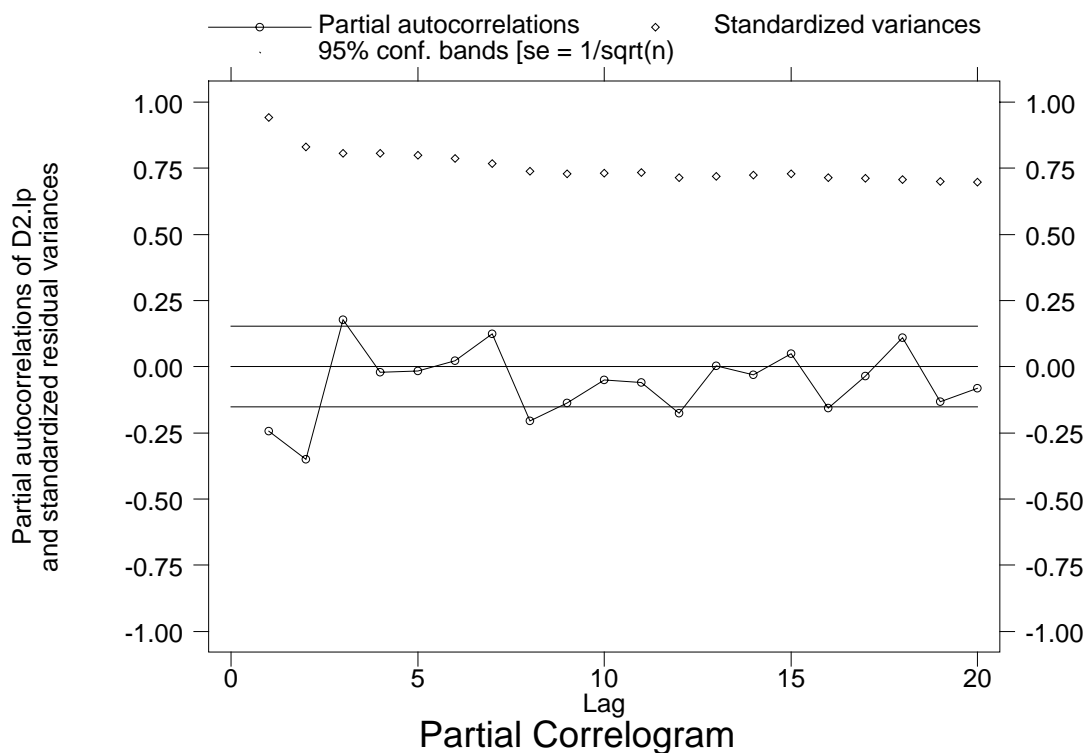
corrgram resar3 , lags(20)

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]		[Partial Autocor]			
-----+-----										
1	0.0091	0.0091	.01375	0.9067						
2	-0.0153	-0.0155	.05273	0.9740						
3	-0.0046	-0.0037	.05634	0.9965						
4	-0.0394	-0.0404	.319	0.9886						
5	0.0637	0.0667	1.0097	0.9618						
6	0.0272	0.0211	1.136	0.9799						
7	0.0059	0.0077	1.1419	0.9922						
8	-0.2327	-0.2364	10.538	0.2293	-			-		
9	-0.0906	-0.0765	11.97	0.2150						
10	-0.0183	-0.0298	12.029	0.2831						
11	-0.0988	-0.1115	13.756	0.2468						
12	-0.1377	-0.1685	17.132	0.1447	-			-		
13	0.0787	0.1057	18.243	0.1485						
14	-0.0531	-0.0499	18.752	0.1747						
15	-0.0656	-0.0704	19.533	0.1906						
16	-0.0021	-0.0628	19.534	0.2419						
17	0.0066	-0.0030	19.542	0.2983						
18	0.0749	0.0558	20.584	0.3009						
19	-0.0988	-0.1969	22.408	0.2644					-	
20	0.0639	-0.0055	23.176	0.2802						

Il modello AR(3) è una adeguata specificazione delle variazioni del tasso di inflazione e, quindi, può essere utilizzato in previsione.

Un modo alternativo di identificare l'ordine p del modello autoregressivo più adatto per la serie storica di interesse è l'ispezione delle sue autocorrelazioni parziali (PAC). In pratica, i ritardi (LAG) significativamente diversi da zero (che escono dall'intervallo di confidenza) indicano con buona approssimazione l'ordine del modello AR da utilizzare.

pac d2.lp , lags(20)



Nel presente caso, differenze prime del tasso di inflazione, solo i primi tre lags sono significativi e, quindi, l'ispezione del correlogramma parziale implica una scelta preliminare di AR(3) che, alla luce dei risultati precedenti, risulta essere la migliore.

**Esercizio controfattuale (per conto vostro):** scegliete di stimare al passo (b) un modello AR(1) ed effettuate i test diagnostici sui residui: ci si aspetta che i residui segnalino la scorretta specificazione dell'ordine  $p=1$  del modello rifiutando  $H_0$ . A mano a mano che aumentate l'ordine del modello AR i residui dovrebbero "mettersi a posto". Svolgimento:

```
reg d2.lp ld2.lp
```

Source	SS	df	MS	Number of obs =	165
Model	.000171343	1	.000171343	F( 1, 163) =	10.26
Residual	.002722186	163	.000016701	Prob > F =	0.0016
Total	.002893529	164	.000017643	R-squared =	0.0592
				Adj R-squared =	0.0534
				Root MSE =	.00409

D2.lp		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp	LD2	-.2429689	.0758547	-3.20	0.002	-.3927535 -.0931842
_cons		.0000221	.0003182	0.07	0.945	-.0006061 .0006504

```
predict resar1, resid
```

corrgram resarl, lag(20)

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	-0.0872	-0.0872	1.2764	0.2586						
2	-0.2785	-0.2882	14.387	0.0008	--			--		
3	0.2673	0.2357	26.539	0.0000		--			--	
4	-0.0195	-0.0810	26.605	0.0000						-
5	-0.1376	0.0091	29.867	0.0000	-					
6	0.1471	0.0512	33.617	0.0000		-				
7	0.0616	0.0625	34.279	0.0000						
8	-0.3012	-0.2508	50.204	0.0000	--			--		
9	-0.0439	-0.0894	50.545	0.0000						
10	0.1267	-0.0430	53.4	0.0000		-				
11	-0.1697	-0.0921	58.552	0.0000	-					
12	-0.1415	-0.1527	62.159	0.0000	-			-		
13	0.1487	0.0333	66.167	0.0000		-				
14	-0.0454	-0.0267	66.543	0.0000						
15	-0.1025	0.0183	68.473	0.0000						
16	0.0066	-0.1704	68.482	0.0000					-	
17	0.0541	0.0347	69.028	0.0000						
18	0.0377	0.0679	69.294	0.0000						
19	-0.0996	-0.1663	71.167	0.0000					-	
20	0.0510	-0.0377	71.66	0.0000						

Evidentemente, il modello AR(1) non dà conto della dinamica di secondo e terzo ordine dei dati e, quindi, i residui segnalano tracce di autocorrelazione ai corrispondenti lags.

### 3.7. Le previsioni univariate e non condizionale con il modello AR

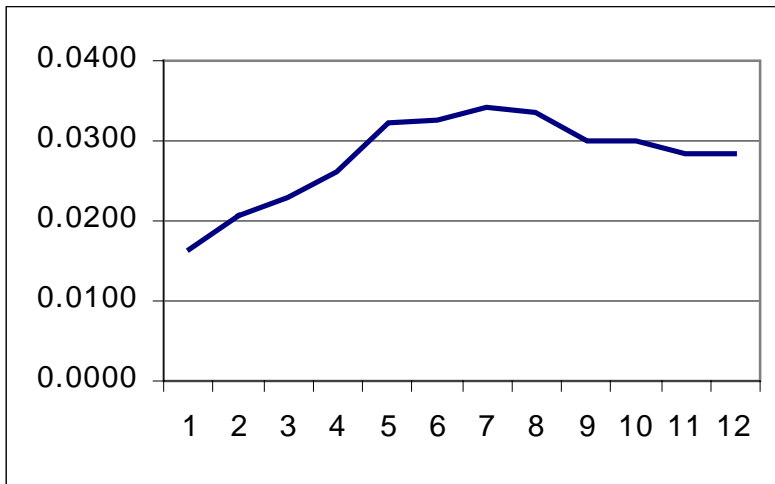
Ottenuto un modello quantitativo univariato per la variabile di interesse  $Y_t$ , la ricerca si può sviluppare in due direzioni alternative: (1) usare il modello in un esercizio di previsione non condizionale (in questo paragrafo); (2) estendere il modello econometrico per dare conto di altre determinanti del tasso di inflazione, oltre alla sua storia passata, seguendo in ciò i suggerimenti della teoria economica (nel paragrafo 4).

La previsione non condizionale del modello AR(3) per le variazioni del tasso di inflazione viene ottenuta nel foglio "forecast" del file Excel datiTS.xls. In pratica le operazioni si articolano nelle seguenti tappe:

- previsione della variabile trasformata (la variabile dipendente del modello AR, nel nostro caso  $d2.lp$ ). Tale previsione è non condizionale perché la sua formulazione richiede solo la conoscenza dei valori storici di fine periodo della variabile da prevedere. La previsione condizionale richiede anche dati e ipotesi esterne sull'andamento storico e futuro di altre variabili economiche.
- ricostruzione della previsione (in grassetto corsivo) dell'inflazione a partire dalla previsione (in grassetto) delle variazioni di inflazione: [ $d.lp = d2.lp + ld.lp$ ] e, ancora all'indietro, ricostruzione della previsione del livello dei prezzi (in grassetto corsivo) a partire dalla previsione dell'inflazione (in grassetto) ottenuta al passaggio precedente: [ $lp = d.lp + l.lp$ ].
- la previsione dei log-livelli dei prezzi può essere comunicata più incisivamente mediante un grafico che presenta, senza soluzione di continuità, dati storici e previsti per il tasso di inflazione annuo tendenziale, ottenuto con la differenza quarta dei dati storico-previsti del log-livello dei prezzi.

In Excel, i dati storici sono disponibili fino al 2000q4 e la nostra previsione è effettuata dal 2001q1 al 2001q4 (un anno - quattro trimestri - avanti). Nella cartella “forecast”, i dati previsti sono in rosso, quelli storici in nero, le stime dei parametri del modello AR(3) sono su sfondo giallo e i calcoli intermedi hanno sfondo verde.

Grafico dell'inflazione tendenziale dal 1999q1  
(storia da 1 a 8, previsione da 9 a 12)



Come evidente, la nostra previsione estrapola, nei quattro trimestri del 2001, una tendenza al declino del tasso di inflazione la cui manifestazione iniziava nell'ultima parte del campione (osservazioni numero 7-8, cioè nel 2000 q3-q4).

### 3.8. Riepilogo dei temi trattati e un esempio finale

In questo paragrafo abbiamo presentato alcune tecniche di analisi e modellazione univariata per serie storiche. Tali tecniche si fondano sulla dipendenza temporale delle realizzazioni delle variabili di questo tipo. In particolare, due temi rivestono grande importanza: (a) la stazionarietà del processo stocastico che genera i dati; (b) l'autocorrelazione sia della variabile di interesse, sia dei residui di stima del modello AR e, quindi, la specificazione dinamica del modello

La validità dell'approccio AR si fonda, come visto, sulla proprietà statistica della stazionarietà. Si potrebbe dire, in una battuta, che la previsione con il modello autoregressivo ipotizza che il futuro sarà come il passato e un tale assunto richiede la stazionarietà della variabile modellata.

L'esistenza di autocorrelazione fra le realizzazioni della variabile di interesse consentono la modellazione e previsione mediante modelli AR. L'autocorrelazione dei residui ha una duplice valenza: da un lato, residui autocorrelati segnalano una scorretta specificazione della dinamica del modello AR; dall'altro, la presenza di autocorrelazione nei residui fa venire meno la proprietà della consistenza dello stimatore OLS (abituamente utilizzato per stimare i parametri del modello). In altri termini, le autocorrelazioni  $\rho_k$  sono la principale fonte di informazione sulla struttura statistica della serie e, pertanto, sono auspicabili, mentre i residui non devono essere autocorrelati se il modello che li genera è una accettabile rappresentazione parametrica dei dati.

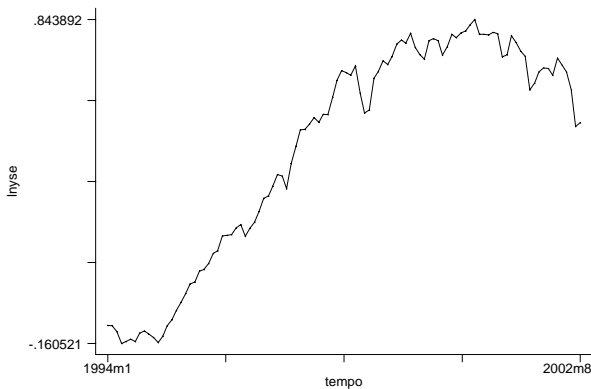
**Esercizio:** *abbiamo scoperto un modo semplice per fare soldi?* Non è poi così difficile: sulla base dell'informazione storica, stimo un modello AR per le quotazioni mensili del New York Stock Exchange e lo uso per prevedere le future realizzazioni delle quotazioni azionarie. Perfetto, almeno in teoria, peccato però che gli studi di Fama, seguendo la teoria dell'efficienza dei mercati

finanziari, spiegano perché la storia delle quotazioni non può costituire una informazione utile per la loro previsione.

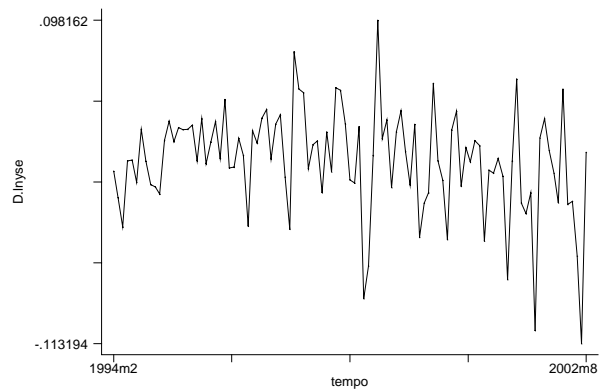
Per provare tale assunto, si copia in Stata il contenuto della cartella “mensile” del foglio Excel “datiTS,xls” (vedi esercizio sopra) e immetto la sequenza di comandi:

```
g tempo = ym(anno,mese)
tsset tempo, monthly
time variable: tempo, 1994m1 to 2002m8
g lnyse=log(nyse)
```

```
graph lnyse tempo, c(1) s(.
```



```
graph d.lnyse tempo, c(1) s(.
```



```
dfuller d.lnyse, lag(13) regress
```

Augmented Dickey-Fuller test for unit root                      Number of obs    =                      89

	Test Statistic	----- Interpolated Dickey-Fuller ----- 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-0.958	-3.525	-2.899	-2.584

\* MacKinnon approximate p-value for Z(t) = 0.7692

D2.lnyse	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----						
lnyse						
LD	-.3639631	.3798397	-0.96	0.341	-1.12081 .3928841	
LD2	-.4941638	.3884899	-1.27	0.207	-1.268247 .2799192	
L2D2	-.5897204	.3827125	-1.54	0.128	-1.352292 .1728508	
L3D2	-.6244028	.3763332	-1.66	0.101	-1.374263 .1254574	
L4D2	-.6746776	.3695495	-1.83	0.072	-1.411021 .0616658	
L5D2	-.5457718	.3628464	-1.50	0.137	-1.268759 .1772155	
L6D2	-.560105	.3510277	-1.60	0.115	-1.259543 .139333	
L7D2	-.6013212	.330402	-1.82	0.073	-1.259662 .0570192	
L8D2	-.4628788	.3125943	-1.48	0.143	-1.085736 .1599789	
L9D2	-.2712002	.287963	-0.94	0.349	-.8449791 .3025786	
L10D2	-.1109798	.2551529	-0.43	0.665	-.6193831 .3974235	
L11D2	-.048841	.219508	-0.22	0.825	-.4862201 .3885382	
L12D2	.0306064	.174721	0.18	0.861	-.3175329 .3787457	
L13D2	.1619072	.134948	1.20	0.234	-.1069825 .4307969	
_cons	.0012618	.0052233	0.24	0.810	-.0091459 .0116696	

dfuller d.lnyse, lag(8) regress

Augmented Dickey-Fuller test for unit root                      Number of obs    =                      94

Test Statistic	----- Interpolated Dickey-Fuller -----		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-1.793	-3.518	-2.895

\* MacKinnon approximate p-value for Z(t) = 0.3840

D2.lnyse	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnyse						
LD	-.5705384	.3182185	-1.79	0.077	-1.203351	.062274
LD2	-.2463712	.3116907	-0.79	0.431	-.8662022	.3734599
L2D2	-.3152173	.2914184	-1.08	0.282	-.8947348	.2643002
L3D2	-.3467735	.2693375	-1.29	0.201	-.8823806	.1888335
L4D2	-.3771079	.2481116	-1.52	0.132	-.870505	.1162892
L5D2	-.2404188	.2206341	-1.09	0.279	-.6791739	.1983363
L6D2	-.2831996	.1862224	-1.52	0.132	-.6535232	.087124
L7D2	-.3061929	.1510608	-2.03	0.046	-.6065938	-.005792
L8D2	-.183995	.1173603	-1.57	0.121	-.4173788	.0493888
_cons	.0035721	.0045617	0.78	0.436	-.0054994	.0126436

dfuller d.lnyse, lag(7) regress

Augmented Dickey-Fuller test for unit root                      Number of obs    =                      95

Test Statistic	----- Interpolated Dickey-Fuller -----		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.343	-3.517	-2.894

\* MacKinnon approximate p-value for Z(t) = 0.1585

D2.lnyse	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnyse						
LD	-.7098368	.3029667	-2.34	0.021	-1.312115	-.1075591
LD2	-.0827008	.2905889	-0.28	0.777	-.6603724	.4949708
L2D2	-.16227	.2691948	-0.60	0.548	-.6974115	.3728716
L3D2	-.1975253	.2479466	-0.80	0.428	-.6904269	.2953763
L4D2	-.2128752	.2201751	-0.97	0.336	-.6505688	.2248183
L5D2	-.0715461	.1864962	-0.38	0.702	-.4422883	.299196
L6D2	-.1169053	.1511343	-0.77	0.441	-.4173503	.1835396
L7D2	-.1562316	.1166134	-1.34	0.184	-.3880514	.0755882
_cons	.0045872	.0044602	1.03	0.307	-.0042795	.0134538

dfuller d.lnyse, lag(0) regress

Dickey-Fuller test for unit root                      Number of obs    =                      102

Test Statistic	----- Interpolated Dickey-Fuller -----		
	1% Critical Value	5% Critical Value	10% Critical Value

Z(t)	-8.270	-3.509	-2.890	-2.580
------	--------	--------	--------	--------

\* MacKinnon approximate p-value for Z(t) = 0.0000

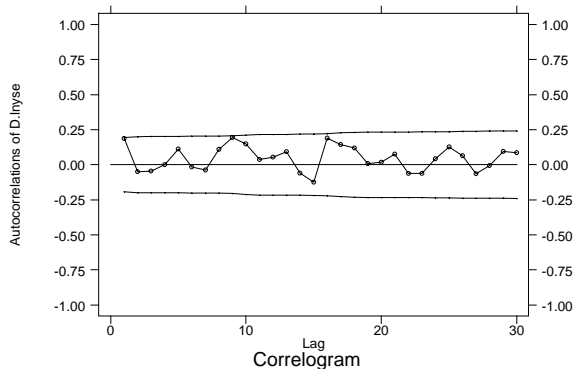
D2.lnyse		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnyse	LD	-.8122833	.0982169	-8.27	0.000	-1.007143 - .6174239
_cons		.0050388	.0034332	1.47	0.145	-.0017726 .0118503

corrgram d.lnyse, lag(30)

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.1877	0.1877	3.7341	0.0533						
2	-0.0502	-0.1025	4.0035	0.1351						
3	-0.0452	-0.0284	4.2249	0.2382						
4	0.0006	0.0050	4.225	0.3764						
5	0.1115	0.1235	5.5959	0.3475						
6	-0.0168	-0.0673	5.6274	0.4662						
7	-0.0390	-0.0117	5.7984	0.5635						
8	0.1086	0.1562	7.142	0.5214						
9	0.1950	0.1840	11.517	0.2419						
10	0.1491	0.1326	14.103	0.1683						
11	0.0362	0.0599	14.258	0.2191						
12	0.0547	0.1147	14.613	0.2633						
13	0.0927	0.1035	15.645	0.2688						
14	-0.0609	-0.1619	16.095	0.3076						
15	-0.1247	-0.1349	18.005	0.2624						
16	0.1884	0.3766	22.416	0.1303						
17	0.1443	-0.0020	25.034	0.0940						
18	0.1181	0.0836	26.809	0.0826						
19	0.0084	0.0712	26.818	0.1090						
20	0.0167	0.1056	26.855	0.1394						
21	0.0753	0.0371	27.602	0.1518						
22	-0.0629	-0.2257	28.13	0.1714						
23	-0.0629	0.0906	28.665	0.1917						
24	0.0419	0.2235	28.905	0.2237						
25	0.1276	0.0815	31.164	0.1836						
26	0.0623	-0.0220	31.71	0.2029						
27	-0.0649	-0.0715	32.309	0.2208						
28	-0.0075	0.1276	32.317	0.2618						
29	0.0954	0.0846	33.648	0.2524						
30	0.0855	0.0280	34.732	0.2525						

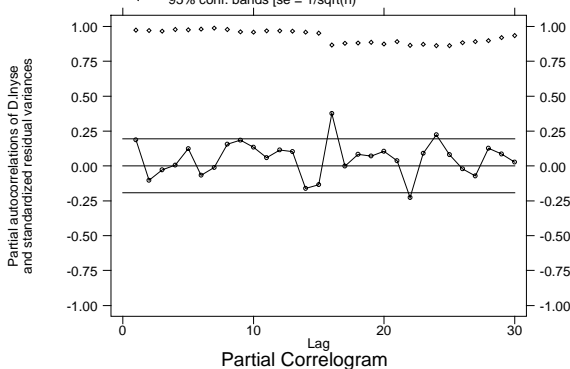
ac d.lnyse, lag(30)

Bartlett's formula for MA(q) 95% confidence bands



pac d.lnyse, lag(30)

Partial autocorrelations and standardized residual variances



L'esito del test conferma lo scetticismo degli econometrici e degli economisti che di solito accompagna qualsiasi tentativo di modellazione statistica dei momenti primi delle quotazioni azionarie.

Una citazione che vale la pena di ricordare è quella del premio Nobel Clive Granger: "L'unico modo certo per far soldi con metodi statistici applicati ai mercati di borsa è quello di scrivere libri di econometria finanziaria, convincere il pubblico che sono utili e incassare i diritti d'autore per le copie vendute."

Se questa è la modesta performance di tecniche statisticamente fondate di tipo ARMA, quale pensate possa essere l'efficacia della analisi tecnica, un metodo molto meno sofisticato da un punto di vista metodologico?

#### **4. Elementi di analisi multivariata: il modello uniequazionale ARDL(p, q)**

L'analisi e la modellazione univariata delle serie storiche economiche offre strumenti semplici e utili per l'analisi preliminare dei dati e per l'ottenimento di previsioni non condizionali. Come già rilevato, uno dei meriti di tale approccio è quello di avere posto l'accento sull'importanza della stazionarietà e della specificazione dinamica del modello, cui gli economisti quantitativi non avevano prestato troppa attenzione (almeno fino alla fine degli anni settanta).

L'essenza dell'analisi multivariata consiste nell'estensione dell'approccio univariato grazie alle informazioni sui nessi di causalità fra variabili che si desumono dalle teorie economiche. L'analisi multivariata che affronteremo in questo paragrafo è di tipo uniequazionale perché si concentra su relazioni economiche misurate in una sola equazione.

Come visto, il modello AR per il tasso di inflazione spiega il tasso di inflazione solo utilizzando l'informazione sulle passate realizzazioni dell'inflazione (analisi univariata della persistenza). In questo paragrafo, invece, studieremo le interrelazioni fra l'andamento del tasso di inflazione e il tasso di disoccupazione.

Prima di affrontare il problema econometrico bisogna prendere in considerazione le indicazioni teoriche a monte di tale interrelazione. Il principio teorico di base che considereremo è il c.d. modello del "triangolo" (cfr. Gordon, 1982), così definito perché individua le tre fondamentali determinanti del tasso di inflazione: inerzia, effetti di domanda e effetti di offerta. In simboli:

$$\pi_t = f(\pi_{t-i}; u_{t-i}; \varepsilon_t) \quad (1)$$

dove  $\pi_t$  è il tasso di inflazione;  $u_t$  è il tasso di disoccupazione;  $\varepsilon_t$  è un disturbo stocastico puramente casuale.

Le tre determinanti dell'inflazione nel modello di Gordon sono indicate nella formula (1):

- (i)  $\pi_{t-i}$  rappresenta l'inerzia (cioè i valori realizzatisi per l'inflazione nei periodi precedenti), ovvero la dinamica temporale;
- (ii)  $u_{t-i}$  rappresenta gli effetti di domanda (misurati dai valori realizzatisi per il tasso di disoccupazione nei periodi precedenti); si suppone che una maggiore domanda implichi un decremento congiunturale (cioè temporaneo) del tasso di disoccupazione rispetto al suo livello di pieno impiego;
- (iii)  $\varepsilon_t$  rappresenta gli effetti di offerta (misurati dallo shock stocastico  $\varepsilon$ ); si suppone che gli shock di offerta siano tra di loro incorrelati nel tempo.

Il passaggio dalla formulazione teorica (1) al modello stimabile, implica una specificazione dinamica generale di tipo AutoRegressive Distributed Lags (ARDL).

Il nome ARDL di questo tipo di modelli deriva dal fatto che in essi si presentano al contempo componenti autoregressive (i ritardi della variabile dipendente) e componenti a ritardi distribuiti relative ad altre variabili esplicative (nel nostro caso il tasso di disoccupazione,  $u$ ). La giustificazione teorica per tale specificazione viene ricondotta all'equazione (1). Vale la pena di sottolineare che il modello ARDL ipotizza la linearità della generica relazione  $f(\dots)$ .

Adottata una specificazione ARDL, la prima scelta da compiere è quella relativa all'ordine della dinamica del processo statistico. Nel caso semplificato di una dinamica del primo ordine sia per la variabile dipendente sia per quella esplicativa, il modello ARDL(1, 1) è:

$$\pi_t = c + b_1 \pi_{t-1} + a_1 u_{t-1} + \varepsilon_t \quad (2)$$

con  $c$ ,  $b_1$ ,  $a_1$  parametri da stimare e  $\varepsilon_t$  identicamente ed indipendentemente distribuito (cioè a varianza costante e incorrelato con le passate realizzazioni).

La soluzione di *steady state* (stato stazionario non stocastico) dell'equazione (2) la si ottiene imponendo:

$$\pi_t = \pi_{t-1} = \pi^*$$

$$u_t = u_{t-1} = u^*$$

$$\varepsilon_t = 0$$

da cui si ottiene

$$\pi^* = c + b_1 \pi^* + a_1 u^* + 0, \text{ e con un po' di algebra:}$$

$$(1 - b_1) \pi^* = c + a_1 u^* \quad (3)$$

La soluzione di stato stazionario non stocastico sub (3) permette di gettare un ponte fra le stime statistiche dell'equazione (2) e i diversi modelli previsti dalla teoria economica. Infatti, la teoria economica difficilmente parla lo stesso linguaggio delle serie storiche ma si limita a delineare i legami di causalità da una variabile all'altra. A conferma di ciò, si valuti la differenza fra la generica scrittura (1) e il modello stimabile (2).

Nel contesto in esame, se nella (3)  $|b_1| < 1$  e  $a_1 < 0$  si ha che:  $\pi^* = c/(1-b_1) + a_1/(1-b_1) u^*$  cioè che esiste un trade off di lungo periodo fra  $\pi$  e  $u$  (cfr curva di Phillips):  $a_1/(1-b_1)$ . In altri termini, l'esistenza di un legame (trade-off) di lungo periodo fra inflazione e disoccupazione può essere valutata alla luce di una combinazione delle stime parametriche del modello statistico ARDL sub (2).

se invece nella (3) si ha che:  $b_1=1$  e  $a_1 < 0$ ; allora, sostituendo, si ha:  $0 = c + a_1 u^*$  da cui l'evidenza che esiste un NAIRU costante nel tempo pari a:  $u^* = -c/a_1$ . In altri termini,  $b_1=1$  implica l'assenza di un trade-off di lungo periodo alla Phillips: dato un qualsiasi livello di inflazione, esiste un livello  $u^*$  in corrispondenza del quale il tasso di inflazione non varia.

I due casi riportati sopra, corrispondenti a  $|b_1| < 1$  oppure a  $b_1=1$ , possono essere incrociati con l'ipotesi che  $a_1=0$  (anziché  $a_1 < 0$ , vedi sopra). In entrambi i casi,  $a_1=0$  evidenzia che il tasso di disoccupazione non interagisce con il tasso di inflazione (assenza di effetti di domanda nel lungo periodo). Quindi, corrispondentemente,

se  $|b_1| < 1$  e  $a_1=0$ , si ha che  $\pi^* = c/(1-b_1)$ , cioè il tasso di inflazione non dipende dalla disoccupazione nel lungo periodo e il miglior modo di rappresentare l'inflazione sarebbe un modello AR(1) stazionario;

se  $b_1=1$  e  $a_1=0$ , si ha che  $\pi$  è un *random walk* e il miglior modo di rappresentare l'inflazione sarebbe un modello AR(1) non stazionario.

In che modo questi risultati possono fare riferimento a ciò che abbiamo ottenuto sviluppando in precedenza l'analisi univariata per il tasso di inflazione? E' semplice: avendo ottenuto sopra che, a livello univariato, l'inflazione è I(1) abbiamo che, in assenza di effetti di disoccupazione, il tasso di inflazione misurato dalla precedente serie storica non ammette una soluzione (costante) di lungo periodo. In altri termini, il campione di dati a nostra disposizione per l'inflazione non è informativo per misurare (stimare) un livello non condizionale di inflazione di lungo periodo verso il quale essa converge in assenza di shock di domanda e/o di offerta. Essendo l'inflazione non stazionaria, I(1), la comprensione del suo comportamento di lungo periodo usando i nostri dati campionari richiede la valutazione delle sue interrelazioni con il tasso di disoccupazione.

D'altro canto, non vi sarà sfuggito che il modello univariato per il tasso di inflazione (in differenze) che abbiamo stimato precedentemente non è del primo ordine. Un secondo fatto che possiamo anticipare a priori è che la stima del modello ARDL(1, 1) presenterà probabilmente residui autocorrelati di ordine superiore al primo. Questa autocorrelazione è attesa scaturire da una scorretta specificazione della dinamica (del primo ordine) del modello (2).

Per valutare questo, possiamo stimare il modello (2) con i nostri dati (Usquarter.dta):

```
. reg d.lp dl.lp l.u
```

Source	SS	df	MS			
Model	.00701564	2	.00350782	Number of obs =	166	
Residual	.00266567	163	.000016354	F( 2, 163) =	214.50	
Total	.00968131	165	.000058675	Prob > F =	0.0000	
				R-squared =	0.7247	
				Adj R-squared =	0.7213	
				Root MSE =	.00404	

D.lp		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lp	LD	.8521137	.0412876	20.64	0.000	.7705862	.9336413
ur	L1	-.0187257	.0212277	-0.88	0.379	-.0606425	.0231911
_cons		.0027401	.0013246	2.07	0.040	.0001245	.0053556

Di cui andiamo subito a verificare l'ipotesi di bianchezza dei residui

```
. hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of D.lp

chi2(1) = 18.48
Prob > chi2 = 0.0000
```

```
. ovtest
```

```
Ramsey RESET test using powers of the fitted values of D.lp
Ho: model has no omitted variables
F(3, 160) = 3.55
Prob > F = 0.0158
```

```
. dwstat
```

```
Durbin-Watson d-statistic( 3, 166) = 2.324371
```

Come noto, il test di Durbin-Watson verifica solo la presenza di autocorrelazione del primo ordine e soffre di distorsione quando applicato ai residui di un modello dinamico.

Con dati a periodicità superiore a quella annuale è possibile che l'ordine dell'autocorrelazione dei residui si spinga oltre al primo ordine. Per verificare questo, si passa all'analisi del correlogramma dei residui:

```
. predict resARDL11, resid
(2 missing values generated)
```

```
. corrgram resARDL11, lags(20)
```

LAG	AC	PAC	Q	Prob>Q	-1 [Autocorrelation]	0	1 -1	0	1 [Partial Autocor]
1	-0.1639	-0.1642	4.538	0.0332	-			-	
2	-0.1931	-0.2265	10.88	0.0043	-			-	
3	0.3491	0.2984	31.73	0.0000		--		--	
4	-0.0053	0.0670	31.735	0.0000					
5	-0.0751	0.0634	32.711	0.0000					
6	0.1812	0.0960	38.436	0.0000		-			
7	0.1241	0.1695	41.137	0.0000					
8	-0.2191	-0.1584	49.606	0.0000	-			-	
9	0.0318	-0.0659	49.785	0.0000					
10	0.1829	0.0317	55.763	0.0000		-			
11	-0.1040	0.0210	57.71	0.0000					
12	-0.0851	-0.0823	59.022	0.0000					
13	0.1995	0.1044	66.276	0.0000		-			
14	-0.0202	0.0587	66.351	0.0000					
15	-0.0445	0.1204	66.716	0.0000					
16	0.0532	-0.0838	67.243	0.0000					
17	0.0588	0.0501	67.89	0.0000					
18	0.0784	0.1661	69.047	0.0000					
19	-0.0830	-0.0875	70.355	0.0000					
20	0.0715	-0.0232	71.331	0.0000					

Come atteso a priori, l'evidenza di autocorrelazione non si concentra tanto sul primo ordine ma soprattutto sul secondo e terzo ordine (con stime dei coefficienti di autocorrelazione rispettivamente pari a -0.1931 e 0.3491).

Per esercizio, effettuate il test di Godfrey per l'autocorrelazione del quarto ordine dei residui. Le precedenti indicazioni di misspecificazione dinamica del modello ARDL sono confermate dall'esito di questo test?

In generale, l'evidenza riportata sopra suggerisce misspecificazione dinamica del modello. Un modello ARDL(1,1) è troppo restrittivo per spiegare in modo statisticamente appropriato l'andamento temporale del tasso di inflazione. Così come già visto per il caso univariato, il modo più immediato di superare il problema della misspecificazione dinamica è quello di stimare un modello ARDL dalla dinamica più articolata.

Nel nostro caso, essendo i dati trimestrali, scegliamo di stimare un modello ARDL(5, 5):

```
. reg d.lp dl(1/5).lp l(1/5).u
```

Source	SS	df	MS	Number of obs =	162
Model	.00783378	10	.000783378	F( 10, 151) =	70.90
				Prob > F =	0.0000

Residual		.001668448	151	.000011049	R-squared	=	0.8244
-----							
Total		.009502229	161	.00005902	Adj R-squared	=	0.8128
					Root MSE	=	.00332

D.lp		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lp							
	LD	.6006062	.0795597	7.55	0.000	.4434122	.7578003
	L2D	.1016128	.0951077	1.07	0.287	-.0863009	.2895265
	L3D	.2993568	.0872537	3.43	0.001	.126961	.4717526
	L4D	-.0968671	.0892109	-1.09	0.279	-.27313	.0793958
	L5D	.1194309	.0805691	1.48	0.140	-.0397574	.2786191
ur							
	L1	-.729002	.1152519	-6.33	0.000	-.9567167	-.5012874
	L2	.9687626	.2144362	4.52	0.000	.5450799	1.392445
	L3	-.1532989	.2307487	-0.66	0.507	-.6092119	.3026141
	L4	-.4861022	.2200327	-2.21	0.029	-.9208426	-.0513618
	L5	.3518809	.1153759	3.05	0.003	.1239213	.5798405
_cons		.0025947	.001237	2.10	0.038	.0001507	.0050386

### Verifica dell'autocorrelazione dei residui del modello ARDL(5, 5)

. dwstat

Durbin-Watson d-statistic( 11, 162) = 2.002607

. predict resARDL55, resid  
(6 missing values generated)

. corrgram resARDL55, lags(20)

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]		[Partial Autocor]			
1	-0.0065	-0.0066	.00701	0.9333						
2	0.0077	0.0079	.01681	0.9916						
3	0.0158	0.0160	.05827	0.9963						
4	-0.0205	-0.0208	.12881	0.9980						
5	0.0095	0.0098	.14424	0.9996						
6	0.0198	0.0200	.21077	0.9998						
7	0.0454	0.0461	.564	0.9992						
8	-0.1731	-0.1768	5.7347	0.6769	-			-		
9	-0.0278	-0.0266	5.869	0.7530						
10	-0.0508	-0.0509	6.3196	0.7877						
11	-0.0248	-0.0191	6.4276	0.8434						
12	-0.1166	-0.1316	8.8351	0.7169					-	
13	0.1825	0.2016	14.776	0.3215		-				-
14	0.1055	0.1188	16.774	0.2684						
15	-0.0723	-0.0654	17.719	0.2777						
16	0.0705	0.0414	18.625	0.2886						
17	0.0203	0.0382	18.7	0.3460						
18	0.0630	0.0581	19.432	0.3657						
19	-0.1211	-0.1829	22.156	0.2766					-	
20	0.0809	0.0396	23.38	0.2705						

Evidentemente, l'allungamento dell'ordine dei ritardi ha sortito all'esito sperato. Ammettendo un dinamica del quinto ordine l'autocorrelazione dei residui non si presenta più.

Ciò che invece rimane è la presenza di eteroschedasticità (attenuato rispetto al modello del primo ordine, ma non completamente superato):

```
. hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
Variables: fitted values of D.lp
```

```
chi2(1)      =      4.92
Prob > chi2   =      0.0266
```

Si noti però che, in fase di verifica della significatività dei parametri stimati, possiamo sempre superare il problema dovuto a residui eteroschedastici utilizzando, come visto, standard error robusti:

```
. reg d.lp dl(1/5).lp l(1/5).u , robust
```

```
Regression with robust standard errors                                Number of obs =      162
                                                                    F( 10,   151) =     50.10
                                                                    Prob > F      =     0.0000
                                                                    R-squared     =     0.8244
                                                                    Root MSE     =     .00332
```

D.lp		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lp							
	LD	.6006062	.0894959	6.71	0.000	.4237804	.7774321
	L2D	.1016128	.0968497	1.05	0.296	-.0897427	.2929683
	L3D	.2993568	.1180618	2.54	0.012	.0660904	.5326231
	L4D	-.0968671	.0899767	-1.08	0.283	-.274643	.0809088
	L5D	.1194309	.0811984	1.47	0.143	-.0410009	.2798626
ur							
	L1	-.729002	.1150408	-6.34	0.000	-.9562996	-.5017045
	L2	.9687626	.2083578	4.65	0.000	.5570895	1.380436
	L3	-.1532989	.2574008	-0.60	0.552	-.6618711	.3552733
	L4	-.4861022	.2853571	-1.70	0.091	-1.04991	.0777061
	L5	.3518809	.13447	2.62	0.010	.0861953	.6175666
_cons		.0025947	.0011576	2.24	0.026	.0003076	.0048817

Superato il problema dell'autocorrelazione e dell'eteroschedasticità, siamo pronti per andare a valutare la significatività delle nostre stime. Ma di quali parametri ha senso verificare la significatività per interpretare in senso economico i risultati se, perso il riferimento al modello ARDL di primo ordine sub (2) perché statisticamente inappropriato, abbiamo perso qualsiasi indicazione (interpretazione) sui parametri che governano la soluzione di stato stazionario del nostro modello? In altri termini, quali e dove sono i parametri di interesse  $b_1$  e  $a_1$  di cui sopra in una specificazione del quinto ordine?

La risposta al quesito la si trova riscrivendo il modello (2) spingendoci fino al quinto ordine:

$$\pi_t = c + b_1 \pi_{t-1} + b_2 \pi_{t-2} + b_3 \pi_{t-3} + b_4 \pi_{t-4} + b_5 \pi_{t-5} + a_1 u_{t-1} + a_2 u_{t-2} + a_3 u_{t-3} + a_4 u_{t-4} + a_5 u_{t-5} + \varepsilon_t \quad (4)$$

dove, come prima,  $c$ ,  $a_i$  e  $b_i$  ( $i=1, 2, 3, 4, 5$ ) sono parametri;  $\varepsilon_t \sim iid(0, \sigma^2)$

La corrispondente soluzione di *steady state* (stato stazionario non stocastico) dell'equazione (4) la si ottiene per:

$$\begin{aligned} \pi_t &= \pi_{t-1} = \pi_{t-2} = \pi_{t-3} = \pi_{t-4} = \pi_{t-5} = \pi^* \\ u_t &= u_{t-1} = u_{t-2} = u_{t-3} = u_{t-4} = u_{t-5} = u^* \\ \varepsilon_t &= 0 \end{aligned}$$

da cui si ottiene il corrispondente della soluzione (3) nel caso di un modello ARDL del quinto ordine:

$$\pi^* = c + (b_1+b_2+b_3+b_4+b_5) \pi^* + (a_1+a_2+a_3+a_4+a_5) u^*$$

e con un po' di algebra:

$$(1-b_1-b_2-b_3-b_4-b_5) \pi^* = c + (a_1+a_2+a_3+a_4+a_5) u^* \quad (5)$$

Definendo:  $\alpha = (a_1+a_2+a_3+a_4+a_5)$  e  $\beta = (b_1+b_2+b_3+b_4+b_5)$  la (5) è riscritta nella:

$$(1-\beta) \pi^* = c + \alpha u^* \quad (6)$$

La soluzione di lungo periodo sub (6) permette di capire che ciò che un modello del primo ordine riconduce a due soli parametri, un modello di ordine superiore al primo riconduce alla somma di tutti i corrispondenti parametri. In generale, per un modello ARDL(p, q) avremo che  $\alpha$  è pari alla somma dei q parametri  $a_i$  e  $\beta$  è la somma dei p parametri  $b_i$ .

Definiamo, per praticità,  $\gamma = (b_1+b_2+b_3+b_4+b_5-1) = (\beta-1)$ ; così facendo, la verifica  $\beta = 1$  può essere sostituita dalla più semplice:  $\gamma = 0$ .

La scomodità di avere a che fare, nell'ARDL(5, 5), con 5 stime di  $a_i$  e  $b_i$  anziché due soli parametri come nell'ARDL(1,1) viene superata da una riparametrizzazione - analoga a quella fatta a livello univariato per passare del modello AR all'equazione per la verifica della stazionarietà col test di Dickey-Fuller - talvolta definta come ECM:

$$\begin{aligned} \Delta\pi_t &= c + d_1 \Delta\pi_{t-1} + d_2 \Delta\pi_{t-2} + d_3 \Delta\pi_{t-3} + d_4 \Delta\pi_{t-4} + \gamma \pi_{t-1} + \\ &g_1 \Delta u_{t-1} + g_2 \Delta u_{t-2} + g_3 \Delta u_{t-3} + g_4 \Delta u_{t-4} + \alpha u_{t-1} + \varepsilon_t \end{aligned} \quad (7)$$

Si noti che la riparametrizzazione del modello ARDL sub (4) e da noi stimato in precedenza nel modello ECM sub (7) e stimato qui sotto permette una stima diretta dei parametri di interesse  $\gamma$  e  $\alpha$ . Sul fatto che la (7) sia una pura riparametrizzazione della (4) non ci sono dubbi: la stima degli errori (cioè i residui) nei due modelli è la stessa (come si può vedere confrontando le stime dei corrispondenti RMSE) e anche la stima delle costanti c è la stessa.

```
. reg d2.lp d2l(1/4).lp ld.lp dl(1/4).u l.u , robust
```

```
Regression with robust standard errors
```

	Number of obs =	162
	F( 10, 151) =	11.57
	Prob > F =	0.0000
	R-squared =	0.4127
	Root MSE =	.00332

D2.lp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lp						
LD2	-.4235333	.095287	-4.44	0.000	-.6118013	-.2352654
L2D2	-.3219205	.1033272	-3.12	0.002	-.5260743	-.1177668
L3D2	-.0225638	.0798013	-0.28	0.778	-.1802351	.1351075

	L4D2	-.1194309	.0811984	-1.47	0.143	-.2798626	.0410009
	<b>LD</b>	<b>.0241396</b>	<b>.0486887</b>	<b>0.50</b>	<b>0.621</b>	<b>-.0720595</b>	<b>.1203387</b>
ur	LD	-.6812425	.1114778	-6.11	0.000	-.9015002	-.4609848
	L2D	.2875202	.1331926	2.16	0.032	.0243583	.550682
	L3D	.1342213	.1648241	0.81	0.417	-.191438	.4598806
	L4D	-.3518809	.13447	-2.62	0.010	-.6175666	-.0861953
	<b>L1</b>	<b>-.0477596</b>	<b>.0196094</b>	<b>-2.44</b>	<b>0.016</b>	<b>-.0865037</b>	<b>-.0090154</b>
<b>_cons</b>		<b>.0025947</b>	<b>.0011576</b>	<b>2.24</b>	<b>0.026</b>	<b>.0003076</b>	<b>.0048817</b>

Evidentemente, essendo un pura riparametrizzazione non avrebbe senso andare a rifare i test di scorretta specificazione del modello ECM perché, basandosi questi sugli stessi residui dell'ARDL(5, 5), porterebbero a risultati analoghi a quelli analizzati prima nel caso dell'ARDL(5, 5). Sempre per lo stesso motivo, abbiamo preferito usare le stime degli standard error robuste al problema (notato sopra) dell'eteroschedasticità.

Allo scopo di gettare un ponte fra i risultati empirici qui sopra con le determinanti teoriche discusse in precedenza bisogna fare inferenza sui parametri di interesse  $c$ ,  $\gamma$  e  $\alpha$  (in grassetto qui sopra). Ma prima di effettuare appropriate analisi di significatività vale la pena di ricordare che lo stimatore OLS dei parametri di un modello che include variabili esplicative irrilevanti è consistente ma inefficiente.

Pertanto, la precedente (sovra)dimensione parametrica viene ordinatamente ridotta con l'ausilio di una batteria di test di significatività allo scopo di eliminare dal modello regressori i cui parametri ( $d_i$  e/o  $g_i$ ) non risultano essere significativamente diversi da zero (nel nostro caso, le stime di  $d_3$ ,  $d_4$  e  $g_3$ ). L'ordine dell'eliminazione va dal più al meno insignificante (in senso statistico):

```
. test l3d2.lp, accum

( 1)  L3D2.lp = 0

      F( 1, 151) = 0.08
      Prob > F = 0.7778

. test l3d.u, accum

( 1)  L3D2.lp = 0
( 2)  L3D.ur = 0

      F( 2, 151) = 0.35
      Prob > F = 0.7019

. test l4d2.lp, accum

( 1)  L3D2.lp = 0
( 2)  L3D.ur = 0
( 3)  L4D2.lp = 0

      F( 3, 151) = 0.82
      Prob > F = 0.4868
```

Ristimo il modello ARDL(5,5) che incorpora le precedenti restrizioni accettate dai dati:

```
. reg d2.lp d2l(1/2).lp ld.lp dl(1/2).u l4d.u l.u

-----+-----
Source |          SS          df          MS          Number of obs =      163
-----+-----
-----+-----
F( 7, 155) = 14.92
```

Model		.001153684	7	.000164812	Prob > F	=	0.0000
Residual		.001712557	155	.000011049	R-squared	=	0.4025
-----							
Total		.002866241	162	.000017693	Adj R-squared	=	0.3755
					Root MSE	=	.00332

D2.lp		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lp							
	LD2	-.4097389	.0787976	-5.20	0.000	-.5653947	-.2540832
	L2D2	-.3031138	.075551	-4.01	0.000	-.4523564	-.1538713
	<b>LD</b> /	<b>.0105182</b>	<b>.0398978</b>	<b>0.26</b>	<b>0.792</b>	<b>-.0682954</b>	<b>.0893318</b>
ur							
	LD	-.659705	.107155	-6.16	0.000	-.8713777	-.4480323
	L2D	.3271787	.1178304	2.78	0.006	.094418	.5599394
	L4D	-.2351129	.088026	-2.67	0.008	-.4089984	-.0612275
	<b>L1</b> /	<b>-.0405048</b>	<b>.0197228</b>	<b>-2.05</b>	<b>0.042</b>	<b>-.0794651</b>	<b>-.0015446</b>
<b>_cons</b>	<b>/</b>	<b>.0022884</b>	<b>.0012045</b>	<b>1.90</b>	<b>0.059</b>	<b>-.000091</b>	<b>.0046679</b>

Il modello stimato qui sopra incorpora restrizioni non significative e, quindi, ci si attende che la batteria di test di scorretta specificazione non segnali nulla di preoccupante. I risultati qui sotto sono in linea con quanto atteso. Vale inoltre la pena di sottolineare il miglioramento della performance dei test di eteroschedasticità e Ramsey RESET ora di poco significativi secondo il tradizionale 5% e non più significativi all'1%.

```
. ovtest
```

```
Ramsey RESET test using powers of the fitted values of D2.lp
Ho: model has no omitted variables
      F(3, 152) =      2.86
      Prob > F =      0.0387
```

```
. hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of D2.lp

      chi2(1)      =      4.07
      Prob > chi2  =      0.0436
```

```
. dwstat
```

```
Durbin-Watson d-statistic( 8, 163) = 1.993439
```

```
. predict resECM, resid
(5 missing values generated)
```

```
. corrgram resECM, lags(20)
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]		[Partial Autocor]			
1	-0.0022	-0.0022	.00079	0.9775						
2	-0.0153	-0.0155	.03981	0.9803						
3	-0.0196	-0.0197	.10468	0.9913						
4	-0.0841	-0.0850	1.302	0.8610						
5	0.0529	0.0538	1.7779	0.8789						
6	0.0491	0.0475	2.1917	0.9013						
7	0.0338	0.0309	2.3885	0.9353						

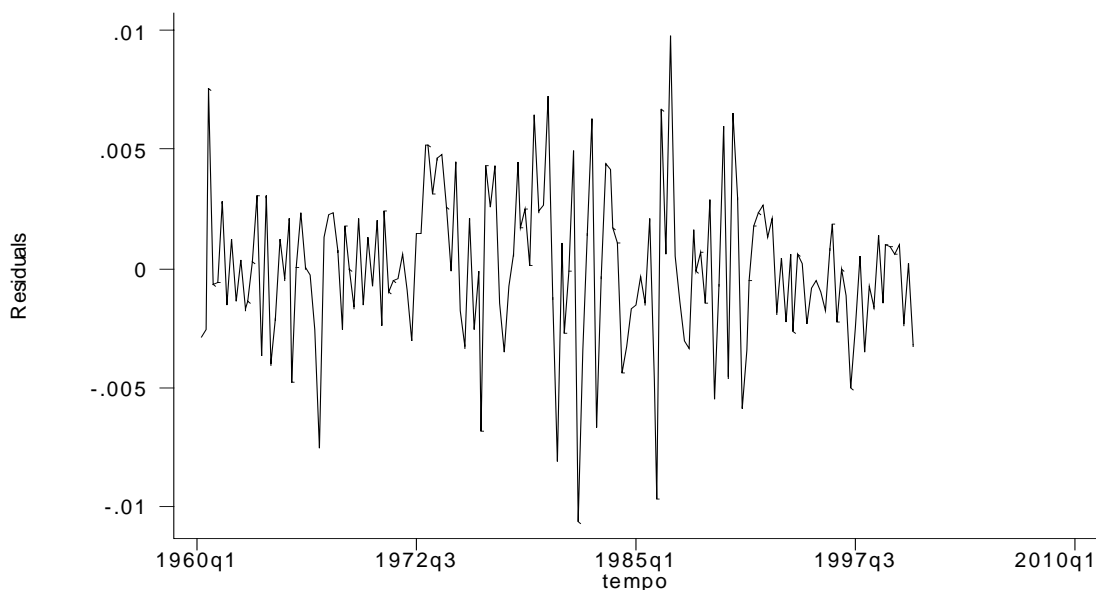
8	-0.1474	-0.1545	6.1597	0.6294	-	-
9	-0.0332	-0.0208	6.3518	0.7043	-	-
10	-0.0492	-0.0456	6.778	0.7462	-	-
11	-0.0236	-0.0307	6.8767	0.8090	-	-
12	-0.0806	-0.1220	8.0333	0.7825	-	-
13	0.1547	0.1739	12.322	0.5014	-	-
14	0.0940	0.1089	13.916	0.4560	-	-
15	-0.0541	-0.0538	14.449	0.4918	-	-
16	0.0838	0.0701	15.734	0.4717	-	-
17	0.0082	0.0603	15.746	0.5419	-	-
18	0.0557	0.0559	16.321	0.5702	-	-
19	-0.1114	-0.1894	18.639	0.4802	-	-
20	0.0446	0.0222	19.012	0.5210	-	-

```
. sktest resECM
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
resECM	0.457	0.039	4.87	0.0875

```
. graph7 resECM tempo, c(1) s(.) xlabel ylabel
```



La dispersione dei residui tende ad essere maggiore nel periodo 1973-1986 che abbraccia i due shock (1973, 1979) e il controshock (1986) petroliferi e questo fatto può avere causato quel segnale (non preoccupante) di eteroschedasticità nei residui e sul test RESET. Dal test di Durbin-Watson di autocorrelazione del primo ordine dei residui non emerge nulla di preoccupante. Inoltre dal correlogramma si evince che i residui non presentano autocorrelazioni significative e, quindi il modello proposto è in grado di dare conto della struttura correlativa dei dati di inflazione.

La performance dei diagnostici di scorretta specificazione sui residui è tale da suggerire di usare questo modello ECM vincolato per analizzare la significatività statistica dei parametri di interesse  $c$ ,  $\gamma$  e  $\alpha$  (in grassetto corsivo qui sopra) e, per questa via, ricollegarci al modello teorico di inflazione più appropriato per il caso da noi esaminato.

Il primo elemento evidente è la non significatività rispetto a zero del parametro  $\gamma$  (ovvero rispetto a uno del parametro  $\beta$ ). Tale risultato è talmente evidente da prescindere da ulteriori considerazioni

sulla natura di tale verifica in un contesto di variabili generate da processi stocastici non stazionari (la teoria della cointegrazione, infatti, va oltre gli obiettivi delle presenti lezioni). Questo fatto ( $\beta=1$ ), alla luce della disamina teorica qui sopra, implica l'assenza di un legame di lungo periodo fra tasso di inflazione e tasso di disoccupazione. Il trade-off delineato dalla teoria sottostante la c.d. curva di Phillips non esiste per il caso statunitense. Impongo anche quest'ultima restrizione al modello precedente ed ottengo la stima:

```
. reg d2.lp d2l(1/2).lp dl(1/2).u l4d.u l.u
```

Source	SS	df	MS	Number of obs = 163		
Model	.001152916	6	.000192153	F( 6, 156)	=	17.50
Residual	.001713325	156	.000010983	Prob > F	=	0.0000
-----				R-squared	=	0.4022
Total	.002866241	162	.000017693	Adj R-squared	=	0.3792
-----				Root MSE	=	.00331

D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lp						
LD2	-.4005545	.0704667	-5.68	0.000	-.5397466	-.2613625
L2D2	-.2973517	.0721044	-4.12	0.000	-.4397786	-.1549248
ur						
LD	-.6522234	.1030201	-6.33	0.000	-.8557178	-.4487291
L2D	.3292217	.1172242	2.81	0.006	.0976702	.5607731
L4D	-.2359843	.0877012	-2.69	0.008	-.4092194	-.0627492
L1	-.0394728	.0192727	-2.05	0.042	-.077542	-.0014037
_cons	.0023418	.0011839	1.98	0.050	3.35e-06	.0046803

Se ci si concentra sulla stima del parametro  $\alpha$  abbiamo che essa è significativamente negativa. Mettendo assieme questi due pezzi di informazione ( $\beta=1$  e  $\alpha<0$ ) delineiamo il caso di un modello per l'inflazione governato dalla teoria del NAIRU. Alla luce di quanto visto, il NAIRU può essere stimato dalla formula:  $u^* = -c/a_1$ .che, tradotta in un comando Stata fornisce il seguente risultato:

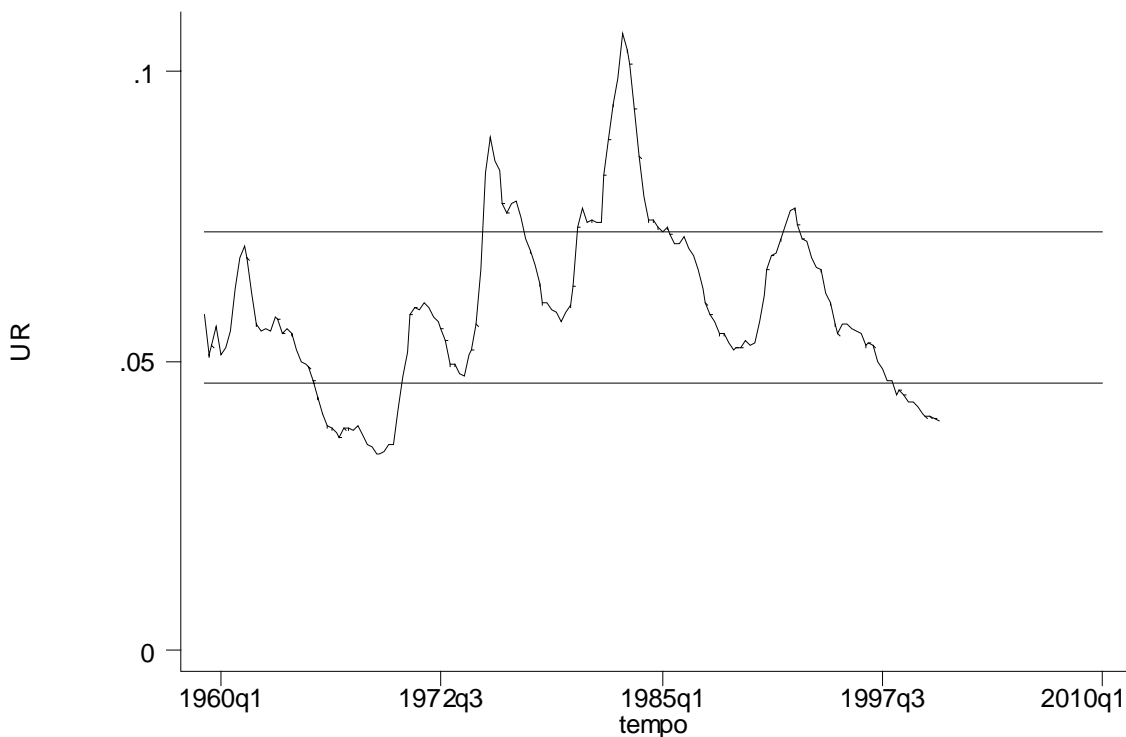
```
. nlcom (NAIRU:-_b[_cons]/_b[l.u])
```

```
NAIRU:  -_b[_cons]/_b[l.u]
```

D2.lp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
NAIRU	.059327	.0065882	9.00	0.000	.0463133	.0723407

Quando il tasso di disoccupazione è inferiore al 5,9% allora si delinea una situazione di eccesso di domanda, con conseguenti spinte inflazionistiche. D'altro canto, bisogna ricordare che la stima proposta per il NAIRU è una combinazione di stime statistiche e, come tale, è soggetta ad un intervallo di confidenza. Qui sotto si riproduce il grafico dell'andamento storico della disoccupazione e due rette orizzontali che delimitano l'intervallo di confidenza al 95% del NAIRU (questi ultimi due dati sono stati ottenuti dalla tabella qui sopra). Evidentemente, i periodi in cui il tasso di disoccupazione storico si trova sopra alla retta più in alto corrispondono a fasi storiche in cui il tasso di inflazione USA ha subito spinte significative verso il basso a causa di una carenza di domanda. Viceversa, spinte inflazionistiche da eccesso di domanda paiono essersi verificate nella seconda metà degli anni sessanta e novanta.

```
graph7 u tempo, yline(.0463133 .0723407) xlabel ylabel c(l) s(.
```



Dal confronto dei risultati di cui sopra con quelli dell'analisi univariata emerge che il modello che spiega le variazioni dell'inflazione utilizzando anche l'informazione del tasso di disoccupazione migliora significativamente la performance del semplice modello AR(3): l'indice  $R^2$  raddoppia, in quanto sale al 40% a partire dal poco meno del 20% dell'AR(3). Grazie all'informazione aggiuntiva fornita dalla teoria economica, la nostra capacità esplicativa del comportamento dell'inflazione negli USA è aumentata considerevolmente.

Il modello ARDL, che incorpora una più articolata struttura causale, implica però che è possibile fare previsioni non condizionali solo un passo avanti in quanto, per prevedere su un orizzonte ( $h$ ) più lungo di un trimestre, è necessario fornire al modello uno scenario per il tasso di disoccupazione e, quindi, per  $h > 1$  il modello ARDL produce solo previsioni condizionali.

Anche se non approfondito in questa sede, si ricordi che il modello di NAIRU da noi proposto genera una stima di  $u^*$  che è *definitivamente* costante nel tempo, mentre nella realtà economica potrebbe essere che, nel tempo, il valore del NAIRU cambia a causa di cambiamenti strutturali dell'economia e delle istituzioni. Pertanto, un'utile analisi aggiuntiva potrebbe essere quella della costanza nel tempo dei parametri del modello ARDL vincolato che sta a monte della nostra stima del NAIRU.

Le successive verifiche che hanno fatto emergere un modello di NAIRU e che, al contempo, hanno scartato il modello di Phillips sono basate su test statistici di significatività interpretati alla luce delle teorie economiche. In questo il modello ARDL/ECM è l'anello di congiunzione fra le teorie economiche e un approccio puramente statistico e, come tale, ateorico. Il modello di NAIRU emerso nel caso degli USA non è detto abbia valenza interpretativa dei casi relativi ad altri paesi.

## 5. Rischi di regressioni spurie e rimedi forniti dalla specificazione dinamica

Lo scopo di questo paragrafo è illustrare i forti rischi che si corrono quando, senza dare conto della stazionarietà delle serie storiche di interesse, si stima un modello statico, cioè un modello che non specifica in modo appropriato i legami fra le serie di interesse. In pratica andiamo a vedere cosa può accadere se un ricercatore è tanto sprovvisto da utilizzare tecniche econometriche con dati di serie storiche, trattandoli come se avesse a che fare con dati di cross section.

Esiste un legame fra consumo e reddito? Molte teorie economiche rispondono affermativamente, ma ovviamente ciò che ci si aspetta è che il consumo di un certo paese (nel nostro caso gli USA) reagisca alle fluttuazioni del reddito dello *stesso* paese, e non di paesi *diversi*.

Come prima cosa, trasportiamo in Stata i dati della cartella "annuale" contenuta nel solito foglio Excel "datiTS.xls". La cosa è già stata richiesta come esercizio. Il contenuto del foglio sono dati annuali a prezzi costanti per il periodo 1960-2001 per i consumi USA (cus) e il reddito del Regno Unito (yuk). Il file "conred.dta" contiene i dati già in formato Stata.

Si noti che il comando risulta semplificato rispetto a ciò che abbiamo visto per dati a maggiore frequenza (mensili e trimestrali). Informiamo Stata della variabile "anno" unico indicatore della dimensione temporale dei dati e calcoliamo la trasformata logaritmica di tutte le variabili (stimeremo così le elasticità del consumo al reddito).

```
tsset anno
  time variable:  anno, 1960 to 2001

g lcus=log(cus)
g lyuk=log(yuk)
```

Supponiamo di stimare un modello statico in cui, i consumi in USA dipendono dal reddito nel Regno Unito. Evidentemente, a priori, non ci si attende una relazione significativa.

Il risultato è stupefacente: la relazione è molto significativa (statistica t altissima) e l'indice  $R^2$  sfiora il suo valore massimo (supera abbondantemente il 99%). Unico neo: un test di Durbin-Watson che segnala forte autocorrelazione di primo ordine dei residui della regressione.

Cosa è successo? Forse è necessario introdurre una nuova teoria sul commercio internazionale, oppure basta fare riferimento a modelli storici sulle origini degli USA? Niente di tutto questo, come vedremo fra poco.

```
reg lcus lyuk
```

Source	SS	df	MS			
Model	6.98900429	1	6.98900429	Number of obs =	42	
Residual	.038130775	40	.000953269	F( 1, 40) =	7331.62	
				Prob > F =	0.0000	
				R-squared =	<b>0.9946</b>	
				Adj R-squared =	0.9944	
Total	7.02713506	41	.171393538	Root MSE =	.03088	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lyuk	1.459444	.0170446	<b>85.62</b>	0.000	1.424996	1.493893
_cons	-11.13127	.2243859	-49.61	0.000	-11.58477	-10.67777

```
dwstat
```

Durbin-Watson d-statistic( 2, 42) = .4183794

Il problema è che le variabili analizzate sono generate da processi stocastici non stazionari e, allo stesso tempo, il ricercatore compie anche l'errore di stimare modelli statici. Spesso, nei lavori empirici, si utilizzano variabili che presentano forti trend e che, quindi, evidenziano una forte non stazionarietà in media. Le regressioni che ne risultano sono definite spurie e mostrano quel fenomeno che Clive Granger e Paul Newbold descrivono in una frase del loro lavoro della metà degli anni '70 (Cfr. Granger, Newbold, 1974 p. 111): "... E' [molto] comune vedere lavori empirici caratterizzati da una apparenterete elevata capacità esplicativa, misurata da alti valori della statistica  $R^2$ , cui però si accompagnano valori della statistica di Durbin-Watson estremamente bassi. E' anche molto strano che nonostante tutti i libri di testo di econometria sottolineino esplicitamente i danni causati dall'autocorrelazione dei residui, tale fenomeno compare [molto di frequente] in rispettabili lavori applicati ..." (si suggerisce una rilettura venti anni dopo di questa frase, comunque attuale, semplicemente omettendo, come aggiornamento, il testo contenuto tra parentesi quadre).

Un valore della statistica Durbin-Watson sensibilmente inferiore al valore dell'  $R^2$  viene interpretato, in un senso più ampio rispetto a quello di regressione spuria, come un trattamento inadeguato della non stazionarietà e della dinamica delle serie storiche nella regressione.

Nella nostra regressione abbiamo che i residui sono fortemente autocorrelati e questo implica la convinzione che il modello statico proposto non sia in grado di spiegare la dinamica dei dati. Il test di Durbin-Watson è il primo sintomo, il correlogramma (più sotto) lo conferma, così come altri test di scorretta specificazione:

```

predict res, resid

. sktest res

                Skewness/Kurtosis tests for Normality
-----+----- joint -----
Variable | Pr(Skewness)  Pr(Kurtosis)  adj chi2(2)  Prob>chi2
-----+-----
res      |      0.096      0.765          3.05         0.2176

. hettest

Cook-Weisberg test for heteroskedasticity using fitted values of lcus
Ho: Constant variance
chi2(1)      =      0.08
Prob > chi2  =      0.7743

. ovtest

Ramsey RESET test using powers of the fitted values of lcus
Ho: model has no omitted variables
F(3, 37)    =     11.91
Prob > F    =      0.0000

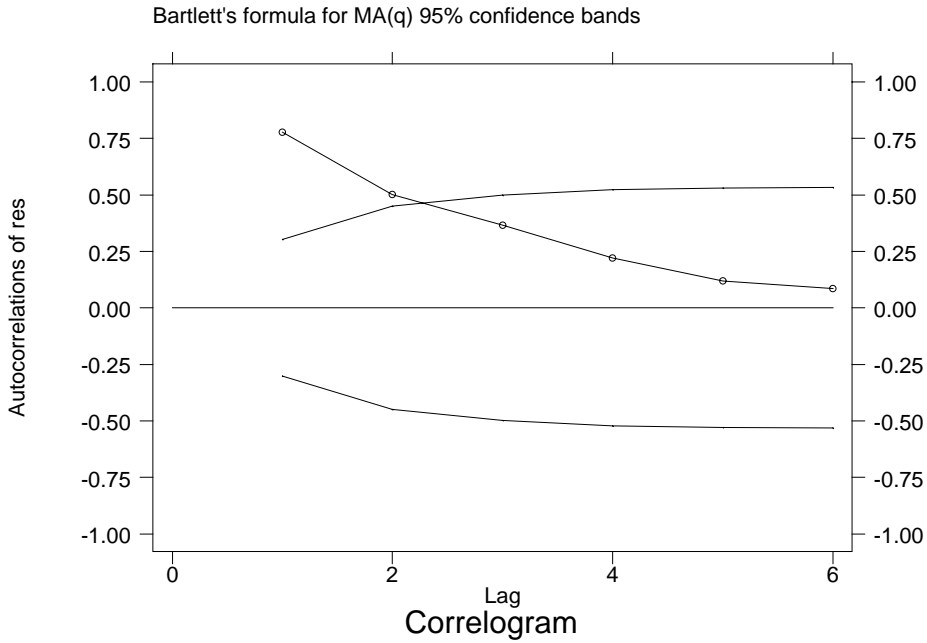
corrgram res, lag(6)

LAG      AC      PAC      Q      Prob>Q      -1      0      1 -1      0      1
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
1      0.7775  0.7926  27.248  0.0000      |-----|-----|
2      0.5016 -0.2562  38.872  0.0000      |----|-----|

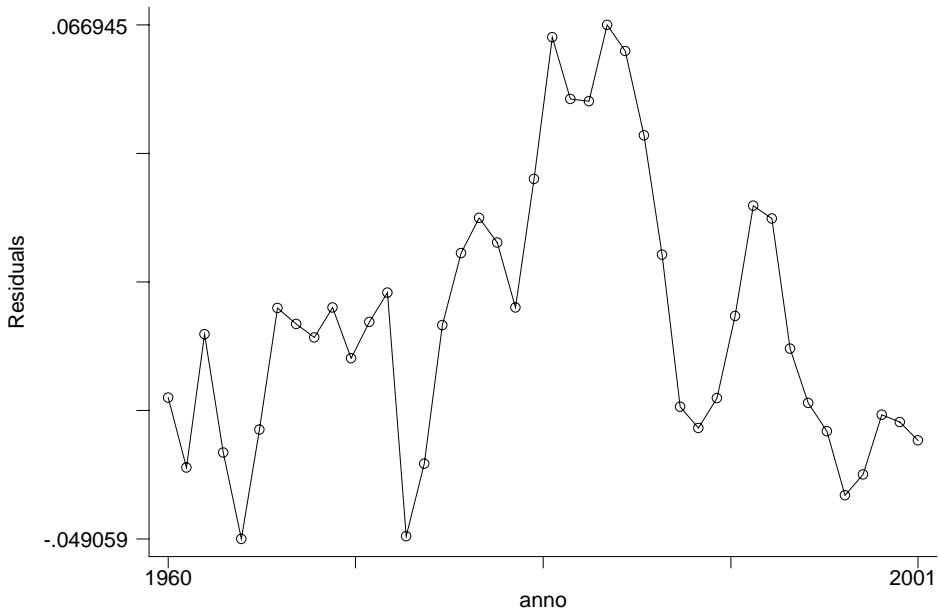
```

3	0.3660	0.2121	45.22	0.0000	--	-
4	0.2202	-0.2384	47.578	0.0000	-	-
5	0.1199	0.1959	48.296	0.0000		-
6	0.0845	-0.1097	48.662	0.0000		

. ac res, lags(6)



. graph res anno, c(1)



dfuller res, regress lags(0)

Dickey-Fuller test for unit root				Number of obs =	41
	Test Statistic	----- Interpolated Dickey-Fuller -----	1% Critical Value	5% Critical Value	10% Critical Value
-----	Z(t)	-2.092	-3.641	-2.955	-2.611

-----  
 \* MacKinnon approximate p-value for Z(t) = 0.2478  
 -----

D.res		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
res	L1	-.2074486	.099168	-2.09	0.043	-.4080348	-.0068624
_cons		-.0001006	.0029956	-0.03	0.973	-.0061596	.0059585

Considerazioni:

La variabile dipendente è molto probabilmente un processo non stazionario, la regressione non ha senso economico, i residui sono non stazionari. La caduta delle ipotesi a monte del CLRM implica scorretta specificazione del modello e stime distorte del parametro di interesse. In altri termini, la relazione che apparentemente è significativa tra consumo inglese e reddito americano è spuria (destituita di senso economico). Infatti, ciò che vede il modello è il trend comune e crescente di entrambe le variabili; a questo trend comune il modello attribuisce erroneamente il significato di legame strutturale (elasticità) tra consumo e reddito, tutto da dimostrare. Anche l'inquinamento in Italia ha un trend crescente e per questo risulterebbe molto probabilmente positivamente correlato con il consumo americano; ma se qui da noi si inquinasse di meno, gli USA consumerebbero meno?

Quindi, un ricercatore accorto, anzichè pensare di introdurre nuove spiegazioni teoriche ex post per spiegare l'arcano, reagisce specificando un modello dinamico di tipo ARDL allo scopo di valutare il contributo di tutte le esplicative (test di significatività) solo dopo che la specificazione di riferimento non presenta problemi nei diagnostici. Ad esempio, questo modello ARDL fornisce buoni risultati in termini di diagnostici dei residui:

```
reg d.lcus ld.lcus ld.lyuk l.lcus l.lyuk
```

Source	SS	df	MS	Number of obs = 40		
Model	.002314119	4	.00057853	F( 4, 35) =	2.19	
Residual	.009224876	35	.000263568	Prob > F =	0.0897	
				R-squared =	0.2005	
				Adj R-squared =	0.1092	
Total	.011538996	39	.000295872	Root MSE =	.01623	

D.lcus		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lcus	LD	.4620318	.2238897	2.06	0.047	.0075115	.916552
lyuk	LD	-.1666111	.2035557	<b>-0.82</b>	0.419	-.5798511	.246629
lcus	L1	.010556	.0895768	<b>0.12</b>	0.907	-.1712944	.1924065
lyuk	L1	-.0287854	.1317032	<b>-0.22</b>	0.828	-.2961572	.2385864
_cons		.3167983	1.013507	<b>0.31</b>	0.756	-1.74073	2.374326

```
dwstat
```

```
Durbin-Watson d-statistic( 5, 40) = 1.768839
```

```
predict res1, resid  

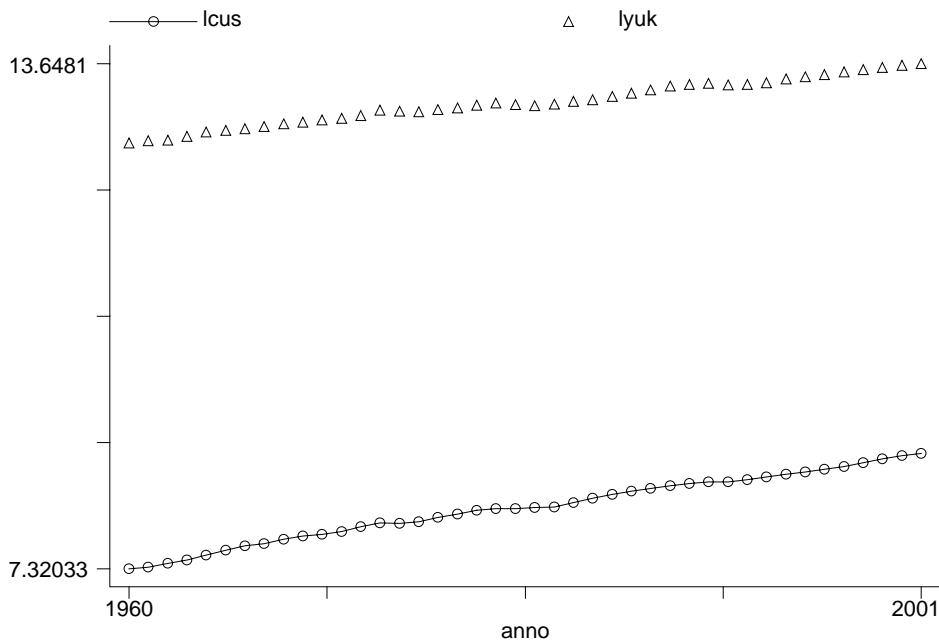
(2 missing values generated)
```

```
corrgram res1, lag(6)
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.1028	0.1037	.45554	0.4997						
2	-0.1674	-0.1845	1.6939	0.4287	-			-		
3	-0.1244	-0.0834	2.3961	0.4944						
4	0.0264	0.0117	2.4287	0.6574						
5	-0.0490	-0.1156	2.5437	0.7699						
6	-0.0722	-0.0877	2.8011	0.8334						

NOTA: Il test di significatività riferiti alla variabile reddito del Regno Unito non rifiutano la nulla.

Per approfondire le precedenti tematiche si effettua l'ispezione del grafico delle serie, del loro correlogramma e dei test di Dickey-Fuller. Ne emerge inequivocabilmente che sia il consumo USA sia il reddito nel Regno Unito sono generati da processi non stazionari e, quindi, lo strano risultato di regressione col modello statico ricade nel classico caso di regressione spuria, come paventato da Granger e Newbold.



```
dfuller lcus, regress lags(2) trend
```

Augmented Dickey-Fuller test for unit root                      Number of obs    =            39

Test                      ----- Interpolated Dickey-Fuller -----  
                             1% Critical                      5% Critical                      10% Critical

	Statistic	Value	Value	Value
Z(t)	-2.910	-4.251	-3.544	-3.206

\* MacKinnon approximate p-value for Z(t) = 0.1590

D.lcus		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lcus	L1	-.2231735	.0766881	-2.91	0.006	-.3790225 - .0673245
	LD	.4220852	.1470002	2.87	0.007	.1233449 .7208254
	L2D	-.1248624	.1532175	-0.81	0.421	-.4362379 .186513
_trend		.0071471	.0025533	2.80	0.008	.0019581 .0123361
_cons		1.674537	.5636851	2.97	0.005	.5289912 2.820083

. dfuller lcus, regress lags(1) trend

Augmented Dickey-Fuller test for unit root                      Number of obs =                      40

Test Statistic	Interpolated Dickey-Fuller		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-3.502	-4.242	-3.540

\* MacKinnon approximate p-value for Z(t) = 0.0392

D.lcus		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lcus	L1	-.2283193	.0651965	-3.50	0.001	-.3605441 - .0960946
	LD	.3922436	.134224	2.92	0.006	.1200248 .6644625
_trend		.0073701	.0021897	3.37	0.002	.0029292 .0118111
_cons		1.707812	.4792647	3.56	0.001	.7358182 2.679806

. dfuller lyuk, regress lags(2) trend

Augmented Dickey-Fuller test for unit root                      Number of obs =                      39

Test Statistic	Interpolated Dickey-Fuller		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-3.080	-4.251	-3.544

\* MacKinnon approximate p-value for Z(t) = 0.1111

D.lyuk		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lyuk	L1	-.3380181	.1097509	-3.08	0.004	-.5610588 - .1149774
	LD	.4716745	.1457791	3.24	0.003	.1754159 .7679332
	L2D	-.0929942	.1607777	-0.58	0.567	-.4197337 .2337454
_trend		.0075177	.0024969	3.01	0.005	.0024433 .0125921
_cons		4.303135	1.388765	3.10	0.004	1.480825 7.125446

. dfuller lyuk, regress lags(1) trend

Augmented Dickey-Fuller test for unit root                      Number of obs =                      40

Test Statistic	----- Interpolated Dickey-Fuller -----		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-3.398	-4.242	-3.540

\* MacKinnon approximate p-value for Z(t) = 0.0517

D.lyuk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lyuk						
L1	-.3278655	.0964974	-3.40	0.002	-.5235713	-.1321596
LD	.4418804	.1459023	3.03	0.005	.1459769	.737784
_trend	.0074004	.0022084	3.35	0.002	.0029216	.0118791
_cons	4.169769	1.221822	3.41	0.002	1.691798	6.64774

Esiste un altro modello statico che evita problemi di non stazionarietà delle variabili; tuttavia, si tratta non di un modello di consumo che stima l'elasticità di lungo periodo del consumo al reddito, bensì di un modello congiunturale di indicatori che stima una relazione tra tassi di crescita del consumo e del reddito dettata probabilmente dal ciclo economico internazionale (eventuale andamento congiunturale in fase nei due paesi). Ma non è ben chiaro a cosa possa servire tale modello e perchè il consumo US debba essere la variabile dipendente e non l'esplicativa. Inoltre, si tacciano i potenziali problemi di regressori endogeni....

. reg d.lcus d.lyuk

Source	SS	df	MS	Number of obs =	41
Model	.005936889	1	.005936889	F( 1, 39) =	39.68
Residual	.005834962	39	.000149614	Prob > F =	0.0000
Total	.011771852	40	.000294296	R-squared =	0.5043
				Adj R-squared =	0.4916
				Root MSE =	.01223

D.lcus	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lyuk						
D1	.6330723	.1004987	6.30	0.000	.4297944	.8363501
_cons	.0197865	.0030949	6.39	0.000	.0135264	.0260466

. predict res1, res  
(1 missing value generated)

. sktest res1

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
res1	0.256	0.165	3.45	0.1781

. hetttest

Cook-Weisberg test for heteroskedasticity using fitted values of D.lcus  
Ho: Constant variance

```

chi2(1)      =      0.00
Prob > chi2  =      0.9529

```

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of D.lcus

Ho: model has no omitted variables

```

F(3, 36) =      2.05
Prob > F =      0.1242

```

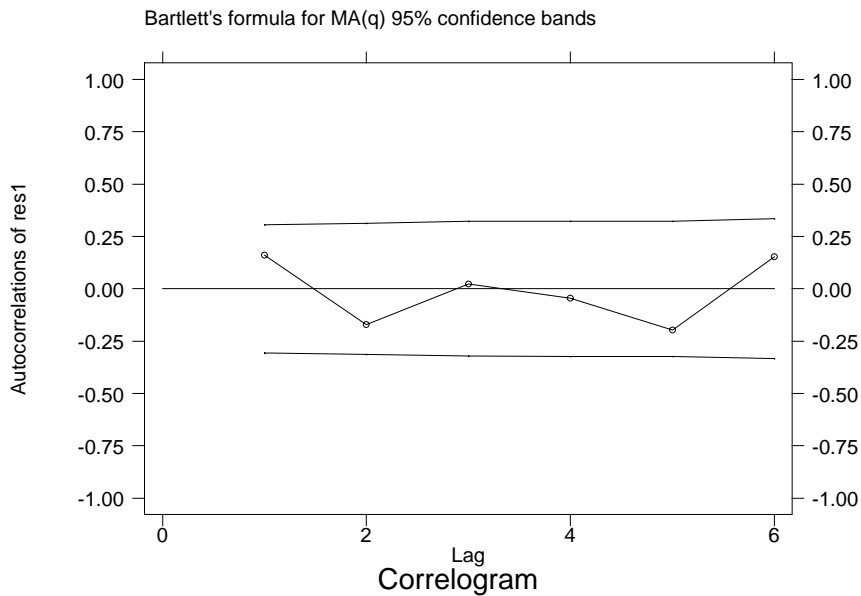
```
. dwstat
```

Durbin-Watson d-statistic( 2, 41) = 1.628796

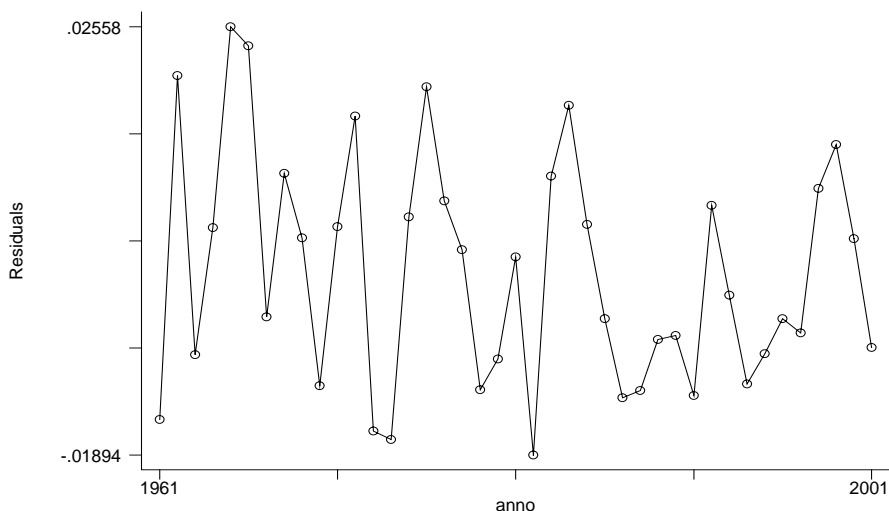
```
. corrgram res1, lags(6)
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.1604	0.1616	1.1344	0.2868						
2	-0.1703	-0.2165	2.4455	0.2944		-			-	
3	0.0235	0.1373	2.4712	0.4805						
4	-0.0446	-0.1856	2.5659	0.6329						
5	-0.1975	-0.1476	4.4759	0.4831		-			-	
6	0.1526	0.2116	5.6487	0.4637						

```
. ac res1, lags(6)
```



```
. graph res1 anno, c(1)
```



NOTA: i residui non presentano più problemi di non stazionarietà, anche se l'assenza della dinamica comporta problemi di autocorrelazione.

Per tornare ad un modello di consumo occorre: reintrodurre l'informazione sui livelli, risolvere il problema della non stazionarietà e dell'autocorrelazione dei residui.

Si noti che la sola aggiunta di dinamica evidenzia che la variabile dipendente ritardata ha un valore molto simile all'unità e che la somma dei parametri del reddito è molto vicina a zero. Entrambi questi risultati portano alla non esistenza di una relazione di lungo periodo significativa tra consumo e reddito.

```
. reg lcus lyuk l.lyuk l.lcus
```

Source	SS	df	MS	Number of obs =	41
Model	6.43464582	3	2.14488194	F( 3, 37) =	14863.51
Residual	.005339293	37	.000144305	Prob > F =	0.0000
Total	6.43998511	40	.160999628	R-squared =	0.9992
				Adj R-squared =	0.9991
				Root MSE =	.01201

lcus		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lyuk	--	.650921	.1027029	6.34	0.000	.4428252 .8590169
	L1	-.5559478	.1195151	-4.65	0.000	-.7981085 -.3137871
lcus	L1	.9282417	.0644627	14.40	0.000	.7976278 1.058856
_cons		-.6510963	.7297171	-0.89	0.378	-2.129643 .8274509

```
. predict res2, res
(1 missing value generated)
```

```
. sktest res2
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
res2	0.867	0.082	3.26	0.1956

. hettest

Cook-Weisberg test for heteroskedasticity using fitted values of lcus

Ho: Constant variance

chi2(1) = 3.41  
Prob > chi2 = 0.0649

. ovtest

Ramsey RESET test using powers of the fitted values of lcus

Ho: model has no omitted variables

F(3, 34) = 0.52  
Prob > F = 0.6737

. dwstat

Durbin-Watson d-statistic( 4, 41) = 1.692594

. list res2 anno

	res2	anno
1.	.	1960
2.	-.0213867	1961
3.	.0137353	1962
4.	-.0136211	1963
5.	-.0019925	1964
6.	.0184968	1965
7.	.0186177	1966
8.	-.0074558	1967
9.	.0071104	1968
10.	.000947	1969
...	omissis	....
37.	-.0029021	1996
38.	-.0047137	1997
39.	.0096762	1998
40.	.0149882	1999
41.	.0062746	2000
42.	-.0046996	2001

Test di Godfrey

. replace res2=0 if anno==1960  
(1 real change made)

. reg res2 l.res2 lyuk l.lyuk l.lcus

Source	SS	df	MS	Number of obs =	41
Model	.000082906	4	.000020727	F( 4, 36) =	0.14
Residual	.005256387	36	.000146011	Prob > F =	0.9654
Total	.005339293	40	.000133482	R-squared =	0.0155
				Adj R-squared =	-0.0939
				Root MSE =	.01208

res2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
res2						
lyuk	L1	.1427131	.1893927	0.75	0.456	-.2413931 .5268192
	--	-.0278246	.109709	-0.25	0.801	-.2503247 .1946755
	L1	.049891	.1372458	0.36	0.718	-.2284564 .3282384
lcus	L1	-.0152612	.0679319	-0.22	0.824	-.1530335 .122511
_cons		-.1664939	.7665508	-0.22	0.829	-1.721131 1.388143

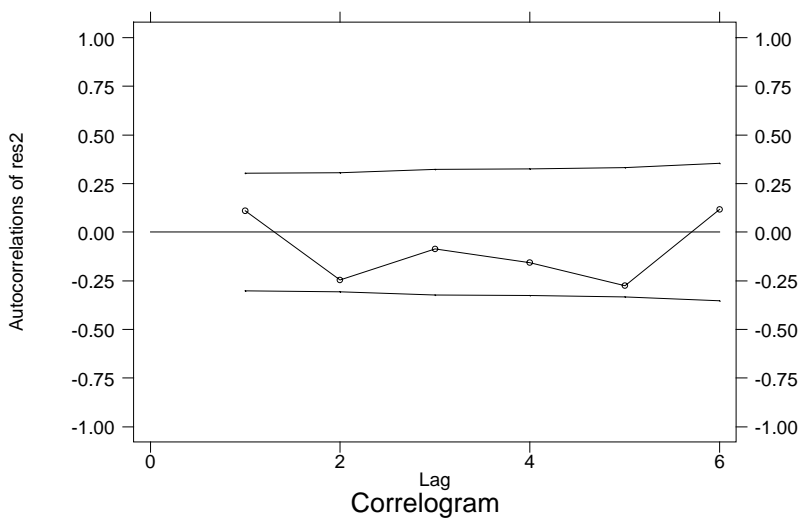
```
. display chi2tail(1,41*0.0155)
.42534503
```

```
. corrgram res2, lags(6)
```

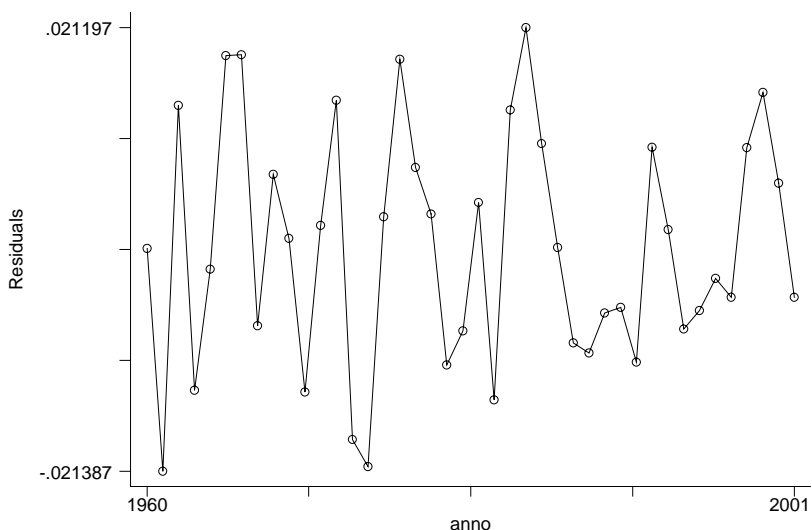
LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.1088	0.1093	.53357	0.4651						
2	-0.2459	-0.2647	3.3265	0.1895	-			--		
3	-0.0863	-0.0114	3.6798	0.2982						
4	-0.1577	-0.2689	4.8888	0.2989	-			--		
5	-0.2752	-0.2570	8.6723	0.1229	--			--		
6	0.1170	0.1309	9.3754	0.1535						-

```
. ac res2, lags(6)
```

Bartlett's formula for MA(q) 95% confidence bands



```
. graph res2 anno, c(1)
```



Il modello che abbiamo stimato è:  $lcus_t = \beta_0 + \beta_1 lyuk_t + \beta_2 lyuk_{t-1} + \gamma_1 lcus_{t-1} + \varepsilon_t$

Riparametrizzazione del modello per dare conto della non stazionarietà della dipendente:  $\pm lcus_{t-1}$

Riparametrizzazione del modello per dare conto di una dinamica più generale:  $\pm \beta_1 lyuk_{t-1}$

Si ottiene il modello a correzione dell'errore (ECM, Sargan 1964):<sup>5</sup>

$$l_{cus_t} - l_{cus_{t-1}} = \beta_0 + \beta_1(l_{yuk_t} - l_{yuk_{t-1}}) + (\beta_1 + \beta_2)l_{yuk_{t-1}} + (\gamma_1 - 1)l_{cus_{t-1}} + \varepsilon_t$$

$$\Delta l_{cus_t} = \beta_0 + \beta_1 \Delta l_{yuk_t} + \alpha [l_{cus_{t-1}} - \beta l_{yuk_{t-1}}] + \varepsilon_t$$

```
. reg d.lcus d.lyuk l.lcus l.lyuk
```

Source	SS	df	MS	Number of obs =	41
Model	.006432559	3	.002144186	F( 3, 37) =	14.86
Residual	.005339293	37	.000144305	Prob > F =	0.0000
				R-squared =	0.5464
				Adj R-squared =	0.5097
Total	.011771852	40	.000294296	Root MSE =	.01201

D.lcus		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lyuk	D1	.650921	.1027029	6.34	0.000	.4428252 .8590169
lcus	L1	-.0717583	.0644627	-1.11	0.273	-.2023722 .0588556
lyuk	L1	.0949732	.0947757	1.00	0.323	-.0970606 .2870071
_cons		-.6510963	.7297171	-0.89	0.378	-2.129643 .8274509

dove:

$\Delta l_{yuk_t}$  è il termine che coglie gli impulsi di breve periodo;

$[l_{cus_{t-1}} - \beta l_{yuk_{t-1}}]$  è il termine che rappresenta la deviazione dall'equilibrio di lungo periodo.

Il parametro  $\alpha = \gamma_1 - 1$  è detto loading parameter e misura la velocità ed il sentiero dell'aggiustamento. Coglie lo stesso effetto della riparametrizzazione per il test DF. Pertanto:

se  $-1 < \alpha < 0 \rightarrow 0 < \gamma_1 < 1$  ed il processo di aggiustamento verso l'equilibrio è stabile;

se  $\alpha = -1 \rightarrow \gamma_1 = 0$  e l'aggiustamento è istantaneo;

se  $-2 < \alpha < -1 \rightarrow -1 < \gamma_1 < 0$  ed il processo di aggiustamento verso l'equilibrio ha overshooting (tipico nei modelli ad informazione incompleta);

se  $\alpha = -2 \rightarrow \gamma_1 = -1$  ed il processo oscilla continuamente senza convergere;

se  $\alpha < -2 \rightarrow \gamma_1 < -1$  ed il processo ha overshooting esplosivo;

se  $\alpha > 0 \rightarrow \gamma_1 > 1$  ed il processo è esplosivo.

Il parametro  $\beta = (\beta_1 + \beta_2) / (1 - \gamma_1)$  è il moltiplicatore di lungo periodo e misura la risposta di equilibrio di lcus a lyuk.

L'equilibrio statico ( $\Delta l_{yuk_t} = \Delta l_{cus_t} = 0$ ) non stocastico ( $\varepsilon_t = 0$ ) è  $0 = \beta_0 + \alpha [l_{cus}^* - \beta l_{yuk}^*]$ , ossia:

$l_{cus}^* = -\beta_0 / \alpha + \beta l_{yuk}^*$ ; da cui emerge che se  $\alpha = 0 \rightarrow \gamma_1 = 1$  il modello non converge verso alcun livello ottimale di lcus e il modello ECM collassa nel modello in differenze.

Se lcus e lyuk sono processi I(1),  $\alpha = 0$  implica assenza di cointegrazione e, quindi, la stazionarietà è raggiunta differenziando i livelli delle variabili di interesse (e non anche combinandone i livelli nella relazione di cointegrazione).

<sup>5</sup> Anche il modello a pagina 44 è un ECM che utilizza nel breve periodo  $\Delta l_{yuk_{t-1}}$ , invece di  $\Delta l_{yuk_t}$ , ed aggiunge  $\Delta l_{cus_{t-1}}$ .