LA MODELLAZIONE EMPIRICA DELLE RELAZIONI ECONOMICHE: APPLICAZIONI IN STATA 7

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

L'econometria delle serie storiche

VIII LEZIONE: paragrafi 1-3.2

1. Alcuni concetti di base

Una serie storica è una successione di dati ordinati nel tempo: Y_t con t=1,2, … , T.

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 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 nella prossima lezione.

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: ΔP_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 ($infl_t$), 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), $infl_t$ può essere annualizzato moltiplicandolo per quattro ($inflA_t = 4 \times infl_t$): nel 1973q3 era circa l'8.3%. Un'altra misura di inflazione annuale è quella annua tendenziale, $inflAT_t$, 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, $inflA_t$ e $inflAT_t$ 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 <u>trasformazione logaritmica</u> 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):

$\Delta \log(P_t) = \log(P_t) - \log(P_{t-1}) \approx infl_t$

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 <u>**datiTS.xls**</u>. 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)  [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  trimes~e          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. = <u>operatore differenza prima</u>: `d.`*`nome-della-serie`* (ad esempio `d.Y` equivale a ΔY_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. Analisi univariata delle serie storiche

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 della media $E(Y_t) = \mu$ di una serie storica è:

$$\hat{\mu} = \frac{\sum_{t=1}^{T} y_t}{T}$$

Mentre i momenti secondi di ordine k (le autocovarianze di ordine k): $COV(Y_t, Y_{t-k}) = \gamma_k$ sono stimati con:[1]

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

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

$\hat{\rho}_k = \dfrac{\hat{\gamma}_k}{\hat{\gamma}_o}$ per definizione $\hat{\rho}_0 = 1$.

3.1 Tecniche esplorative per serie storiche

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

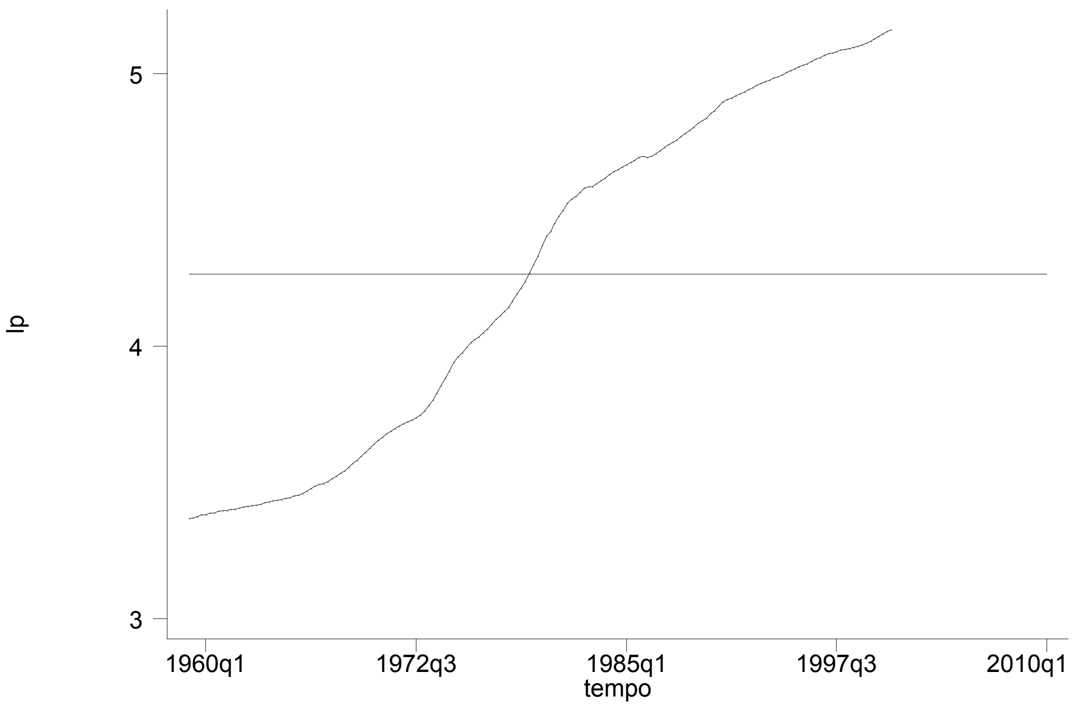
[1] In generale, $k \geq 0$; in particolare, quando k=0, l'autocovarianza di ordine zero è la $VAR(Y_t)$.

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.

```
graph lp tempo, c(l) 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.

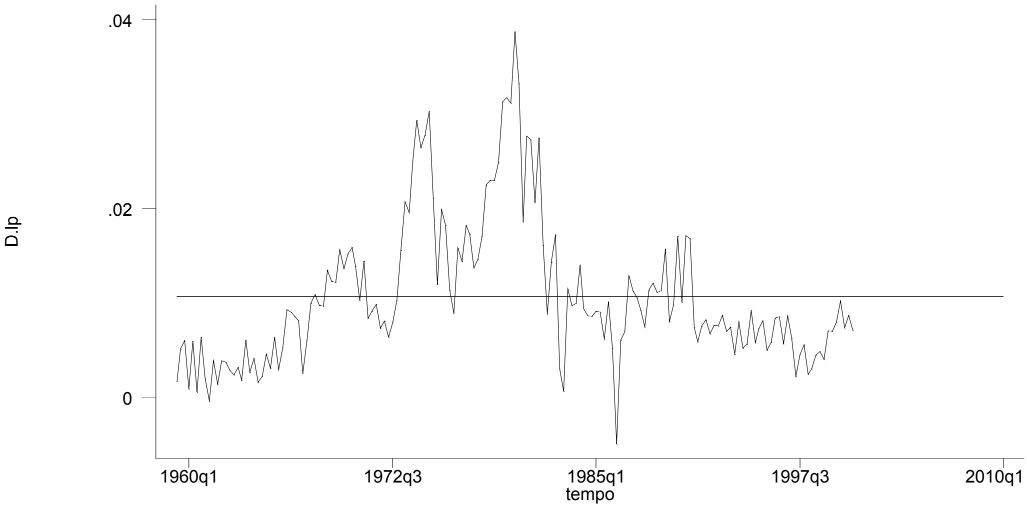
Si pensi all'esempio della serie storica W_t del peso di un individuo dalla nascita fino a 60 anni (T=60): le realizzazioni temporali di W_t non sono estratte dalla stessa distribuzione statistica perché il peso varia strutturalmente con l'età e lo sviluppo dell'individuo. Al contrario, in una cross-section 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 ($t=1,2, \ldots, 60$) 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: $dlp_t = \log(P_t)-\log(P_{t-1})$.

```
      summm d.lp

    Variable |      Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
lp           |
          D1 |      167    .0107397    .007669  -.0048857   .0386982
```

```
      graph d.lp tempo, c(l) s(.) xlabel ylabel yline(0.0107)
```

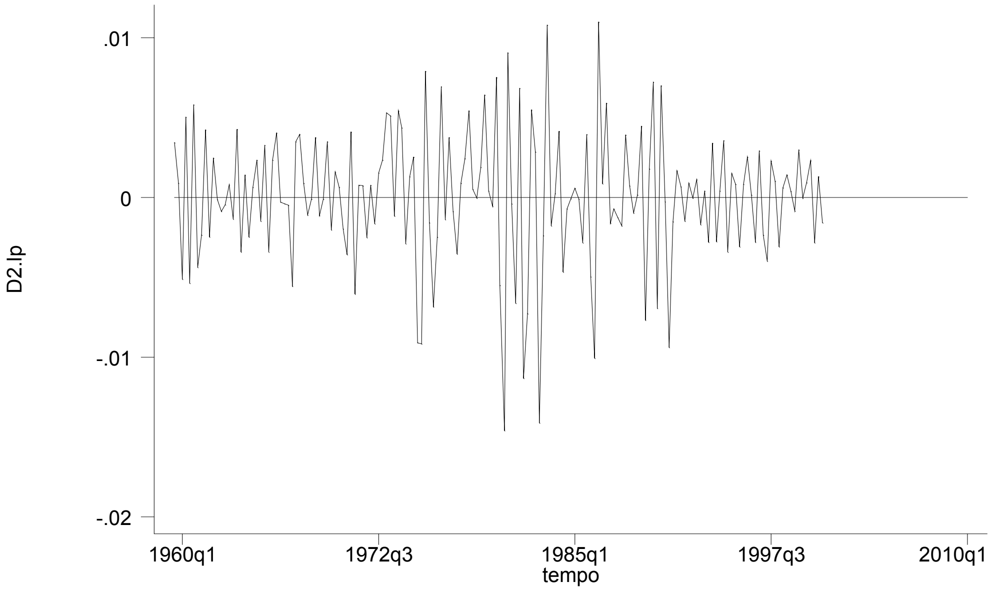


Non c'è trend ma, nonostante ciò, le realizzazioni persistono a lungo sopra/sotto la media campionaria: potrebbe essere che media e varianza di dlp_t (= d.lp) tendano ancora a variare nel tempo. Si ricorda che, oltre ad un trend, anche una forte persistenza al di fuori della media campionaria suggerisce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

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

```
      summ d2.lp

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

```
graph d2.lp tempo, c(l) 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 $\log(P_t)$: $\Delta^2\log(P_t) = \Delta\log(P_t)-\Delta\log(P_{t-1})$.

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 eteogenei (non sommabili) in quanto prodotti nel tempo da diversi processi stocastici.

3.2 La verifica statistica della stazionarietà: il test di 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).

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

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

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.835            -4.019            -3.442            -3.142
------------------------------------------------------------------------------
* MacKinnon approximate p-value for Z(t) = 0.6875

------------------------------------------------------------------------------
D.lp         |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lp           |
          L1 |  -.0056673   .0030888    -1.83   0.068    -.0117692    .0004347
          LD |   .6861981   .0802166     8.55   0.000     .5277313    .8446649
         L2D |  -.0471866   .0957877    -0.49   0.623    -.2364141    .1420409
         L3D |   .4423717   .0886942     4.99   0.000     .2671574    .6175859
         L4D |  -.1809553   .0958216    -1.89   0.061    -.3702497     .008339
         L5D |   .0096405   .0801159     0.12   0.904    -.1486275    .1679085
_trend       |   .0000705   .0000407     1.73   0.085    -9.89e-06    .0001508
_cons        |   .0191888   .0097061     1.98   0.050     .0000145    .0383631
------------------------------------------------------------------------------
```

L'opzione *regress* del comando `dfuller` permette di visualizzare la stima del modello di riferimento per il test statistico, allo scopo di valutare, come prima cosa, la validità della scelta dell'ordine di augmentation: se al quinto lag (quello più elevato) corrisponde una t-Student (non standard) in valore assoluto minore di 1.645 (nel nostro caso t=0.12), si rieffettua il test scegliendo una augmentation ridotta (nel nostro caso 4) e così via finché il lag (ritardo) più elevato è significativo (|t|>1.645).

```
      dfuller lp, lags(4) trend 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)             -1.748            -4.019            -3.442            -3.142
------------------------------------------------------------------------------
* MacKinnon approximate p-value for Z(t) = 0.7288

------------------------------------------------------------------------------
D.lp         |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lp           |
          L1 |  -.0052657   .0030127    -1.75   0.082    -.0112166    .0006853
          LD |   .6795243   .0782437     8.68   0.000     .5249705    .8340781
         L2D |  -.0384192   .0881595    -0.44   0.664    -.2125595    .1357212
         L3D |   .4432444   .0882193     5.02   0.000     .2689859    .6175028
         L4D |  -.1790887   .0784455    -2.28   0.024     -.334041   -.0241364
_trend       |   .0000646   .0000396     1.63   0.105    -.0000137    .0001428
_cons        |   .0180505   .0094943     1.90   0.059    -.0007035    .0368044
------------------------------------------------------------------------------
```

(c) dopo avere definito il modello di riferimento, la statistica DF è la t-Student corrispondente al livello ritardato (L1), nel nostro caso –1.748 (arrotondato nel tabulato a –1.75);

(d) confronto della statistica test (-1.75) con il valore critico al 5% (-3.442): se la statistica è maggiore (con segno) del valore critico, come nel nostro caso, non si rifiuta l'ipotesi nulla di non stazionarietà di lp (log-livelli dei prezzi). In modo alternativo, non si rifiuta H_0 perché il p-value della statistica test (pari al 72.88%) indica una probabilità di errore di prima specie ben al di sopra del 5%.

Lo stesso test per il tasso di inflazione (differenza prima di lp) fornisce un risultato analogo; si noti che in questo caso, data l'evidenza grafica, il modello di riferimento non ha trend:

```
      dfuller d.lp, lags(5) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =       161

                               ---------- 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     0.32   0.752    -.1372289    .1896146
        L5D2 |   .0227313   .0792953     0.29   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)             -2.565            -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)             -7.889            -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.

IX LEZIONE: paragrafi 3.3-3.5

3.3 Definizione e stima dei momenti primi e secondi di serie stazionarie: il correlogramma

Una variabile casuale Y si dice stazionaria in covarianza se i suoi momenti primi e secondo non dipendono dal tempo:

- valore atteso (media), momento primo: $E(Y_t) = \mu$
- autocovarianza di ordine k, momenti secondi: $E[\, (Y_t - \mu)(Y_{t-k} - \mu)\,] = \gamma_k$ con $k=0,1,2,\ldots$. (ricorda che, quando k=0, γ_o è la variaza)

Allo scopo di superare il problema dell'unità di misura dei dati di autocovarianza, si usa un indice (numero puro, $-1 < \rho_k < 1$):

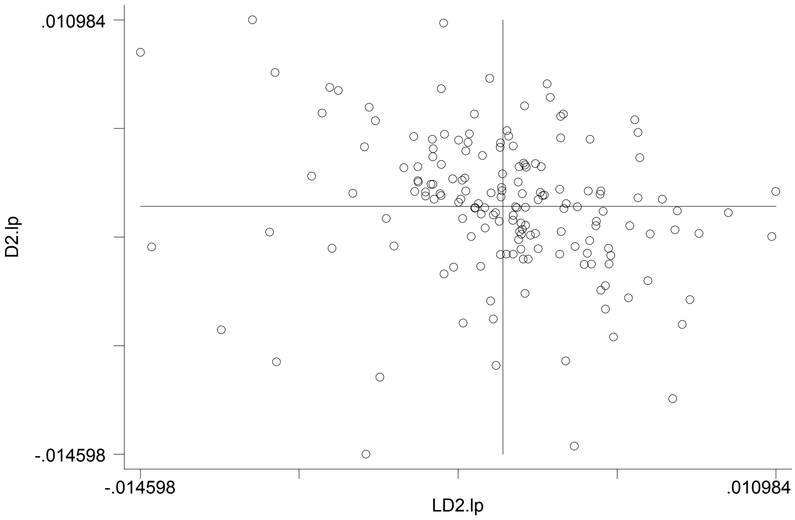
- coefficiente di autocorrelazione di ordine k: $\rho_k = \gamma_k/\gamma_o$ (se $\rho_k = 0$, Y_t non è correlata con Y_{t-k}).

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, 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: $COV(Y_t, Y_{t-k})$, con $k=1,2,\ldots$ 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: $|COV(Y_t, Y_{t-1})| > |COV(Y_t, Y_{t-10})|$. 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.

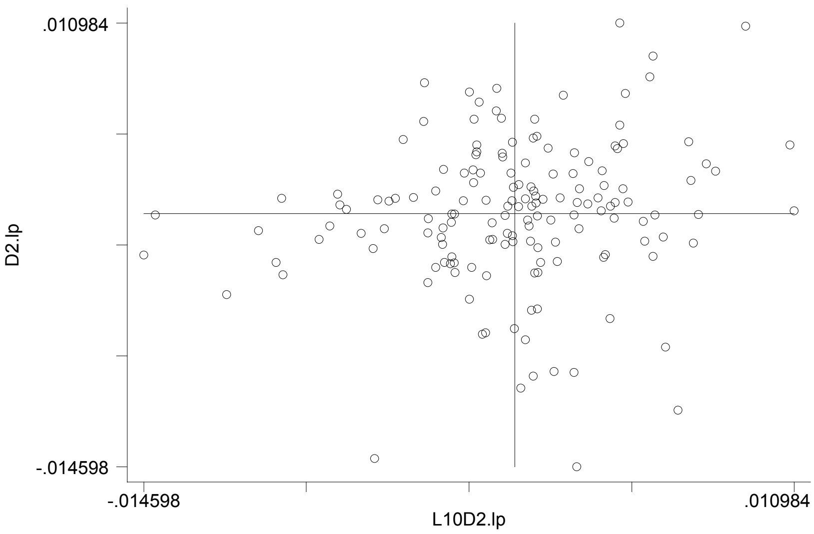
```
        summm d2.lp ld2.lp

    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)
```

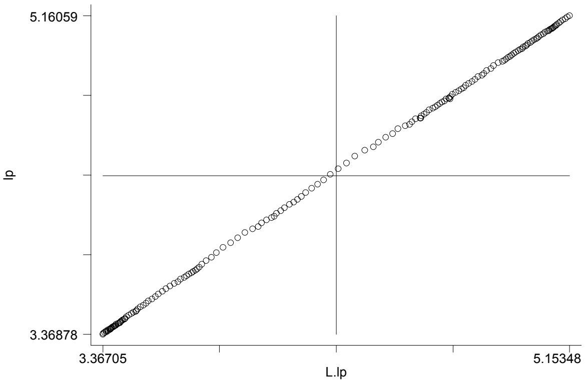


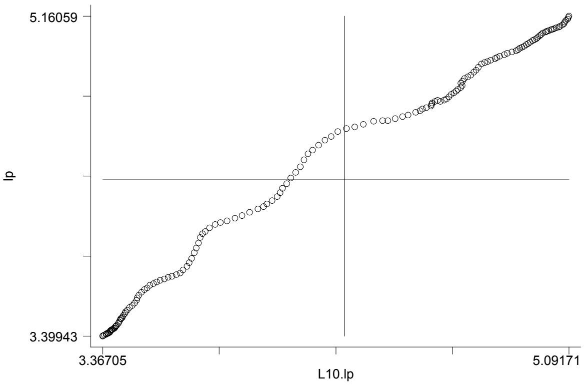
Lo scatter indica l'esistenza di un legame negativo, $\hat{\rho}_1 < 0$, fra Δdlp_t (d2.lp) e Δdlp_{t-1} (ld2.lp), mentre il legame sparisce se ci si concentra su Δdlp_t (d2.lp) e Δdlp_{t-10} (l10d2.lp), $\hat{\rho}_{10} \cong 0$:

```
        graph d2.lp l10d2.lp, xline(0) yline(0)
```



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 orrizzontale 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)
```

```
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [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.

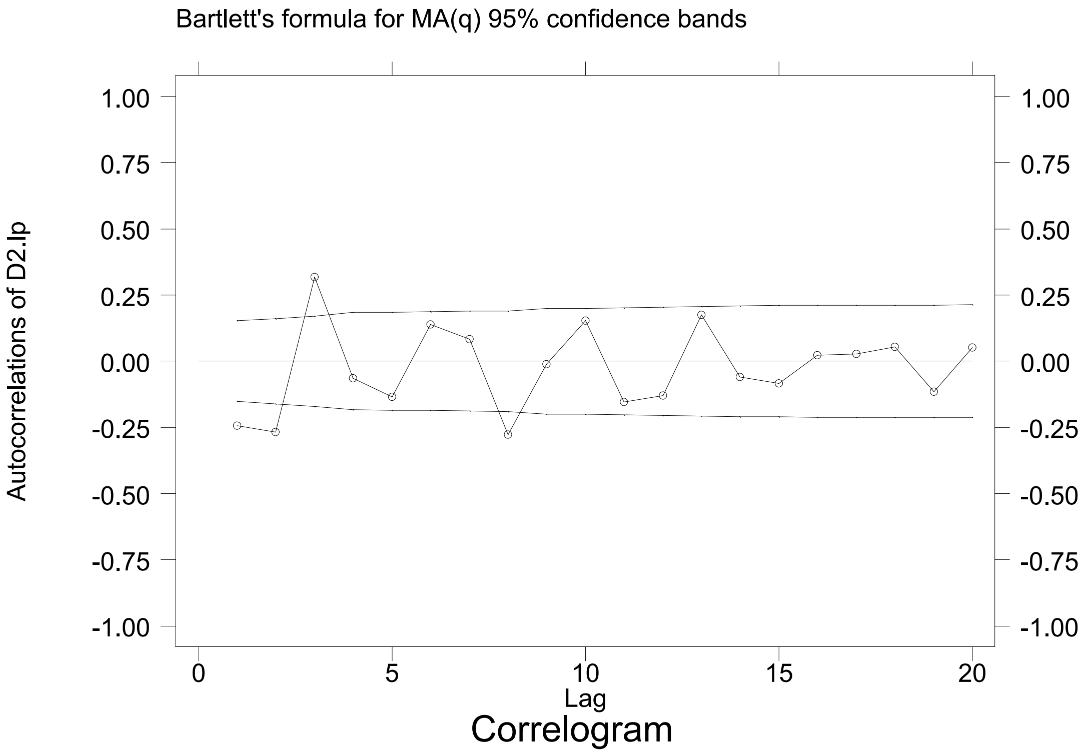
Le stime dei coefficienti di autocorrelazione parziale (PAC) vengono spiegati in seguito, quando si tratterà di scegliere l'ordine dei ritardi del modello autoregressivo (AR).

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 Ho, 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 ≈ 0 suggerisce il rifiuto di Ho: alcuni dei primi 20 coefficienti ρ_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":

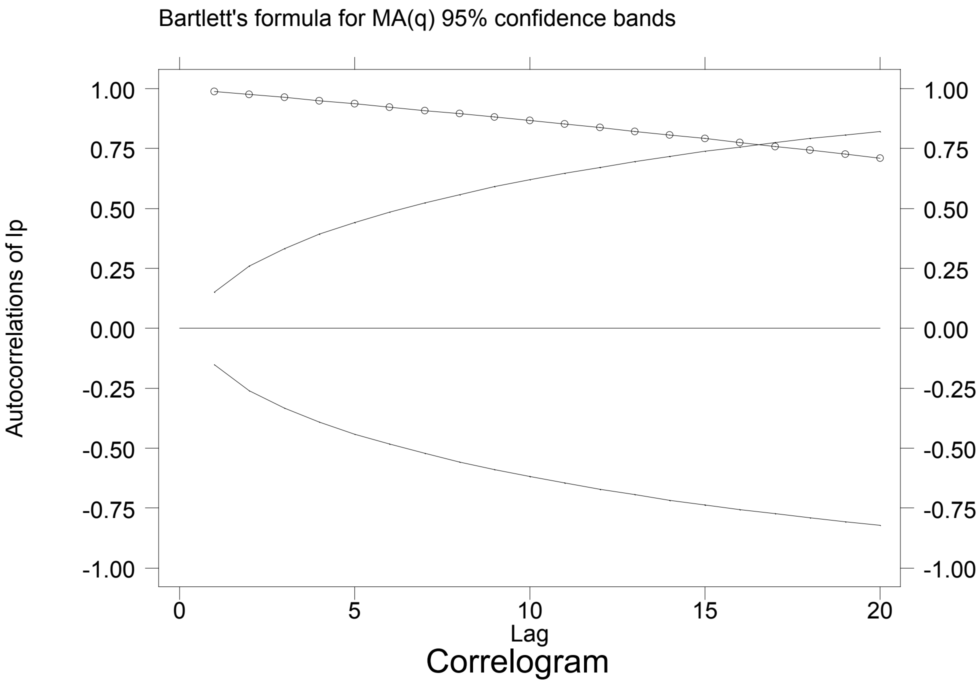
```
ac d2.lp , lags(20)
```



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

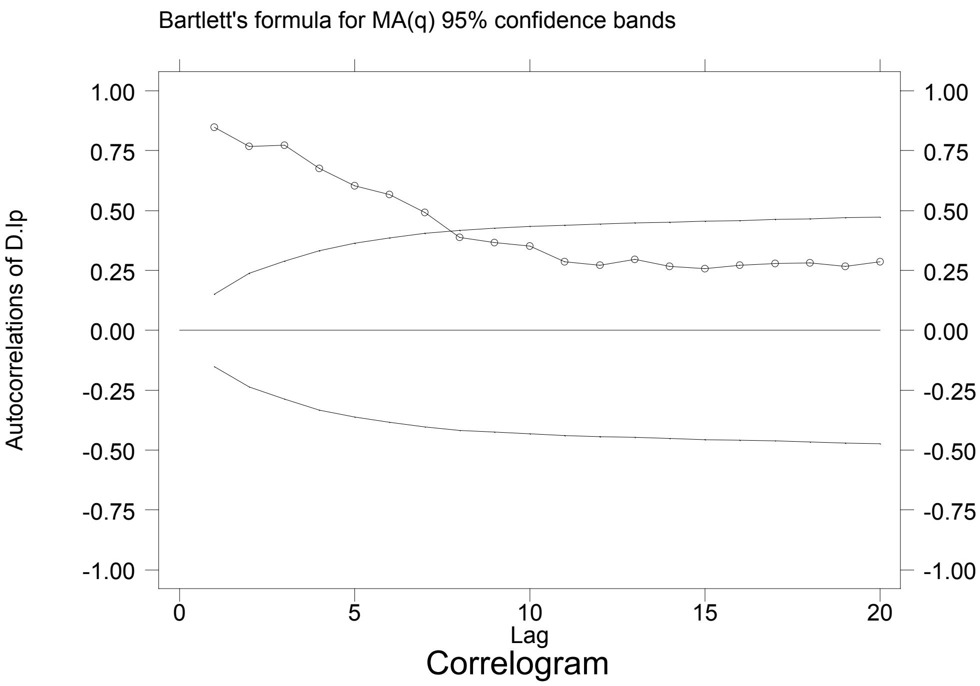
In precedenza si è notato graficamente (scatter) che, quando un processo non è stazionario, le sue autocorrelazioni sono sempre molto forti (decrescono molto, molto lentamente con k). Il correlogramma di lp ci fa apprezzare visivamente questo fatto.

```
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 …

3.4 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, già discusso nella lezione precedente, 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 affrontato 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) Scelta della trasformazione di Y_t che la rende stazionaria. Nel caso in cui si scelga di renderla stazionaria per differenze successive, la differenziazione a livello della quale il test DF rifiuta Ho fornisce la variabile trasformata, punto di partenza per l'identificazione del modello generale AR(p).

(b) Scelta dell'ordine (p) del modello AR. Una regola pratica è quella di iniziare da un ordine collegato alla periodicità dei dati (in modo simile a ciò che si è visto per augmentation del test DF).

(c) Analisi dei test diagnostici del modello AR(p). Si salvano i residui e si verifica se sono white noise, oppure se sono ancora autocorrelati (nel qual caso si ristima un modello in cui compaiono altri 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 e al successivo controllo diagnostico dei residui del modello vincolato.

Evidentemente, la fase di modellazione AR termina con un modello il cui correlogramma dei residui è "vuoto" (white noise) perché i test Q non rifiutano H_o per qualsiasi ordine e i cui parametri stimati (tranne la costante, da lasciare comunque) sono significativamente diversi da zero.

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

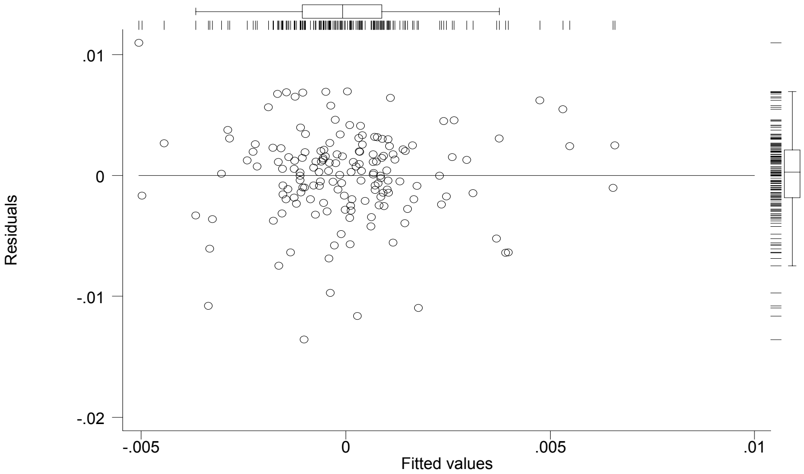
(a) alla luce dei test di stazionarietà, i log-livelli dei prezzi devono essere differenziati due volte per essere resi stazionari, quindi ci si concentra sulla modellazione AR della variazione del tasso di inflazione. Inoltre, l'analisi del correlogramma per $\Delta infl_t$ suggerisce il rifiuto di Ho 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 d2l.lp d2l2.lp d2l3.lp d2l4.lp d2l5.lp

      Source |       SS       df       MS              Number of obs =     161
-------------+------------------------------           F(  5,   155) =    7.72
       Model |  .000560383     5  .000112077           Prob > F      =  0.0000
    Residual |  .002251216   155  .000014524           R-squared     =  0.1993
-------------+------------------------------           Adj R-squared =  0.1735
       Total |  .002811599   160  .000017572           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
------------------------------------------------------------------------------
```

(c) Test diagnostici sui residui di stima.

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

Infine, si verifica l'autocorrelazione dei residui (sotto Ho sono white noise, cioè il correlogramma dei residui è "vuoto"):

```
      corrgram resar5 , lags(20)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
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 Ho 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 Δdlp_t .

(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.

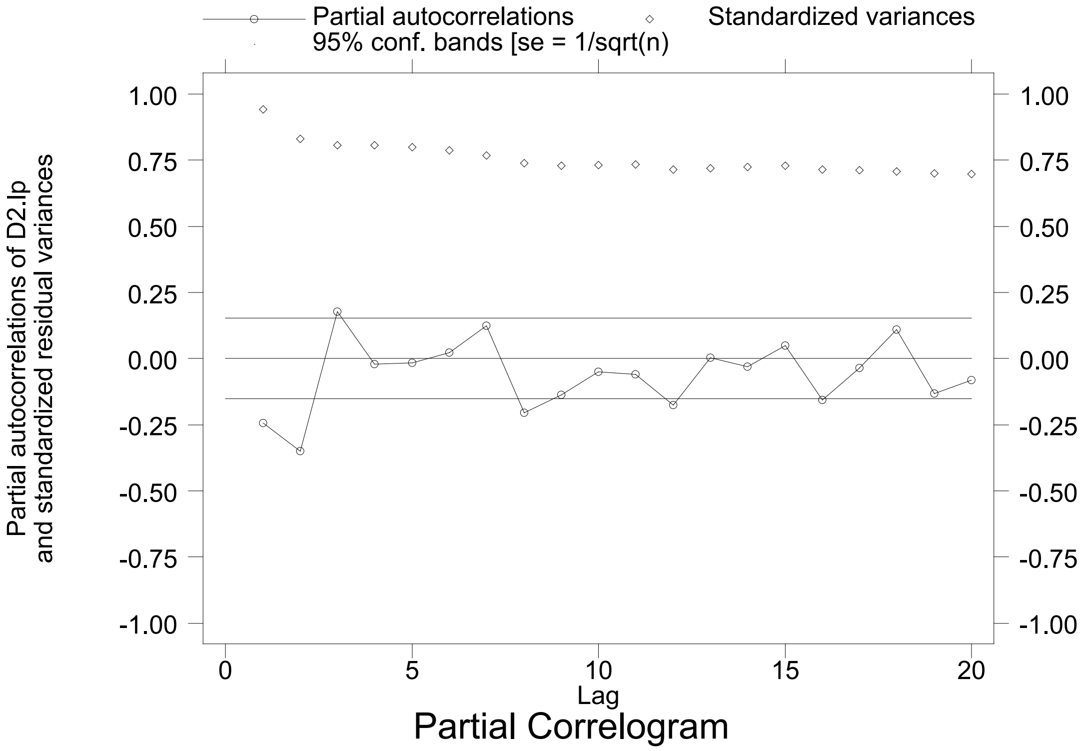
```
      reg d2.lp d2l.lp d2l2.lp d2l3.lp

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

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [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 (A CASA): sceglietе 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 Ho. 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
-------------+------------------------------           F(  1,   163) =   10.26
       Model |  .000171343     1  .000171343           Prob > F      =  0.0016
    Residual |  .002722186   163  .000016701           R-squared     =  0.0592
-------------+------------------------------           Adj R-squared =  0.0534
       Total |  .002893529   164  .000017643           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 resar1, lag(20)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [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.

PARAGRAFO FACOLTATIVO:

3.5 Le previsione univariata 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:

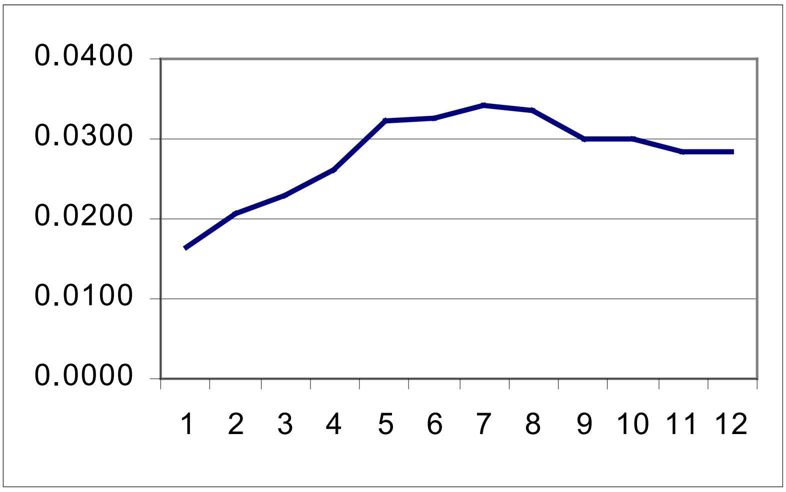
(a) previsione della variabile trasformata (la variabile dipendente del modello AR, nel nostro caso d2.lp). Tale previsione e' non condizionale perché la sua formulazione richiede solo la conoscenza dei valori storici di fine periodo della variabile da prevedere. La previsione condizionale richiede anche dati e ipotesi esterne sull'andamento storico e futuro di altre variabili economiche.

(b) 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].

(c) 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).

X LEZIONE: paragrafi 3.6-5

3.6 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.

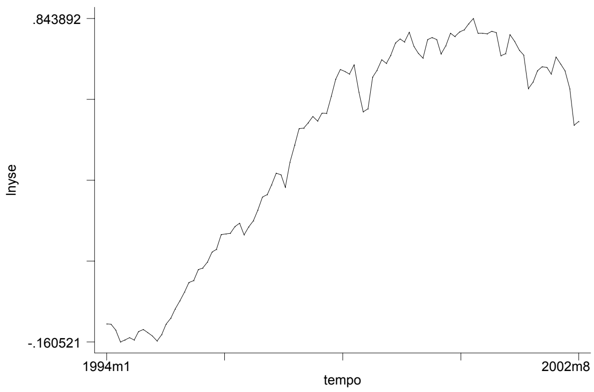
L'esistenza di autocorrelazione fra le realizzazioni della variabile di interesse consentono la modellazione e previsione mediante modelli AR. L'autocorrelazione dei residui ha una duplice valenza: da un lato, residui autocorrelati segnalano una scorretta specificazione della dinamica del modello AR; dall'altro, la presenza di autocorrelazione nei residui fa venire meno la proprietà della consistenza dello stimatore OLS (abitualmente utilizzato per stimare i parametri del modello). In altri termini, le autocorrelazioni ρ_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.

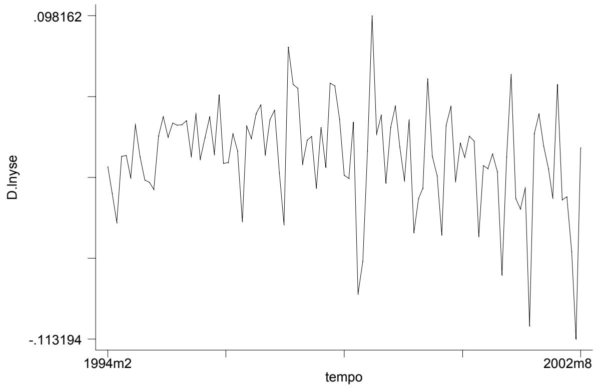
Per provare tale assunto, si copia in Stata il contenuto della cartella "mensile" del foglio Excel "datiTS,xls" (vedi esercizio sopra) e immetto la sequenza di comandi:

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

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




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



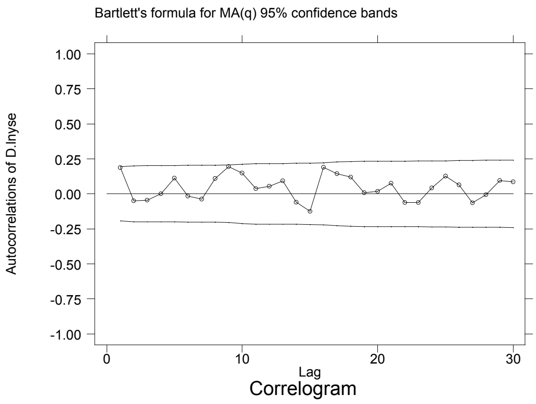

```
dfuller d.lnyse, lag(13) regress
```

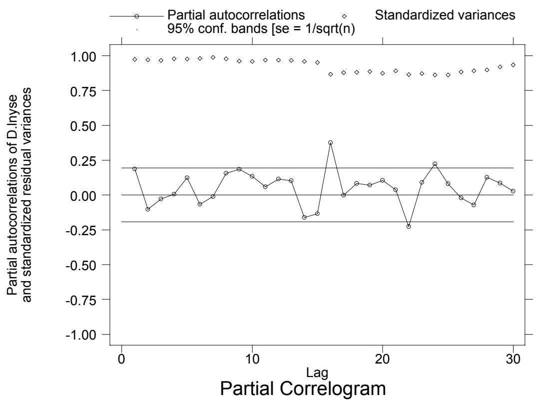
```
Augmented Dickey-Fuller test for unit root         Number of obs   =        89

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

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.793            -3.518            -2.895            -2.582
------------------------------------------------------------------------------
* 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

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.343            -3.517            -2.894            -2.582
------------------------------------------------------------------------------
* 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

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

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [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)                     pac d.lnyse, lag(30)
```





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 di uno dei più famosi econometrici viventi, Clive Granger, che non molto tempo fa ha affermato che: "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

L'analisi e la modellazione univariata delle serie storiche economiche offre strumenti semplici e utili per l'analisi preliminare e 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).

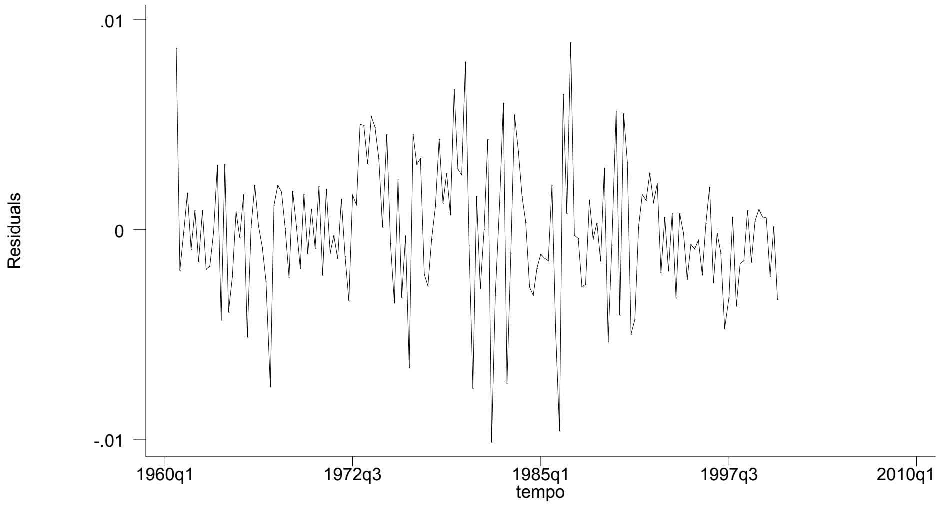
L'essenza dell'analisi multivariata consiste nell'estensione dell'approccio univariato grazie alle informazioni sui nessi di causalità fra variabili che si desumono dalle teorie economiche. L'analisi multivariata che affronteremo in questo paragrafo è di tipo uniequazionale perché si concentra su relazioni economiche misurate in una sola equazione. Ad esempio, il precedente modello AR per il tasso di inflazione può essere esteso allo scopo di sfruttare l'informazione contenuta nel tasso di disoccupazione; dalla teoria della curva di Phillips è possibile infatti ricavare una formulazione 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, ur)

Il modello dinamico di partenza muove da una dinamica ricca di regressori che, verificata la bianchezza dei residui con test di scorretta specificazione, viene ordinatamente ridotta con l'ausilio di una batteria di test di significatività.

```
reg d2.lp ld2.lp l2d2.lp l3d2.lp l4d2.lp l5d2.lp l.ur l2.ur l3.ur l4.ur l5.ur
```

```
      Source |       SS       df       MS              Number of obs =     161
-------------+------------------------------           F( 10,   150) =   10.36
       Model |  .001148657    10  .000114866           Prob > F      =  0.0000
    Residual |  .001662942   150  .000011086           R-squared     =  0.4085
-------------+------------------------------           Adj R-squared =  0.3691
       Total |  .002811599   160  .000017572           Root MSE      =  .00333

------------------------------------------------------------------------------
D2.lp        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lp           |
         LD2 |  -.3990169   .0796808    -5.01   0.000    -.5564586   -.2415751
        L2D2 |  -.3038825   .0862994    -3.52   0.001     -.474402    -.133363
        L3D2 |  -.0045659   .0893011    -0.05   0.959    -.1810165    .1718846
        L4D2 |  -.1099272   .0816498    -1.35   0.180    -.2712594     .051405
        L5D2 |  -.0158374   .0737752    -0.21   0.830    -.1616102    .1299354

ur           |
          L1 |  -.6904383    .111317    -6.20   0.000    -.9103902   -.4704864
          L2 |   .9063062   .2214601     4.09   0.000     .4687221     1.34389
          L3 |  -.1092277   .2379083    -0.46   0.647    -.5793119    .3608565
          L4 |  -.4963373   .2208568    -2.25   0.026    -.9327295   -.0599451
          L5 |   .3438748   .1157663     2.97   0.003     .1151315    .5726181
_cons        |   .0027616   .0012728     2.17   0.032     .0002466    .0052767
------------------------------------------------------------------------------

      hettest

Cook-Weisberg test for heteroskedasticity using fitted values of D2.lp
     Ho: Constant variance
         chi2(1)      =      4.20
         Prob > chi2  =      0.0403

      ovtest

Ramsey RESET test using powers of the fitted values of D2.lp
       Ho:  model has no omitted variables
                  F(3, 147) =      2.02
                   Prob > F =      0.1131

      dwstat
Durbin-Watson d-statistic( 11,   161) =  1.920004

      predict resardl, resid
(7 missing values generated)

      graph resardl tempo, c(l) 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 troppo preoccupante) di eteroschedasticità nei residui.

Dal test di Durbin-Watson di autocorrelazione del primo ordine dei residui riportato sopra non emerge nulla di preoccupante. Inoltre dal correlogramma sotto, si evince che i residui non presentano autocorrelazioni significative e, quindi, come atteso (vista la dimensione molto elevata) il modello è in grado di dare conto della struttura correlativa dei dati sulla differenza prima dell'inflazione.

```
      corrgram resardl, lag(20)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.0142   0.0144   .03314  0.8555          |                  |
2        0.0075   0.0073   .04255  0.9790          |                  |
3        0.0213   0.0214   .11757  0.9896          |                  |
4       -0.0085  -0.0092   .12954  0.9980          |                  |
5        0.0411   0.0417   .41392  0.9949          |                  |
6        0.0290   0.0276   .55643  0.9971          |                  |
7        0.0454   0.0446   .90841  0.9962          |                  |
8       -0.1545  -0.1607   5.0051  0.7570         -|                 -|
9       -0.0226  -0.0172   5.0935  0.8261          |                  |
10      -0.0434  -0.0454   5.4214  0.8613          |                  |
11      -0.0188  -0.0130   5.4833  0.9055          |                  |
12      -0.1054  -0.1174    7.439  0.8273          |                  |
13       0.1767   0.2099   12.978  0.4495          |-                 |-
14       0.1059   0.1152    14.98  0.3795          |                  |
15      -0.0733  -0.0620   15.946  0.3856          |                  |
16       0.0644   0.0435   16.697  0.4055          |                  |
17       0.0257   0.0457   16.817  0.4668          |                  |
18       0.0745   0.0606   17.837  0.4665          |                  |
19      -0.1132  -0.1823   20.206  0.3823          |                 -|
20       0.0733   0.0446   21.207  0.3850          |                  |
```

La performance dei diagnostici di scorretta specificazione sui residui è tale da permetterci di passare ai test di significatività dei parametri. In particolare sembrano particolarmente poco significative le stime dei ritardi di ordine 3, 4 e 5 della componente autoregressiva e il terzo ritardo del tasso di disoccupazione. Supponendo che tutti i corrispondenti parametri siano simultaneamente nulli, si stima il modello vincolato:

```
      reg d2.lp ld2.lp l2d2.lp  l.ur l2.ur l4.ur l5.ur

      Source |       SS       df       MS              Number of obs =     163
-------------+------------------------------           F(  6,   156) =   17.39
       Model |  .001148682     6  .000191447           Prob > F      =  0.0000
    Residual |  .001717559   156   .00001101           R-squared     =  0.4008
-------------+------------------------------           Adj R-squared =  0.3777
       Total |  .002866241   162  .000017693           Root MSE      =  .00332

------------------------------------------------------------------------------
D2.lp        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lp           |
         LD2 |  -.4385675   .0676407    -6.48   0.000    -.5721772   -.3049577
        L2D2 |  -.2901734   .0736948    -3.94   0.000    -.4357418    -.144605
ur           |
          L1 |  -.6653684   .1001241    -6.65   0.000    -.8631423   -.4675945
          L2 |   .8091745   .1462322     5.53   0.000     .5203238    1.098025
          L4 |  -.4913805   .1457863    -3.37   0.001    -.7793504   -.2034106
          L5 |    .306413   .0986296     3.11   0.002     .1115912    .5012347
_cons        |   .0024396   .0011873     2.05   0.042     .0000944    .0047848
------------------------------------------------------------------------------

      hettest
Cook-Weisberg test for heteroskedasticity using fitted values of D2.lp
     Ho: Constant variance
         chi2(1)      =     3.34
         Prob > chi2  =     0.0677

      ovtest
Ramsey RESET test using powers of the fitted values of D2.lp
       Ho:  model has no omitted variables
                 F(3, 153) =      2.74
                  Prob > F =      0.0454

      dwstat
Durbin-Watson d-statistic(  7,   163) =  1.904991


      predict res, resid
(5 missing values generated)

      corrgram res, lag(20)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.0414   0.0416   .28416  0.5940          |                  |
2       -0.0496  -0.0515    .6959  0.7061          |                  |
3        0.0121   0.0167   .72031  0.8684          |                  |
4       -0.0875  -0.0924   2.0146  0.7331          |                  |
5        0.0492   0.0616   2.4266  0.7875          |                  |
6        0.0686   0.0551   3.2327  0.7791          |                  |
7        0.0323   0.0337   3.4121  0.8444          |                  |
8       -0.1387  -0.1502   6.7512  0.5637         -|                 -|
9       -0.0403  -0.0142   7.0352  0.6335          |                  |
10      -0.0403  -0.0444   7.3211  0.6948          |                  |
11      -0.0286  -0.0226   7.4654  0.7602          |                  |
12      -0.0849  -0.1280   8.7503  0.7241          |                 -|
13       0.1682   0.2019   13.824  0.3864          |-                 |-
14       0.0956   0.0872   15.473  0.3466          |                  |
15      -0.0483  -0.0329   15.897  0.3889          |                  |
```

```
16       0.0679   0.0492    16.74  0.4026          |                  |
17       0.0214   0.0673   16.825  0.4663          |                  |
18       0.0558   0.0611   17.402  0.4956          |                  |
19      -0.1126  -0.2047   19.772  0.4084          |                 -|
20       0.0566   0.0512   20.376  0.4347          |                  |
```

Esercizio: calcolate la statistica F per verificare l'accettabilità delle 4 restrizioni che permettono il passaggio dal modello iniziale a quello vincolato.

Si noti che le stime dei parametri dei ritardi distribuiti del tasso di disoccupazione presentano segni alterni a coppie: questo fatto dipende dalla forte persistenza di ur_t che ne suggerisce la differenziazione per spiegare una variabile stazionaria qual è Δdlp_t.

Il modello che spiega la variazione 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.

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 sprovveduto 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).

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              Number of obs =      42
-------------+------------------------------           F(  1,    40) = 7331.62
       Model |  6.98900429     1  6.98900429           Prob > F      =  0.0000
    Residual |  .038130775    40  .000953269           R-squared     =  0.9946
-------------+------------------------------           Adj R-squared =  0.9944
       Total |  7.02713506    41  .171393538           Root MSE      =  .03088

------------------------------------------------------------------------------
        lcus |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lyuk |   1.459444   .0170446    85.62   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 lo conferma:

```
      predict res, resid

      corrgram res, lag(6)
```

```
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
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          |                 |
```

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
-------------+------------------------------           F(  4,    35) =    2.19
       Model |  .002314119     4   .00057853           Prob > F      =  0.0897
    Residual |  .009224876    35  .000263568           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    -0.82   0.419    -.5798511     .246629
lcus         |
          L1 |    .010556   .0895768     0.12   0.907    -.1712944    .1924065
lyuk         |
          L1 |  -.0287854   .1317032    -0.22   0.828    -.2961572    .2385864
_cons        |   .3167983   1.013507     0.31   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)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [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          |                 |
```

Come esercizio, si lascia agli studenti l'ispezione del grafico delle serie, del loro correlogramma e dei test di Dickey-Fuller. Ne emergerà 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.

Infine, si lascia ancora come esercizio, la stima del modello che muove dal modello ARDL precedente ed elimina tutti i parametri non significativi (evidentemente riferiti alla variabile reddito del Regno Unito, come è più che lecito attendersi).