ECON532 Applied Econometrics

Lecture 9

Introduction to Time Series and Forecasting

Time Series and Regression

Time series data are data observed over multiple time periods on the same variable(s):

Why time series data?

1. To develop models to allow forecasting
   - the rate of inflation next year?
   - the risk level on the stock market tomorrow?
   - the demand level for Coca Cola in 2019?

2. To estimate <u>dynamic</u> causal effects

- of an increase in cigarette taxes on cigarette consumption this year, next year, in 5 years;
- of a change in the Fed Funds rate on inflation, this month, in 6 months, and 1 year;
- of a freeze in Florida on the price of orange juice concentrate in 1 month, 2 months, 3 months…

Regression Models for Forecasting

Forecasting and estimation of causal effects are quite different objectives.

For forecasting:

- Theoretical justification not as important
- $\bar{R}^2$, SER is not as relevant
- Omitted variable bias isn't as much of an issue!
- "External validity": the model estimated using historical data must hold into the (near) future

New technical issues

- Time series data are rarely i.i.d
  - Correlation over time (*serial correlation*, a.k.a. *autocorrelation*)
    - OLS estimator still unbiased, consistent and approximately normal distributed, but is not BLUE and the SE is different.
  - Estimating standard errors under serially correlated residuals

- Models need not have a causal interpretation
- **Stationarity** is very, very important
  - estimation biases and inconsistencies
  - Are we estimating something that even exists?

- Data are observed only ONCE in each time period and for ONE realization
- Choice of time frequency

Time series data

- Time lags
- *Autocorrelation*, lagged cross-correlation

- Forecasting from regression methods:
  - *autoregressive* (AR) models
  - *autoregressive distributed lag* (ADL) models
  - need not have a causal interpretation

- Conditions under which dynamic effects can be estimated, and how to estimate them

- Validity of tests when the data are serially correlated

Notations and Terminologies for Time Series

First, some notation and terminology.

Notation for time series data

- Y_t = value of Y in period t.
- Data set: $Y_1,\ldots,Y_T$ = T consecutive observations on the random variable Y
- We consider only consecutive, evenly-spaced observations

(missing and non-evenly spaced data introduce even MORE technical complications)

Notations and Terminologies for Time Series

Lags, First Differences, Logarithms, and Growth Rates

- The first lag of a time series Y_t is Y_{t-1}; its j^{th} lag is Y_{t-j}.
- The first difference of a series, ΔY_t, is its change between periods $t - 1$ and t, that is, $\Delta Y_t = Y_t - Y_{t-1}$.
- The first difference of the logarithm of Y_t is $\Delta\ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1})$.
- The percentage change of a time series Y_t between periods $t - 1$ and t is approximately $100\Delta\ln(Y_t)$, where the approximation is most accurate when the percentage change is small.

Transformations mainly applied to achieve:

STATIONARITY and/or

to observe GROWTH

Forecast uncertainty and forecast intervals

Why do you need a measure of forecast uncertainty?

- To construct forecast intervals
- To let users of your forecast (including yourself) know what degree of accuracy to expect

Consider the forecast

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\beta}_1 X_T$$

The forecast error is:

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_1 - \beta_2)X_T]$$

The *mean squared forecast error (MSFE)* is,

$$\mathrm{E}(Y_{T+1} - \hat{Y}_{T+1|T})^2 = E(u_{T+1})^2 +$$

$$+ E[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_1 - \beta_2)X_T]^2$$

- MSFE = var(u_{T+1}) + uncertainty arising because of estimation error
- If the sample size is large, the part from the estimation error is (much) smaller than var(u_{T+1}), in which case

$$\text{MSFE} \approx \text{var}(u_{T+1})$$

- The ***root mean squared forecast error (RMSFE)*** is the square root of the MS forecast error:

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

The root mean squared forecast error (RMSFE)

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

- The RMSFE is a measure of the spread of the forecast error distribution.
- The RMSFE is like the standard deviation of u_t, except that it explicitly focuses on the forecast error using estimated coefficients, not using the population regression line.
- The RMSFE is a measure of the magnitude of a typical forecasting "mistake"

Three ways to estimate the RMSFE

1. Use the approximation RMSFE = σ_u, so estimate the RMSFE by the SER.
2. Use an actual forecast history for $t = t_1,\ldots, T$, then estimate by

$$MSFE = \frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} (Y_{t+1} - \hat{Y}_{t+1|t})^2$$

   Usually, this isn't practical – it requires having an historical record of actual forecasts from your model
3. Use a simulated forecast history, that is, simulate the forecasts you would have made using your model in real time….then use method 2, with these ***pseudo out-of-sample forecasts***…

The method of *pseudo out-of-sample forecasting*

- Re-estimate your model every period, $t = t_1 - 1, \ldots, T-1$
- Compute your "forecast" for date $t+1$ using the model estimated through t
- Compute your pseudo out-of-sample forecast at date t, using the model estimated through $t-1$. This is $\hat{Y}_{t+1|t}$.
- Compute the *poos* forecast error, $Y_{t+1} - \hat{Y}_{t+1|t}$
- Plug this forecast error into the MSFE formula,

$$MSFE = \frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} (Y_{t+1} - \hat{Y}_{t+1|t})^2$$

Why the term "pseudo out-of-sample forecasts"?

Using the RMSFE to construct forecast intervals

If u_{T+1} is normally distributed, then a 95% forecast interval can be constructed as

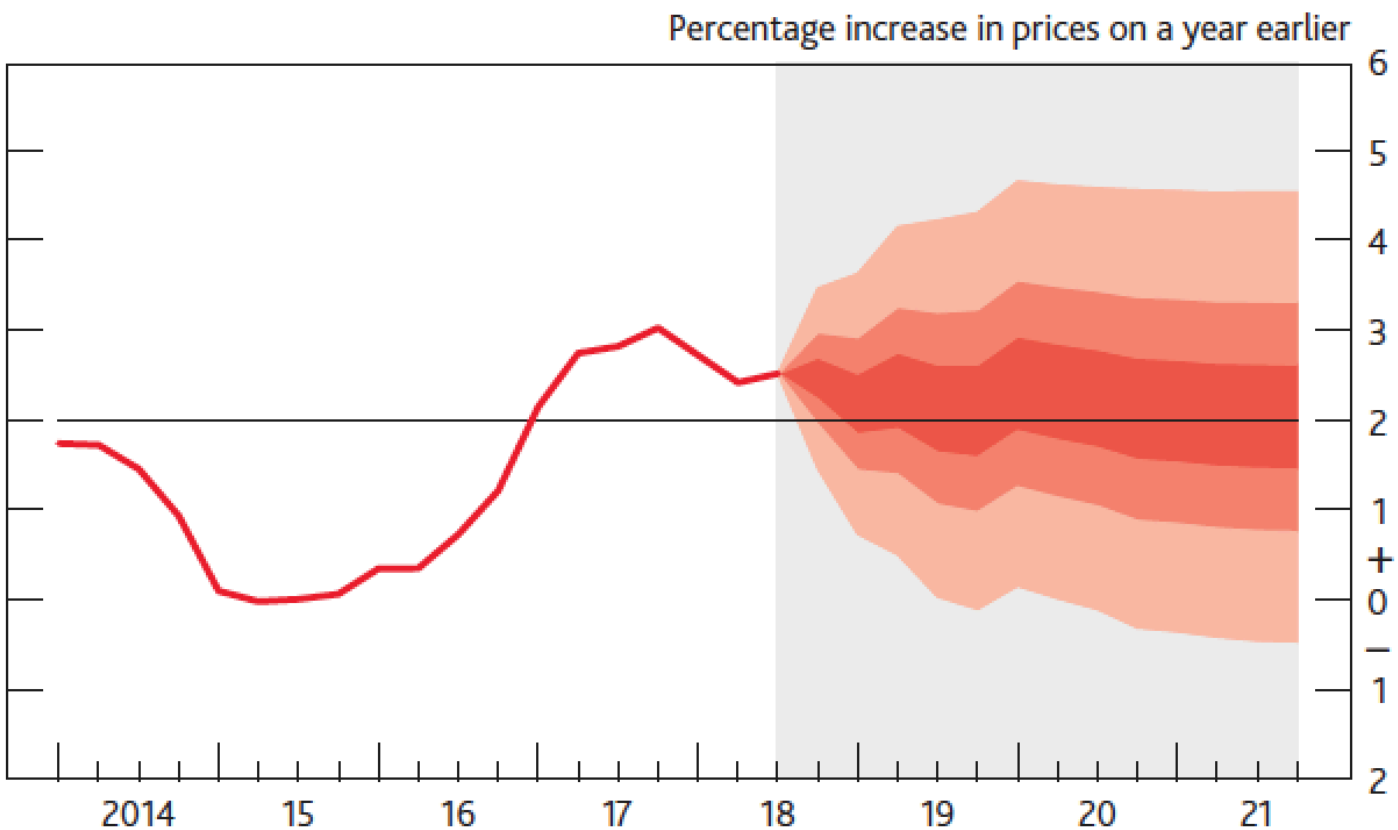
$$\hat{Y}_{T|T-1} \pm 1.96 \times RMSFE$$

Note:

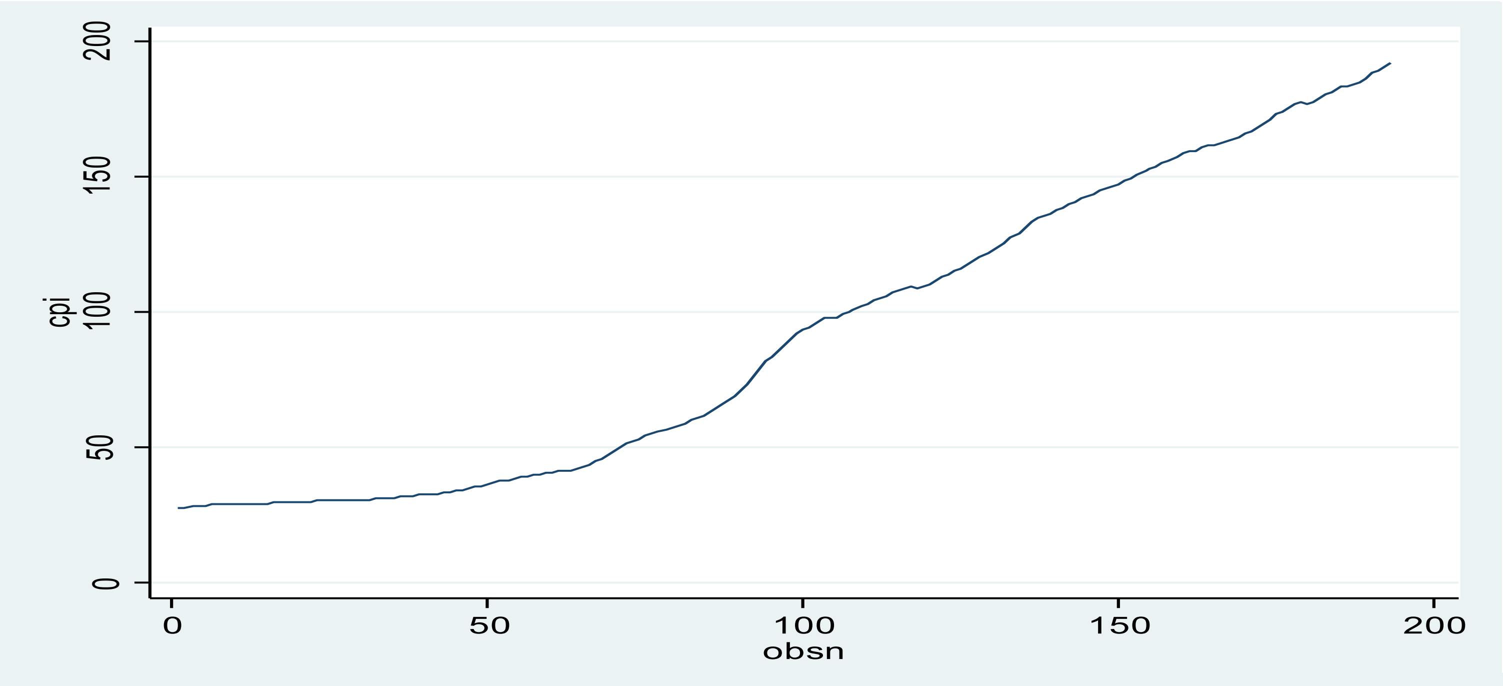
1. A 95% forecast interval is not a confidence interval (Y_{T+1} isn't a nonrandom coefficient, it is random!)
2. This interval is only valid if u_{T+1} is normal – but still might be a reasonable approximation and is a commonly used measure of forecast uncertainty
3. Often "67%" forecast intervals are used: $\pm\ RMSFE$

Example : the Bank of England "Fan Chart"

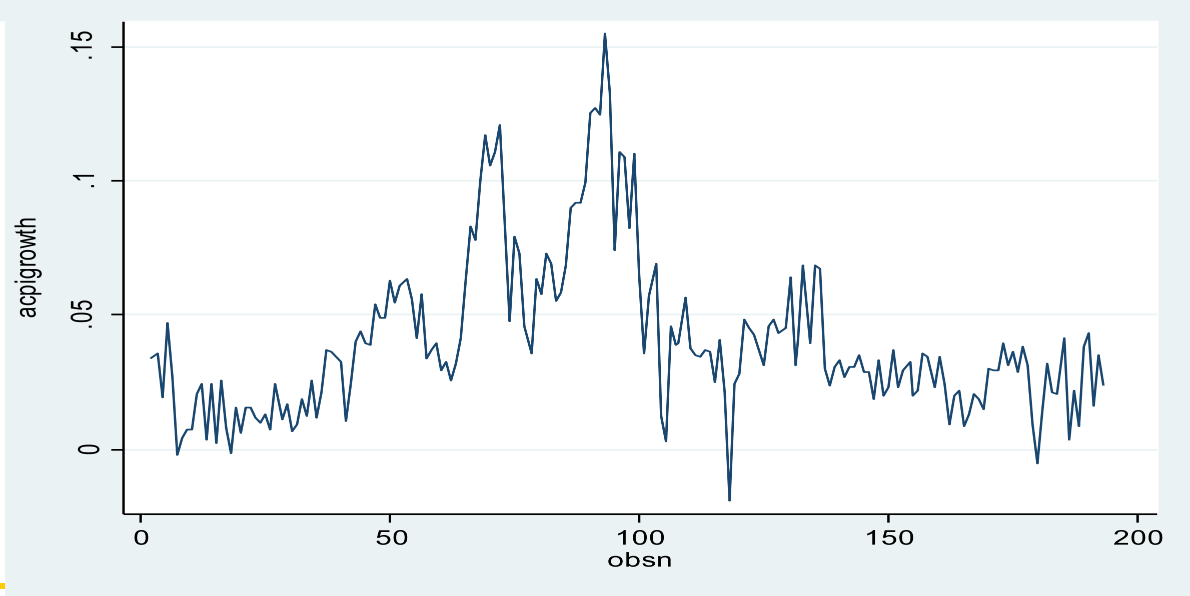
Chart 5.3 CPI inflation projection based on market interest rate expectations, other policy measures as announced



Example: CPI and inflation (U.S.)




```
use "macro_2e.dta", clear
gen obsn=_n
tsset obsn
gen cpi=punew
tsline cpi
gen lcpi=ln(cpi)
tsline lcpi
gen acpigrowth=100*4*d.lcpi
tsline  acpigrowth
```



Example: inflation (U.S.)

CPI = Consumer Price Index (Bureau of Labor Statistics)

- 1st quarter, 2004 = 186.57
- 2nd quarter, 2004 = 188.60
- % change

$$= 100 \times \left(\frac{188.60 - 186.57}{186.57} \right) = 100 \times \left(\frac{2.03}{186.57} \right) = 1.088\%$$

- Now *at an annual rate*

$$= 100((1+1.088/100)^4 \text{ -}1) = \mathbf{4.42}\% \approx \mathbf{4.4}\% \text{ per annum}$$

- Approx with logs yields

$4 \times 100 \times [\log(188.60) - \log(186.57)] = \mathbf{4.329}\%$

Annualizing (quarterly) growth rates

$$R_{t+4,t} = \frac{y_{t+4} - y_t}{y_t} = \frac{\Delta y_{t+1} + \Delta y_{t+2} + \Delta y_{t+3} + \Delta y_{t+4}}{y_t}$$

$$y_{t+4} = y_t\left(1+R_{t+1}\right)\left(1+R_{t+2}\right)\left(1+R_{t+3}\right)\left(1+R_{t+4}\right)$$

$$= y_t\left(1+R_{t+1}\right)^4 \qquad \text{under a constant growth rate}$$

$$\rightarrow R_{t+4,t}(annual) = \left(1+R_{t+1}\right)^4 - 1$$

With *ln* approximation

$$r_{t+4,t} = \ln\left(\frac{y_{t+4}}{y_t}\right) = \ln\left(y_{t+4}\right) - \ln(y_t)$$

$$= \Delta\ln\left(y_{t+1}\right) + \Delta\ln\left(y_{t+2}\right) + \Delta\ln\left(y_{t+3}\right) + \Delta\ln\left(y_{t+4}\right)$$

$$= r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4}$$

$$\therefore \ln\left(y_{t+4}\right) = \ln\left(y_t\right) + r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4}$$

$$= \ln\left(y_t\right) + 4r_{t+1} \qquad \text{under a constant growth rate}$$

$$\rightarrow r_{t+4,t}(annual) = 4r_{t+1}$$

Autocorrelation

The correlation of a series with its own lagged values is called ***autocorrelation*** or ***serial correlation***.

- The first *autocorrelation* of Y_t is corr(Y_t,Y_{t-1})
- Thus

$$\text{corr}(Y_t,Y_{t-1}) = \frac{\text{cov}(Y_t,Y_{t-1})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-1})}} = \rho_1$$

- These are population correlations – they describe one aspect of the joint distribution of (Y_t, Y_{t-1})

Sample autocorrelations

The ***j*^th^ *sample autocorrelation*** is estimated by:

$$\hat{\rho}_j = \frac{\widehat{\mathrm{cov}}(Y_t, Y_{t-j})}{\widehat{\mathrm{var}}(Y_t)}$$

i.e.

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2}$$

NB:

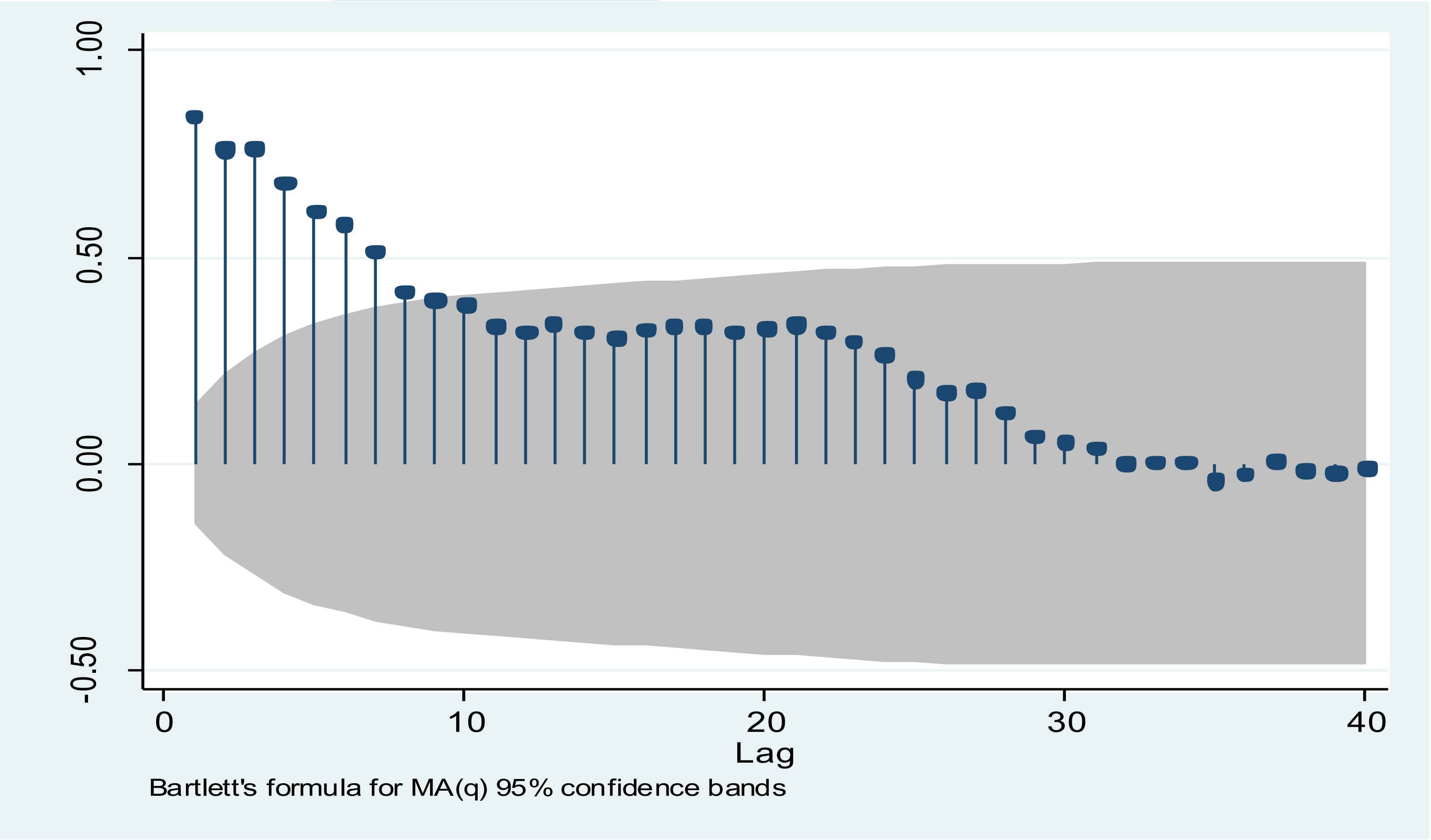
- the summation is over $t=j+1$ to T *(why?)*
- there is no divisor (e.g. *n* or *n-1* or *n-j-1*, etc)

What two things are implicitly assumed by this formula?

ACF plots: Inflation rate

We can put the correlation numbers into a plot

ac acpigrowth



Shows +- 2 SEs for each lag j

What does the plot tell us?

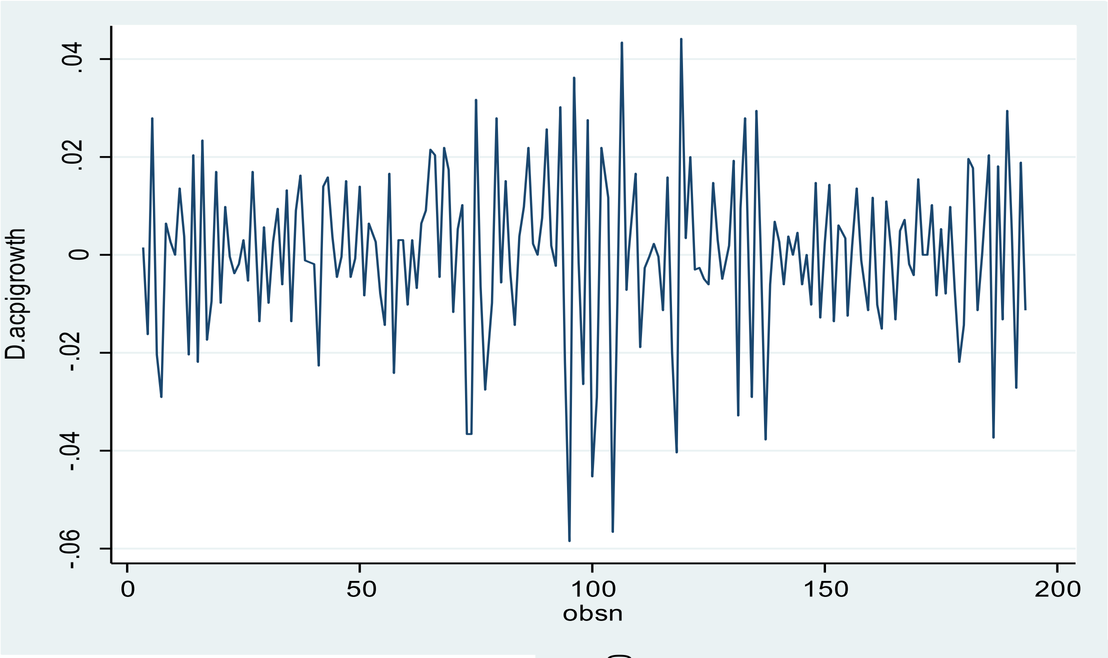
SEs for auto-correlation plots

$$SE(r_j) = \begin{cases} \dfrac{1}{\sqrt{n}}\ , & j = 1 \\ \dfrac{\sqrt{1 + 2\sum_{k=1}^{j-1} r_k^2}}{\sqrt{n}}\ , & j > 1 \end{cases}$$

$$t_{r_j} = \frac{r_j}{SE(r_j)} \quad \rightarrow \quad N(0,1) \quad \text{if } \rho_j = 0$$

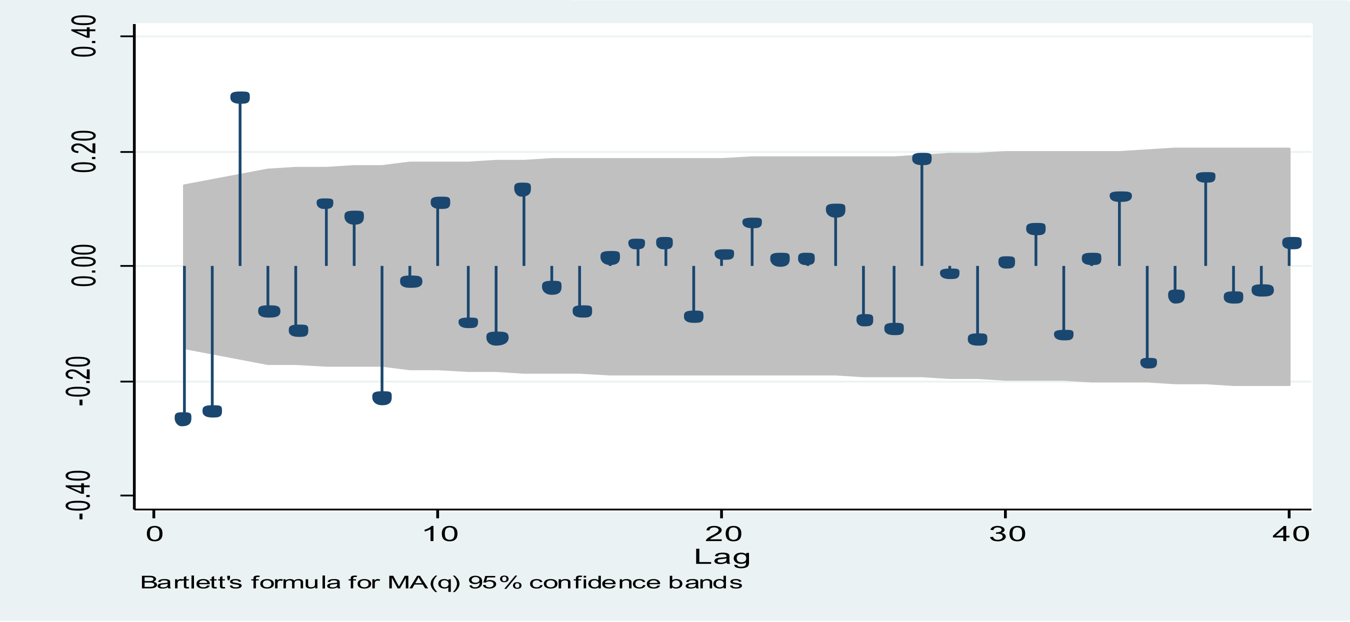
Here is your first test for autocorrelation !

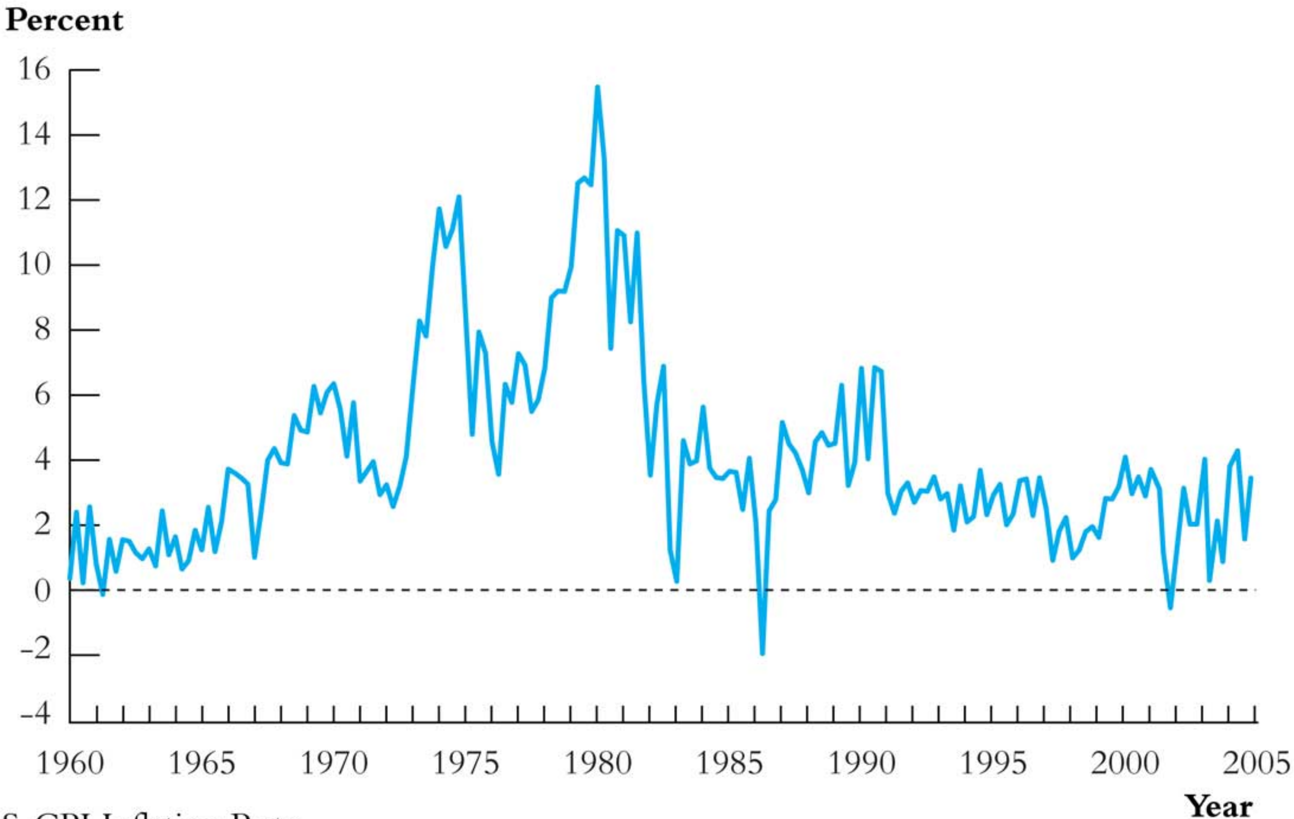
ACF plots: Changes in Inflation rate



What do these plots tell us?

tsline d.acpigrowth
ac d.acpigrowth

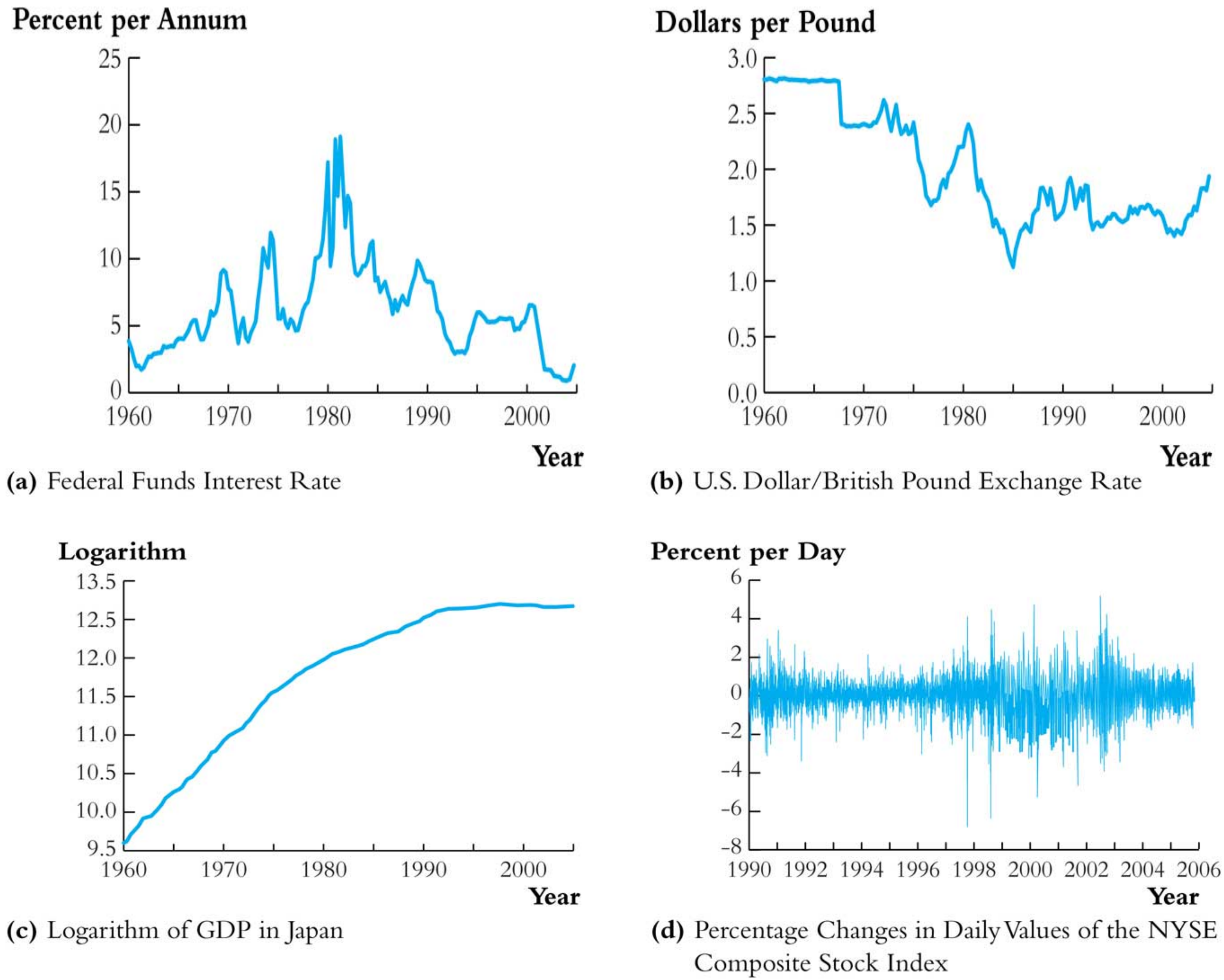




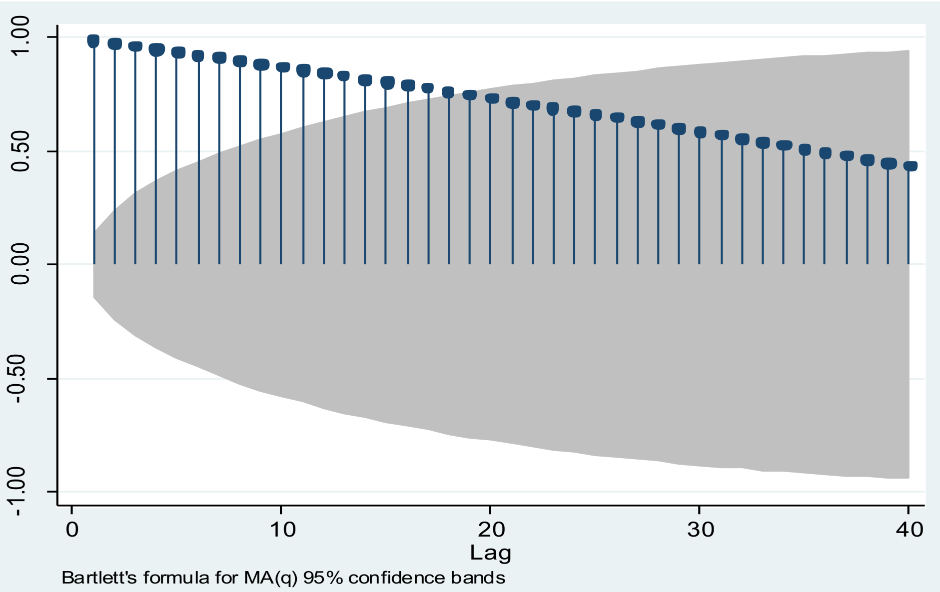
(a) U.S. CPI Inflation Rate

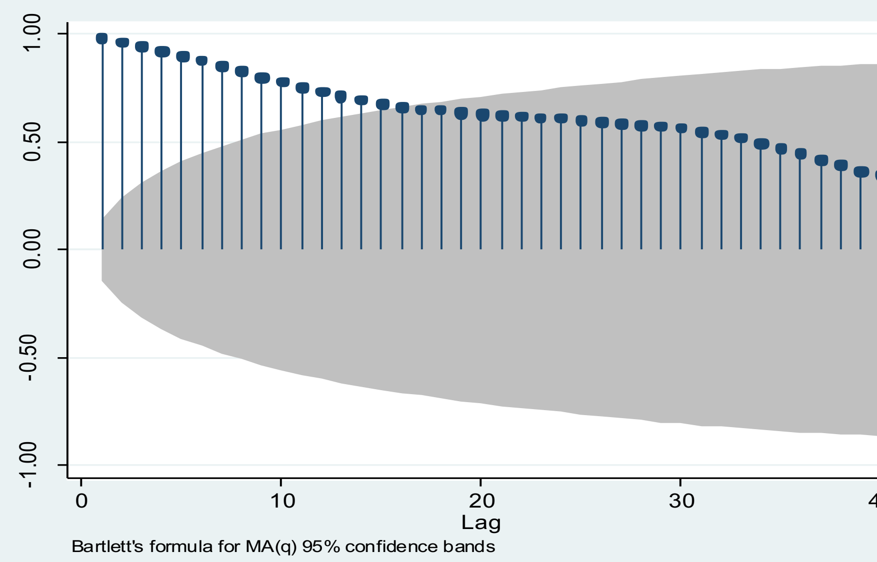
- Inflation rate is highly positively auto-correlated (e.g. $\rho_1 = .84$, t > 2)
  → Last quarter's inflation rate contains much information about this quarter's inflation rate (may not be causal!)
- Lots of persistent rises and falls and some surprise movements

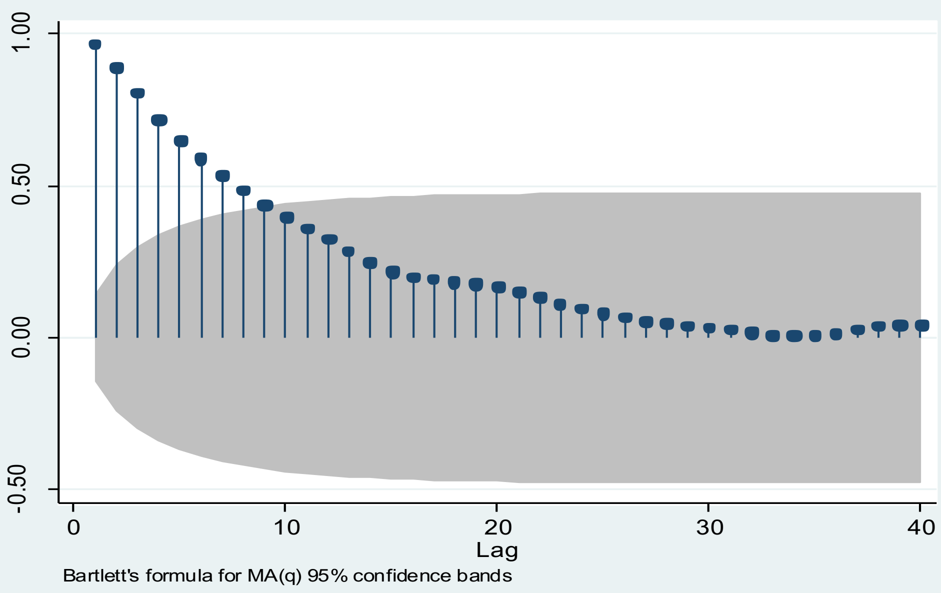
How would you characterize or describe these series?

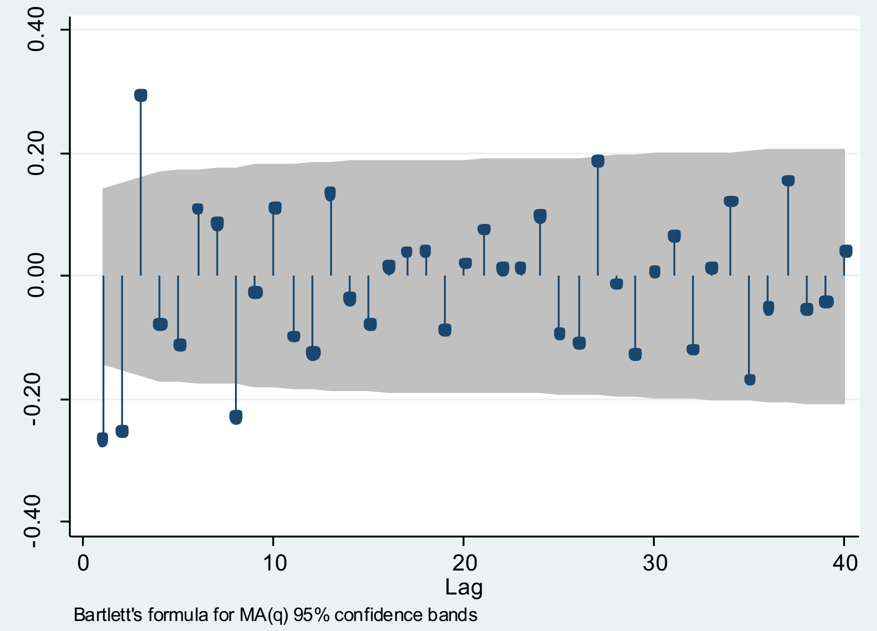


ACFs


1.00
0.50
0.00
-0.50
-1.00
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands


1.00
0.50
0.00
-0.50
-1.00
0
10
20
30
Lag
Bartlett's formula for MA(q) 95% confidence bands


1.00
0.50
0.00
-0.50
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands


0.40
0.20
0.00
-0.20
-0.40
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands

Stationarity: a key requirement for time series models

Definition for STRICT Stationarity:

The joint distribution (of the data) does not change over time.

i.e. the joint *distribution* of $y_t, y_{t+1}, \cdots, y_{t+k}$ is *independent* of t.

Definition for WEAK Stationarity:

A process is **weakly stationary** if its ***unconditional*** mean, variance and covariance functions (are finite and) do not change with time.

i.e.

$$E(Y_t) = \mu \quad \text{and} \quad Var(Y_t) = \sigma^2$$

AND $Cov(Y_{t_1}, Y_{t_2}) = Cov(Y_{t_1+k}, Y_{t_2+k}) = \gamma_{t_1-t_2}$

ACF of stationary data

A** <u>stationary</u> **in mean series has an ACF that 'dies down' or 'cuts off' reasonably quickly

*A **non-stationary** in mean process has an ACF that 'dies down' extremely slowly*

Weak stationarity

Almost all formal time series model estimators assume the observation process is **weakly stationary**

Often a **transformation** is required to better satisfy this assumption

Box and Jenkins (1976) advocate using the ACF (and PACF) plots to assess stationarity and identify a suitable model. A suitable variance stabilising transform SHOULD be done first.

Autoregressive models

A natural starting point for forecasting is to use past values of Y to forecast Y_{t+1}.

- ***Autoregression***: a regression model where Y_t is regressed against its lagged values.
- The number of lags used as regressors is called the ***order*** of the autoregression.
  - In a ***first order autoregression***, Y_t is regressed against Y_{t-1}
  - In a ***p^th^ order autoregression***, Y_t is regressed against *a subset of* $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p+1}$ AND Y_{t-p}.

The First Order Autoregressive AR(1) Model

The population AR(1) model is

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

- β_0 and β_1 *do not* have causal interpretations
- if $\beta_1 = 0$, Y_{t-1} is not useful for forecasting Y_t
- The AR(1) model can be estimated by OLS: Y_t against Y_{t-1}
- Testing $\beta_1 = 0$ v. $\beta_1 \neq 0$ provides a test of the hypothesis that Y_{t-1} is not useful for forecasting Y_t

The AR(1) Model assumptions

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

1. $E\left(\varepsilon_t \mid Y_{t-1}\right) = 0$

2. (a) The series $Y_1, Y_2, \ldots, Y_T$ is weakly stationary

   (b) Each pair of observations Y_{t-j}, Y_t tend to being independent as j gets large

3. Large outliers are rare, i.e. $E\left(Y_t^4\right) < \infty$

What aspects should we focus on when testing a time series model's fit to the data??

Example: AR(1) model for inflation

reg acpigrowth l.acpigrowth

```
      Source |       SS           df       MS      Number of obs   =       191
-------------+----------------------------------   F(1, 189)       =    446.07
       Model |  1198.12255         1  1198.12255   Prob > F        =    0.0000
    Residual |  507.643225       189  2.68594299   R-squared       =    0.7024
-------------+----------------------------------   Adj R-squared   =    0.7008
       Total |  1705.76578       190  8.97771461   Root MSE        =    1.6389

------------------------------------------------------------------------------
  acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  acpigrowth |
         L1. |   .8386746   .0397092    21.12   0.000     .7603444    .9170048
             |
       _cons |   .6460874   .1994396     3.24   0.001     .2526738    1.039501
------------------------------------------------------------------------------
```

Example: AR(1) model for inflation

Estimated using data from 1962:I – 2004:IV:

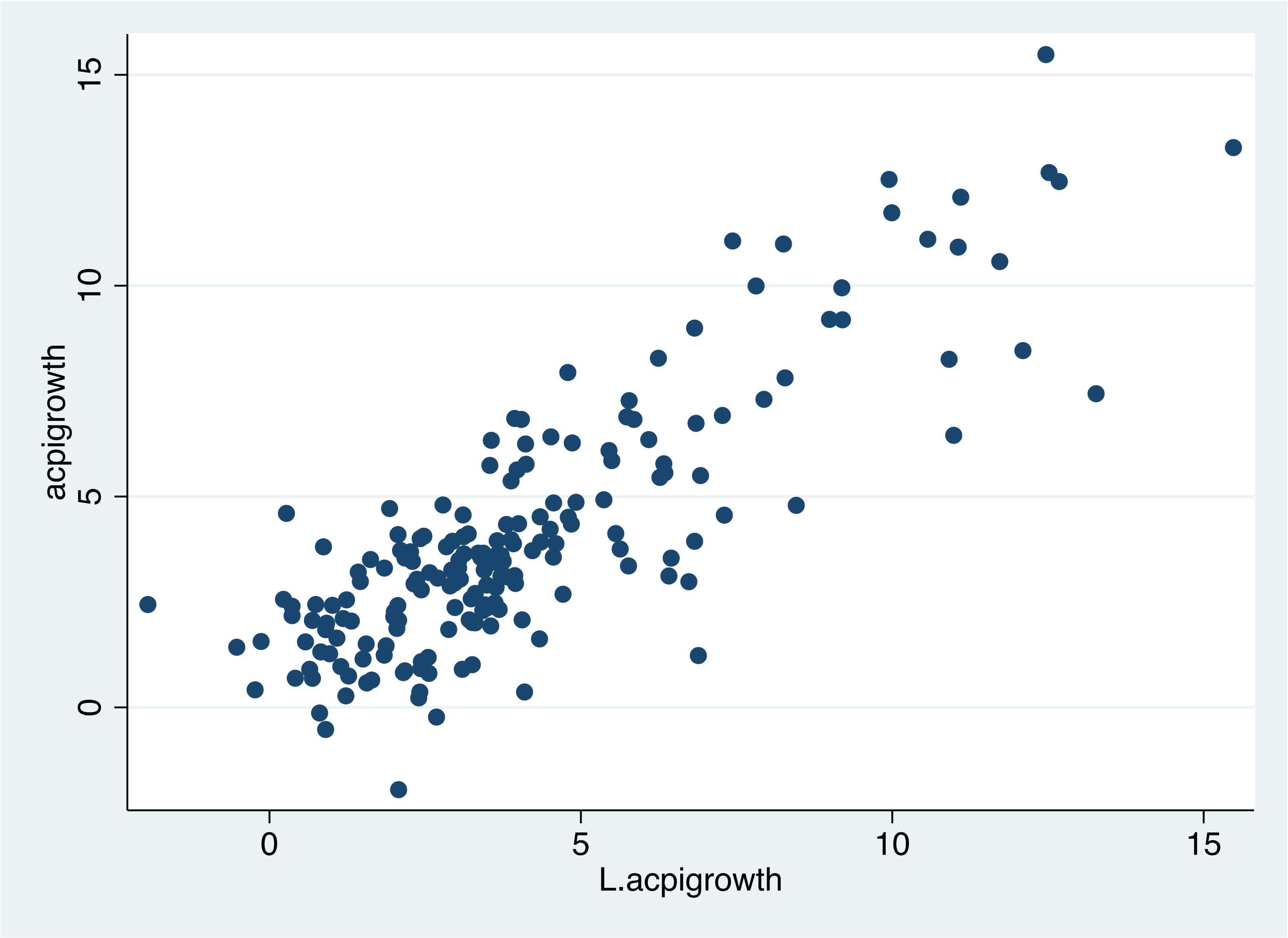
$$\widehat{Inf}_t = 0.646 - 0.839\, Inf_{t-1}\ ;\ \bar{R}^2 = 0.701;\ \text{SER}=1.639$$
$$\quad(0.20)\quad(0.040)$$

Is the lagged inflation a useful predictor of the current inflation?

- $t = 21.12 > 2$ (in absolute value, p-val ~=0)
- Reject H_0: $\beta_1 = 0$ at the 1% significance level
- $\bar{R}^2$ is quite reasonably high

- Yes, lagged inflation is a useful predictor of current inflation, but SER = 1.6% (is this too high?)

Scatterplots still work here!



White noise processes

- A sequence of *independently and identically distributed* random variables with mean 0 and finite variance σ^2

- What we hope and plan for the error component in a quantitative model

- What would the ACF plot look like for a white noise process?

Testing for Autocorrelation (in residuals)

- Breusch-Godfrey's LM test
  - Easily accommodates tests for higher order autocorrelation
  - Valid in presence of lagged dependent variables
  - Stata command: `estat bgodfrey`

- Ljung-Box Q test
  - Test the null that the correlation coefficients between lags are all zero
  - Need to determine the highest order of the autocorrelation in the test
  - Stata command: `wntestq` or `corrgram`

- Durbin-Watson test (DW-test) – used to be very popular
  - Only refers to a static model with AR(1) errors and is not applicable to models with lagged y's
  - Assumes a Gaussian error distribution!
  - Stata command: `estat dwatson`
  - STATA does NOT give a p-value!

Breusch-Godfrey (BG) tests

Suppose

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \ldots.. + \rho_p \varepsilon_{t-p} + u_t$$

and want to test

$$H_0 : \rho_1 = \rho_2 = \ldots. = \rho_p = 0$$

$$p-val = \Pr(\chi_p^2 > (n-p)R^2)$$

As ε_t *is unknown, use OLS residuals in auxiliary regression*

(i) Regress Y on X to yield OLS residuals, $\hat{\varepsilon}_t = Y_t - X_t \hat{\beta}$

(ii) Regress $\hat{\varepsilon}_t$ on all the X's and on $\hat{\varepsilon}_{t\text{-}1},\ \hat{\varepsilon}_{t\text{-}2},\ \hat{\varepsilon}_{t\text{-}3}$ etc.

Suppose we've included p lags on RHS.

(iii) For large n, under H_0 , $(n-p) * R^2 \sim \chi^2$ with p degrees of freedom

(iv) Regressors can include lagged dependent variables, i.e. $y_{t\text{-}1}, y_{t-2}, y_{t-3}$

Is there any remaining auto-correlation in the residuals from the Inflation AR(1) model ?

```
estat bgodfrey
estat bgodfrey, lags(4)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	7.685	1	0.0056

H0: no serial correlation

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
4	30.360	4	0.0000

H0: no serial correlation

Why choose to test 4 lags?

Ljung-Box Q test

Assume $\rho_0, \rho_1, \ldots, \rho_p$ are the correlation coefficients between lag 1 to p in the residules.

$H_0: \rho_0 = \rho_1 = \cdots = \rho_p = 0$

Under H_0, test statistic $Q_{LB} = n(n+2)\sum_{j=1}^{p}\frac{\hat{\rho}_j^2}{n-j} \xrightarrow{d} \chi^2\,(p)$

p is chosen, so that it is large enough to cover potential higher order autocorrelation, and as small as possible (compared with sample size n) to be close to χ^2 distribution.

The default choice for p in Stata is $\min\{floor(n/2) - 2{,}40\}$, where $floor(n/2)$ calculates the largest integer that no larger than $n/2$.

LB Q test on residuals

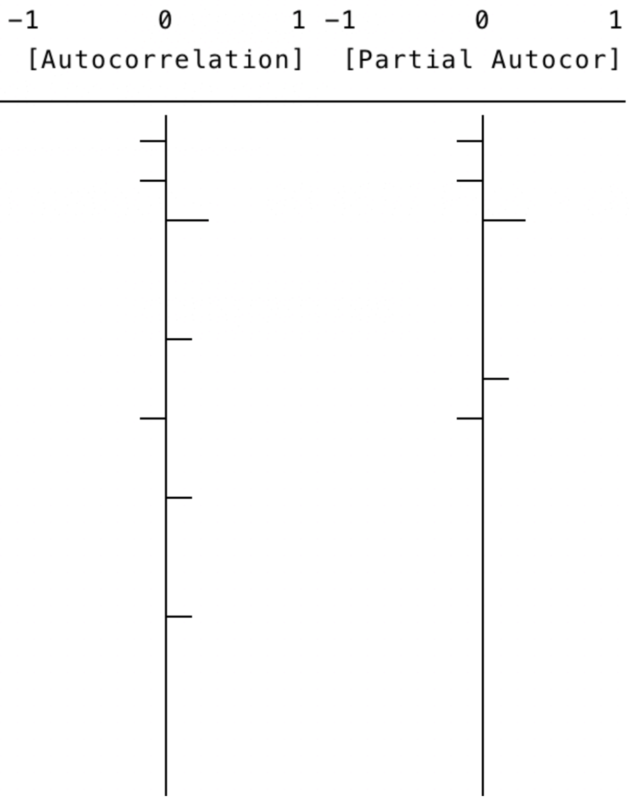
predict res, resid
wntestq, res

Portmanteau test for white noise

Portmanteau (Q) statistic = 99.4697 Prob > chi2(40) = 0.0000

corrgram res

LAG	AC	PAC	Q	Prob>Q
1	-0.1680	-0.1685	5.4756	0.0193
2	-0.1623	-0.1968	10.615	0.0050
3	0.3351	0.2972	32.633	0.0000
4	-0.0110	0.0629	32.657	0.0000
5	-0.0466	0.0618	33.087	0.0000
6	0.1511	0.0784	37.634	0.0000
7	0.1245	0.1689	40.741	0.0000
8	-0.1724	-0.1280	46.726	0.0000
9	0.0114	-0.0842	46.752	0.0000
10	0.1362	-0.0009	50.532	0.0000
11	-0.0608	0.0270	51.289	0.0000
12	-0.0843	-0.0813	52.754	0.0000
13	0.1548	0.0901	57.716	0.0000
14	-0.0031	0.0539	57.718	0.0000
15	-0.0443	0.0953	58.129	0.0000
16	0.0424	-0.0634	58.509	0.0000
17	0.0648	0.0549	59.398	0.0000



Durbin-Watson Test

$$r_1 = Corr\left(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}\right) = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2}$$

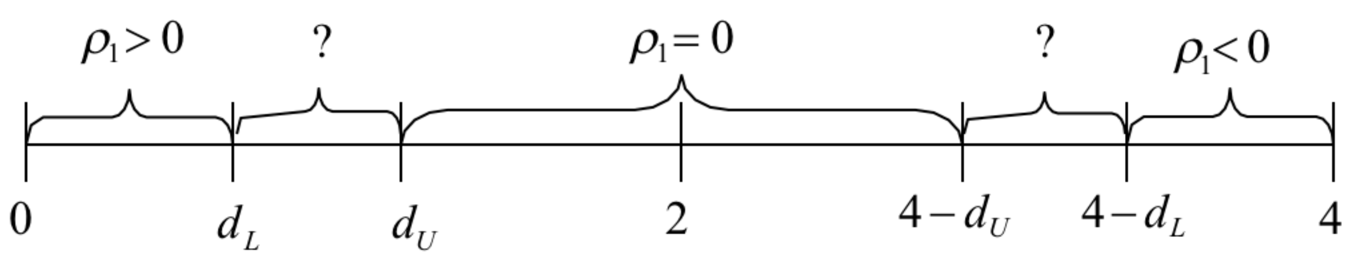
$$d(\equiv \mathrm{DW}) = \frac{\sum \hat{\varepsilon}_t^2 + \sum \hat{\varepsilon}_{t-1}^2 - 2\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2}$$

$$\therefore d \cong 2(1 - r_1)$$

$$\text{when } r_1 \approx 1 \quad \Rightarrow \quad d \approx 0$$

$$\text{when } r_1 \approx -1 \quad \Rightarrow \quad d \approx 4$$

$$\text{when } r_1 \approx 0 \quad \Rightarrow \quad d \approx 2$$

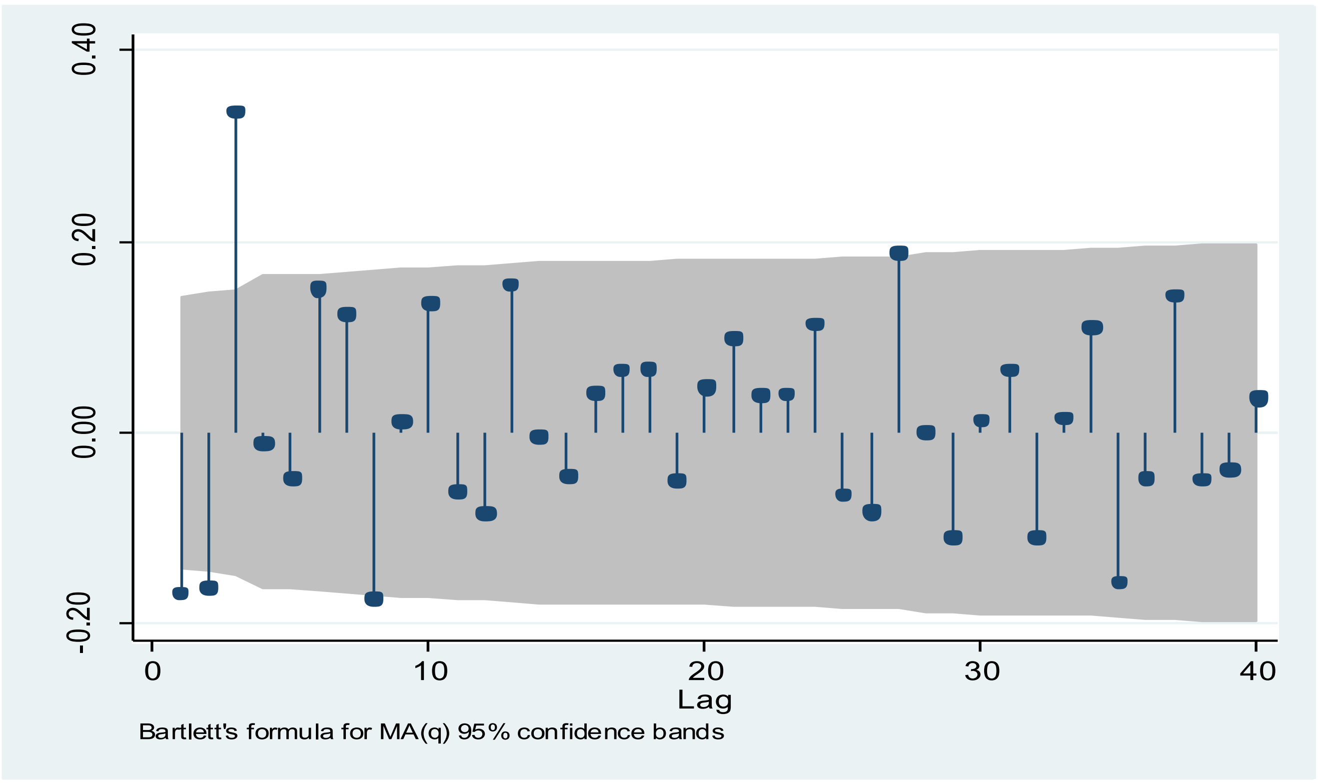


Rule of thumb, cannot reject null if 1.75<d<2.15

DW test on residuals

estat dwatson
ac res

Durbin-Watson d-statistic(2,   191) = 2.333052



AR(p): using multiple lags

The p^{th} order autoregressive model (AR(p)) is

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

- The AR(p) model uses p lags of Y as regressors
- To test the hypothesis that $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ do not help forecast Y_t, use an F-test
- Can use t- or F-tests to determine the lag order p
- Or, better, determine p using an "information criterion" (*more on this later…*)

Example: AR(4) model of inflation – STATA

```
reg  acpigrowth l(1/4).acpigrowth
```

```
      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  4,   183) =  139.78
       Model |  .128127703     4  .032031926           Prob > F      =  0.0000
    Residual |  .041935207   183  .000229154           R-squared     =  0.7534
-------------+------------------------------           Adj R-squared =  0.7480
       Total |  .170062911   187  .000909427           Root MSE      =  .01514

------------------------------------------------------------------------------
  acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  acpigrowth |
         L1. |   .6722326   .0725569     9.26   0.000      .529077    .8153882
         L2. |  -.0190465   .0818239    -0.23   0.816     -.180486    .1423931
         L3. |   .4209105   .0823805     5.11   0.000     .2583728    .5834481
         L4. |  -.1764868   .0729409    -2.42   0.017    -.3204001   -.0325736
             |
       _cons |   .0040124   .0019563     2.05   0.042     .0001527    .0078721
```

NOTES

- `L(1/4).` `acpigrowth` *is a convenient way to say "use lags 1–4 of inf as regressors"*
- *L1,…,L4 refer to the first, second,… 4th lags of* `acpigrowth`

Example: AR(4) model of inflation

$$\hat{Inf}_t = .004 + .67Inf_{t-1} - .02Inf_{t-2} + .42Inf_{t-3} - .18Inf_{t-4},$$
$$\quad (.002) \quad (.07) \qquad\quad (.08) \qquad\quad (.08) \qquad\quad (.07)$$

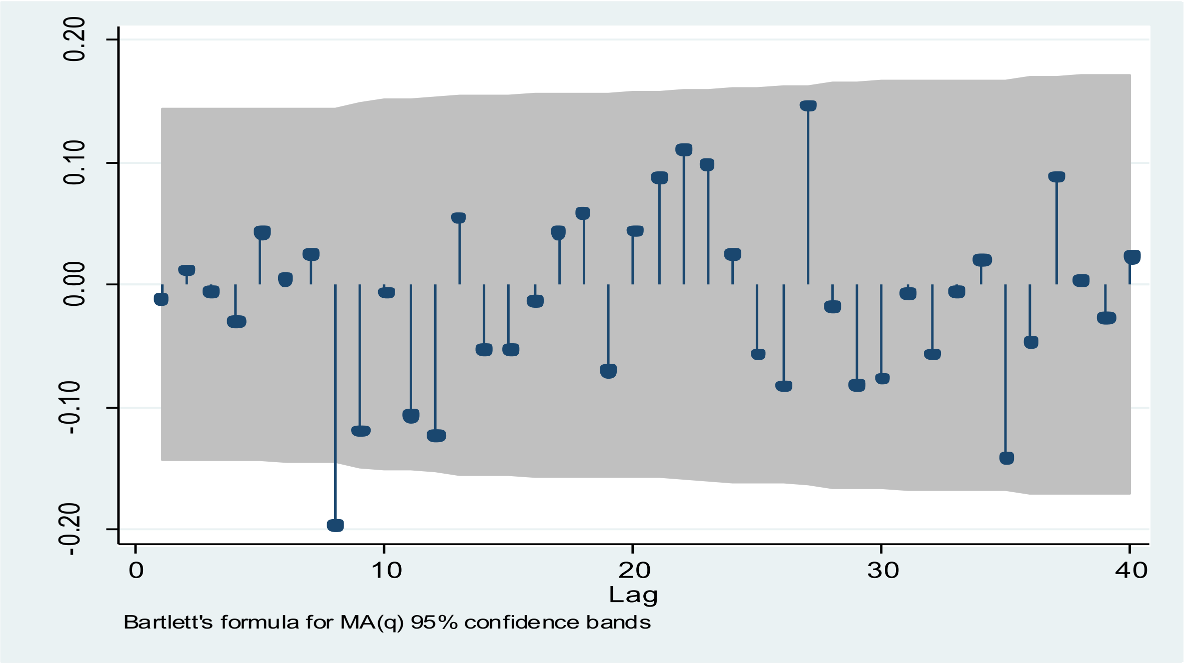
$$\bar{R}^2 = 0.748 \text{ , SER=0.015}$$

- F-statistic testing lags 2, 3, 4 is 11.55 (p-value ~= 0)
- $\bar{R}^2$ increased from .70 to .75 by adding lags 2, 3, 4
- So, lags 2, 3, 4 (jointly) help to predict inflation – both in a statistical and in a practical sense.

AR(4) model of inflation – Test

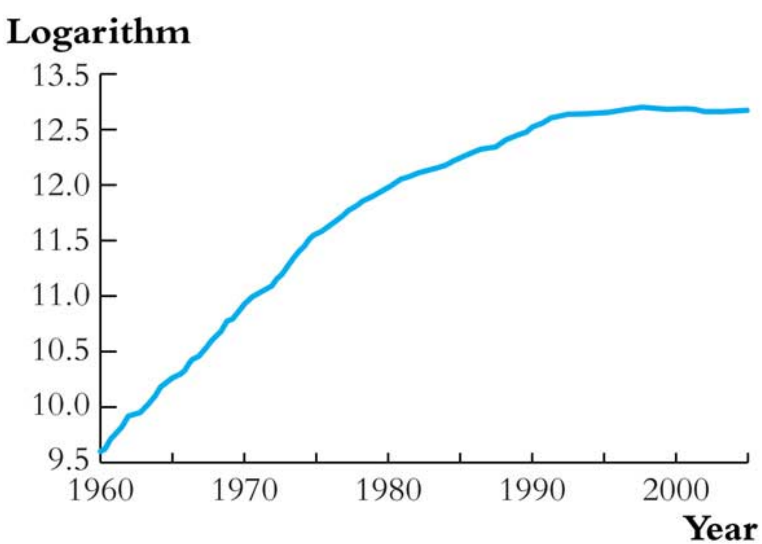
```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)    |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4       |          3.473               4                   0.4820
---------------------------------------------------------------------------
                        H0: no serial correlation

Durbin-Watson d-statistic(  5,   187) =  2.003734

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    39.2847
 Prob > chi2(40)           =     0.5023
```

```
estat bgodfrey, lags(4)
estat dwatson
predict res4, resid
wntestq res4
ac res4
```

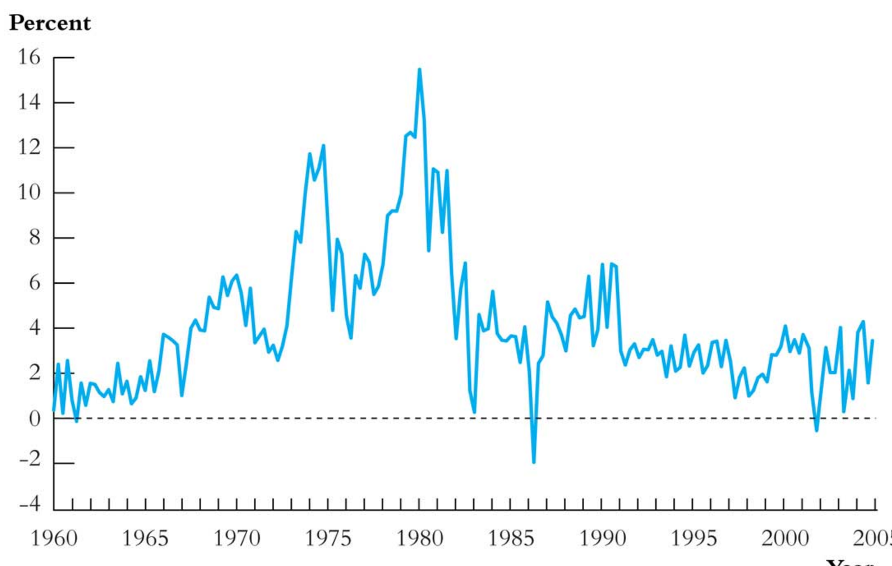


Nonstationarity: Trends

- Two important types of nonstationarity, in practice, are:
  - Trends
  - Structural breaks (model instability)
- A trend is a long-term movement or tendency in the data.
- Trends need not be just a straight line!
- Which of the following two series has a trend?



(c) Logarithm of GDP in Japan



(a) U.S. CPI Inflation Rate

Deterministic and stochastic trends

- A **deterministic trend** is a nonrandom function of time (e.g. $y_t = t$, or $y_t = t^2$).
- A **stochastic trend** is random and varies over time
- An important example of a stochastic trend is a **random walk**:

$$Y_t = Y_{t-1} + u_t, \text{ where } u_t \text{ is serially uncorrelated}$$

If Y_t follows a random walk, then the value of Y tomorrow is the value of Y today, plus an unpredictable disturbance.

Deterministic and stochastic trends, ctd.

Two key features of a random walk:

(i) $Y_{T+h|T} = Y_T$

- Your best prediction of the value of Y in the future is the value of Y today
- To a first approximation, log stock prices follow a random walk (more precisely, stock returns are unpredictable)

(ii) $\text{var}(Y_{T+h|T} - Y_T) = h\sigma_u^2$

- The variance of your forecast error increases linearly in the horizon. The more distant your forecast, the greater the forecast uncertainty. (Technically this is the sense in which the series is "nonstationary")

Deterministic and stochastic trends, ctd.

A **random walk with drift** is

$$Y_t = \beta_0 + Y_{t-1} + u_t, \text{ where } u_t \text{ is serially uncorrelated}$$

The "drift" is β_0: If $\beta_0 \neq 0$, then Y_t follows a random walk around a linear trend. You can see this by considering the h-step ahead forecast:

$$Y_{T+h|T} = \beta_0 h + Y_T$$

The random walk model (with or without drift) is a good description of stochastic trends in many economic time series.

If Y_t has a random walk trend, then ΔY_t is stationary and regression analysis should be undertaken using ΔY_t instead of Y_t.

Stochastic trends and unit autoregressive roots

Random walk (with drift): $Y_t = \beta_0 + Y_{t-1} + u_t$

AR(1): $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

- The random walk is an AR(1) with $\beta_1 = 1$.
- The special case of $\beta_1 = 1$ is called a unit root*.
- When $\beta_1 = 1$, the AR(1) model becomes

$$\Delta Y_t = \beta_0 + u_t$$

*This terminology comes from considering the equation

$1 - \beta_1 z = 0$ – the "root" of this equation is $z = 1/\beta_1$, which equals one (unity) if $\beta_1 = 1$.

Unit roots in an AR(2)

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

Use the "rearrange the regression" trick:

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

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

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

Subtract Y_{t-1} from both sides:

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

or

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + u_t,$$

where $\delta = \beta_1 + \beta_2 - 1$ and $\gamma_1 = -\beta_2$.

Unit roots in an AR(2), ctd.

If $1 - \beta_1 z - \beta_2 z^2 = 0$ has a unit root, then $\beta_1 + \beta_2 = 1$ (*you can show this yourself!*)

Thus, if there is a unit root, then $\delta = 0$ and the AR(2) model becomes,

$$\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + u_t$$

If an AR(2) model has a unit root, then it can be written as an AR(1) in first differences.

AR(p): $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t$$

This regression can be rearranged as,

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

Unit roots in the AR(p) model

$$\delta = \beta_1 + \beta_2 + \ldots + \beta_p - 1$$
$$\gamma_1 = -(\beta_2 + \ldots + \beta_p)$$
$$\gamma_2 = -(\beta_3 + \ldots + \beta_p)$$
$$\ldots$$
$$\gamma_{p-1} = -\beta_p$$

The AR(p) model can be written as,

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

where $\delta = \beta_1 + \beta_2 + \ldots + \beta_p - 1$.

If there is a unit root in the AR(p) model, then $\delta = 0$ and the AR(p) model becomes an AR(p–1) model in first differences:

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

What problems are caused by trends?

There are three main problems with stochastic trends:

1. AR coefficients can be badly biased towards zero. This means that if you estimate an AR and make forecasts, if there is a unit root then your forecasts can be poor (AR coefficients biased towards zero)

2. Some *t*-statistics don't have a standard normal distribution, even in large samples (more on this later)

3. If *Y* and *X* both have random walk trends then they can look related even if they are not – you can get "spurious regressions." Here is an example…

How do we detect trends?

1. Plot the data.
2. There is a regression-based test for a random walk – the Dickey-Fuller test for a unit root.

<u>The Dickey-Fuller test in an AR(1)</u>

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

or

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

H_0: $\delta = 0$ (that is, $\beta_1 = 1$) v. H_1: $\delta < 0$

(note: this is 1-sided: $\delta < 0$ means that Y_t is stationary)

DF test in AR(1), ctd.

- Under H_0, this t statistic does **NOT** have a normal distribution!
- You need to compare the t-statistic to the table of Dickey-Fuller critical values. There are two cases:

(a) $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$ (intercept only)

(b) $\Delta Y_t = \beta_0 + \mu t + \delta Y_{t-1} + u_t$ (intercept & time trend)

- The two cases have different critical values!
- Reject if the DF t-statistic (the t-statistic testing $\delta = 0$) is less than the specified critical value. This is a 1-sided test.

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

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

The Dickey-Fuller test in an AR(p)

In an AR(p), the DF test is based on the rewritten model,

$$Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t \qquad (*)$$

where $\delta = \beta_1 + \beta_2 + \ldots + \beta_p - 1$. If there is a unit root (random walk trend), $\delta = 0$; if the AR is stationary, $\delta < 1$.

<u>The DF test in an AR(p) (intercept only):</u>

1. Estimate (*), obtain the t-statistic testing $\delta = 0$
2. Reject the null hypothesis of a unit root if the t-statistic is less than the DF critical value in Table 14.5

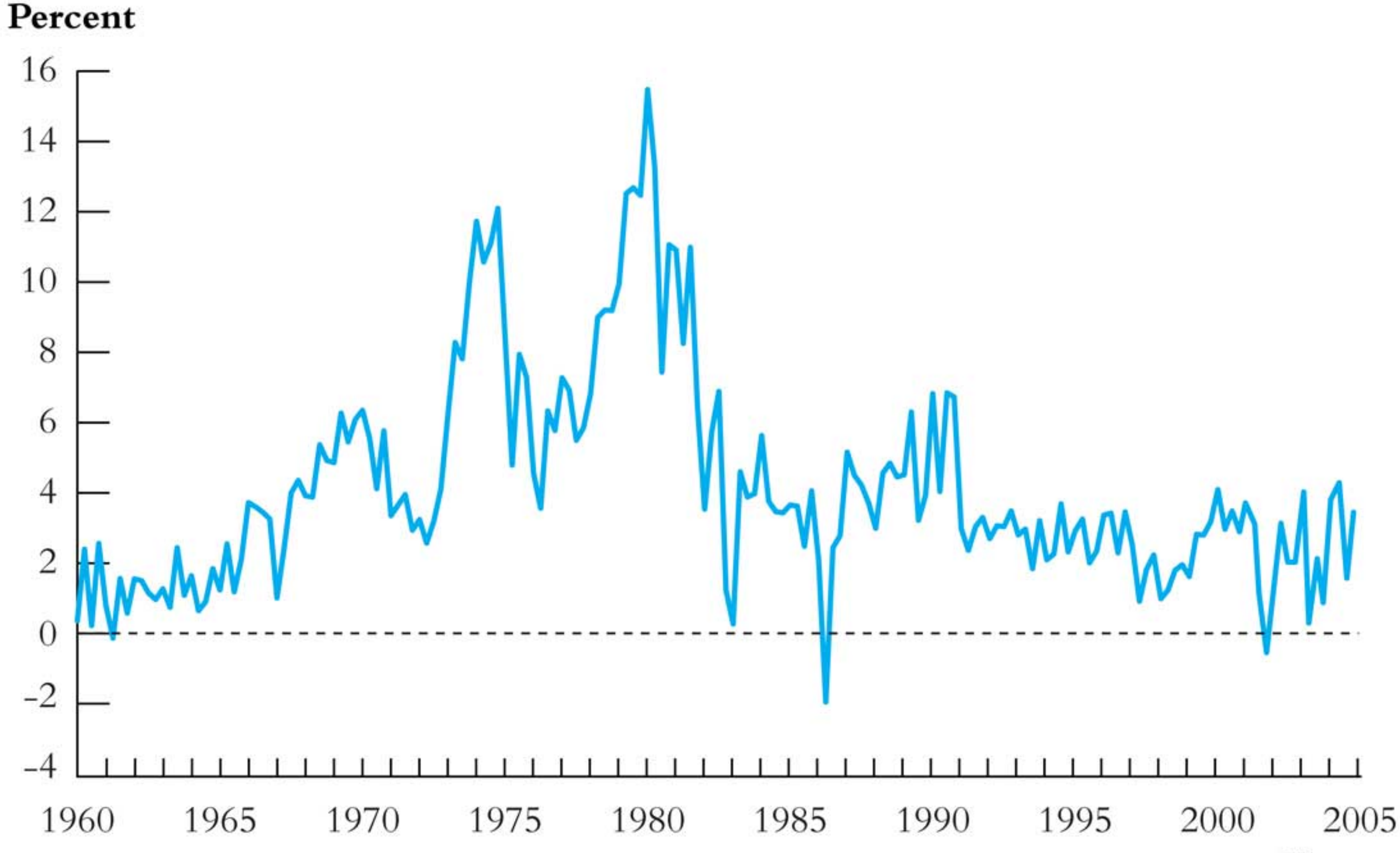
<u>Modification for time trend</u>: include t as a regressor in (*)

When should you include a time trend in the DF test?

The decision to use the intercept-only DF test or the intercept & trend DF test depends on what the alternative is – and what the data look like.

- In the intercept-only specification, the alternative is that Y is stationary around a constant
- In the intercept & trend specification, the alternative is that Y is stationary around a linear time trend.

Example: Does U.S. inflation have a unit root?



(a) U.S. CPI Inflation Rate

The alternative is that inflation is stationary around a constant

Example: Does U.S. inflation have a unit root? ctd

DF test for a unit root in U.S. inflation – using p = 4 lags

```
. reg dinf L.inf L(1/4).dinf if tin(1962q1,2004q4);

      Source |       SS       df       MS              Number of obs =     172
-------------+------------------------------           F(  5,   166) =   10.31
       Model |  118.197526     5  23.6395052           Prob > F      =  0.0000
    Residual |  380.599255   166   2.2927666           R-squared     =  0.2370
-------------+------------------------------           Adj R-squared =  0.2140
       Total |  498.796781   171  2.91694024           Root MSE      =  1.5142

------------------------------------------------------------------------------
        dinf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inf |
         L1. |  -.1134149   .0422339    -2.69   0.008    -.1967998     -.03003
        dinf |
         L1. |  -.1864226   .0805141    -2.32   0.022    -.3453864   -.0274589
         L2. |   -.256388   .0814624    -3.15   0.002     -.417224   -.0955519
         L3. |    .199051   .0793508     2.51   0.013     .0423842    .3557178
         L4. |   .0099822   .0779921     0.13   0.898     -.144002    .1639665
       _cons |   .5068071    .214178     2.37   0.019     .0839431     .929671
------------------------------------------------------------------------------
```

DF *t*-statstic = –2.69

Don't compare this to –1.645 – use the Dickey-Fuller table!

DF *t*-statstic = –2.69 (intercept-only):

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

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

–2.69 rejects a unit root at 10% level but not the 5% level

- Some evidence of a unit root – not clear cut.
- This is a topic of debate – what does it mean for inflation to have a unit root?
- We model inflation as having a unit root.

How to address and mitigate problems raised by trends

If Y_t has a unit root (has a random walk stochastic trend), the easiest way to avoid the problems this poses is to model Y_t in first differences.

- In the AR case, this means specifying the AR using first differences of Y_t (ΔY_t)
- This is what we did in our initial treatment of inflation – the reason was that inspection of the plot of inflation, plus the DF test results, suggest that inflation plausibly has a unit root – so we estimated the ARs using ΔInf_t

Summary: detecting and addressing stochastic trends

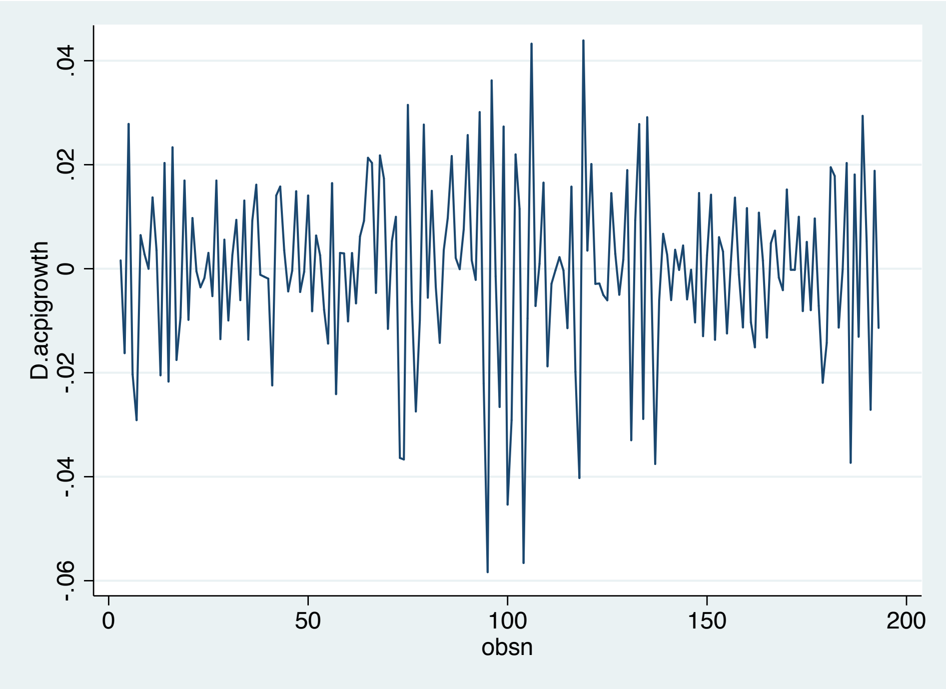
1. The random walk model is the workhorse model for trends in economic time series data

2. To determine whether Y_t has a stochastic trend, first plot Y_t, then if a trend looks plausible, compute the DF test (decide which version, intercept or intercept+trend)

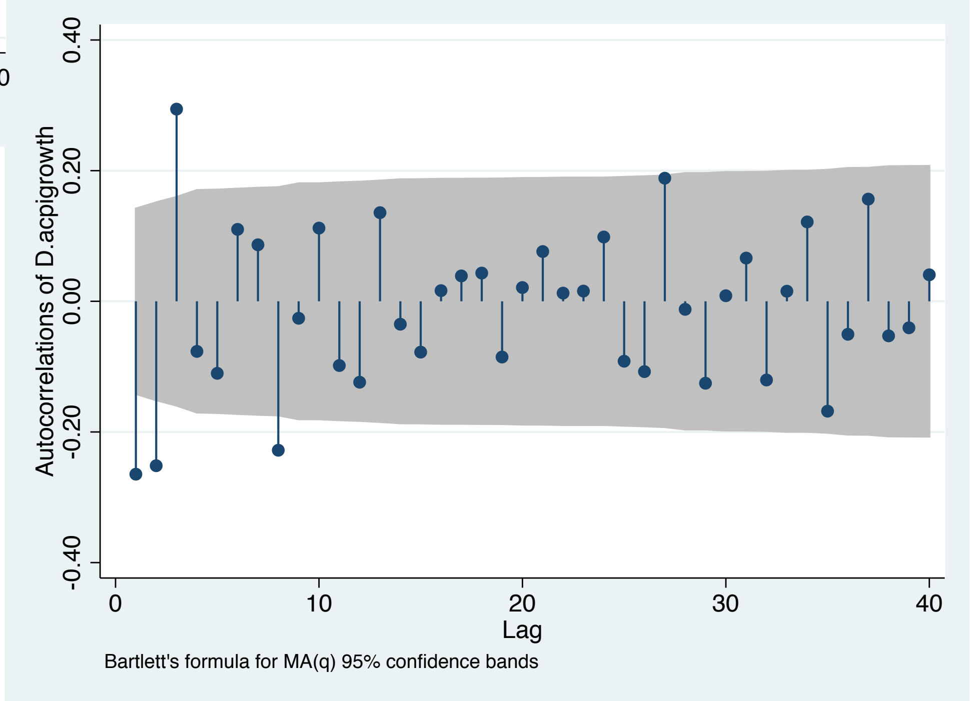
3. If the DF test fails to reject, conclude that Y_t has a unit root (random walk stochastic trend)

4. If Y_t has a unit root, use ΔY_t for regression analysis and forecasting.

Eg: TS plot and ACF for change in inflation



tsline d.acpigrowth
ac d.acpigrowth



Example: AR(1) model for Δinflation

reg d.acpigrowth l.d.acpigrowth

```
. reg  d.acpigrowth l.d.acpigrowth

      Source |       SS       df       MS              Number of obs =     190
-------------+------------------------------           F(1, 188)     =   14.20
       Model |  38.7624957     1  38.7624957           Prob > F      =  0.0002
    Residual |  513.185424   188   2.7297097           R-squared     =  0.0702
-------------+------------------------------           Adj R-squared =  0.0653
       Total |   551.94792   189  2.92035937           Root MSE      =  1.6522

------------------------------------------------------------------------------
D.acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  acpigrowth |
         LD. |  -.2653108   .0704056    -3.77   0.000    -.4041973   -.1264242
             |
       _cons |   -.006096    .119862    -0.05   0.959    -.2425433    .2303512
------------------------------------------------------------------------------
```

Eg: AR(1) model for change in inflation

```
reg  d.acpigrowth l.d.acpigrowth
```

Estimated using data from 1962:I – 2004:IV:

$$\Delta \hat{Inf}_t = \text{-}0.006 - 0.265\Delta Inf_{t-1} \;\; ; \;\; \bar{R}^2 = 0.05, \text{ SER}=1.653$$

$$\qquad (0.120) \; (0.070)$$

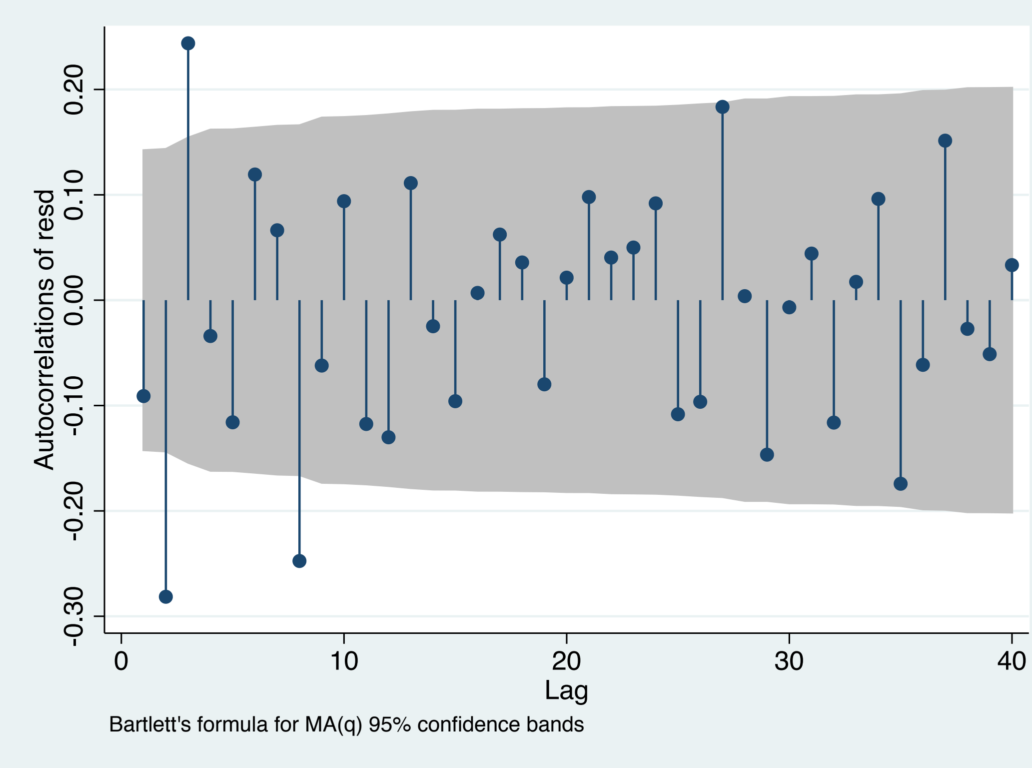
Is the lagged change in inflation a useful predictor of the current change in inflation?

- $t = -3.77 > 1.96$ (i.e. p-val < 0.05)
- Reject H_0: $\beta_1 = 0$ at the 5% significance level
- Yes, the lagged change in inflation is a useful predictor of current change in inflation–but the $\bar{R}^2$ is pretty low!

AR(1) model for Δinflation– Test

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4     |         27.842               4                   0.0000
---------------------------------------------------------------------------
                        H0: no serial correlation

Durbin-Watson d-statistic(  2,   190) =  2.176543

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =   105.7256
 Prob > chi2(40)           =     0.0000
```

```
estat bgodfrey, lags(4)
estat dwatson
predict resd, resid
wntestq resd
ac resd
```



Example: AR(4) model of Δinflation

```
reg d.acpigrowth d.l(1/4).acpigrowth
```

```
------------------------------------------------------------------------------
D.acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
  acpigrowth |
         LD. |  -.2827175   .0739956    -3.82   0.000     -.428717    -.13671
        L2D. |  -.3102006   .0760401    -4.08   0.000    -.4602341   -.160167
        L3D. |   .1367869   .0763387     1.79   0.075    -.0138357    .287409
        L4D. |  -.0405382   .0739884    -0.55   0.584    -.1865235    .105447
             |
       _cons |  -.0000451   .0011251    -0.04   0.968    -.0022651    .002174
------------------------------------------------------------------------------
```

```
test L2.d.acpigrowth L3.d.acpigrowth L4.d.acpigrowth

      ( 1)  L2D.acpigrowth = 0
      ( 2)  L3D.acpigrowth = 0
      ( 3)  L4D.acpigrowth = 0

            F(  3,   182) =   10.21
                 Prob > F =    0.0000
```

Example: AR(4) model of Δinflation

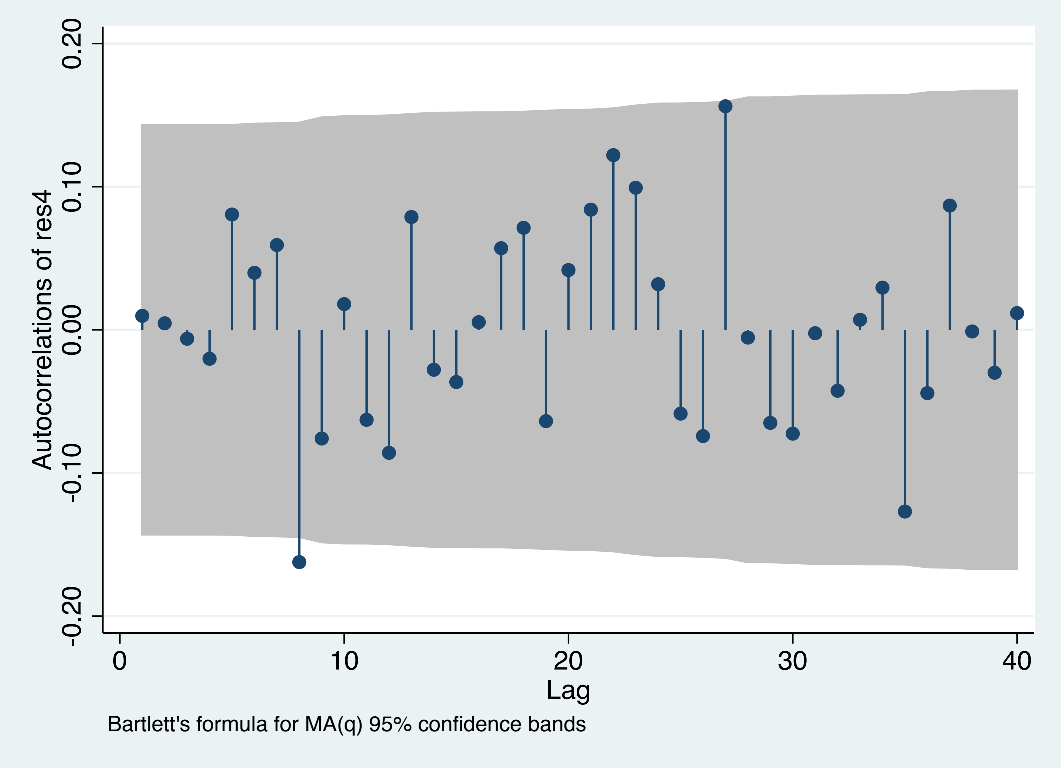
$$\Delta \widehat{Inf}_t = .02 - .26\Delta Inf_{t-1} - .32\Delta Inf_{t-2} + .16\Delta Inf_{t-3} - .03\Delta Inf_{t-4},$$
$$\quad (.12)\ (.09) \qquad\quad (.08) \qquad\quad (.08) \qquad\quad (.09)$$

$$\bar{R}^2 = 0.181 \quad ; \quad \text{SER} = 0.0154$$

- F-statistic testing lags 2, 3, 4 is 6.91 (p-value < .001)
- $\bar{R}^2$ increased from .05 to .18 by adding lags 2, 3, 4
- So, lags 2, 3, 4 (jointly) help to predict the change in inflation – both in a statistical sense (are statistically significant) and in a practical sense (substantial increase in the $\bar{R}^2$)

AR(4) model of Δinflation – Test

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)   |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4      |          3.473               4                   0.4820
---------------------------------------------------------------------------
                        H0: no serial correlation

Durbin-Watson d-statistic(  5,   187) =  2.003734

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    45.1889
 Prob > chi2(40)           =     0.2641
```




```
estat bgodfrey, lags(4)
estat dwatson
predict resd4, resid
wntestq resd4
ac res4
```

So which to use: ΔInf_t OR Inf_t?

$$\Delta Inf_t = \beta_0 + \beta_1 \Delta Inf_{t-1} + u_t$$

or

$$Inf_t - Inf_{t-1} = \beta_0 + \beta_1 (Inf_{t-1} - Inf_{t-2}) + u_t$$

or

$$Inf_t = Inf_{t-1} + \beta_0 + \beta_1 Inf_{t-1} - \beta_1 Inf_{t-2} + u_t$$

$$= \beta_0 + (1+\beta_1) Inf_{t-1} - \beta_1 Inf_{t-2} + u_t$$

Note that the AR(2) model here is **restricted** and has **only 1 slope** parameter

So which to use: ΔInf_t OR Inf_t?

- When Y_t is strongly, positively auto-correlated, the OLS estimator is biased towards 0 (see next slide).
- In the extreme case that $\beta_1 = 1$, Y_t isn't stationary: the ε_t's accumulate and Y_t "blows up".
- If Y_t isn't stationary, the estimation theory we are working with here doesn't really work (re: 3rd LSA)
- Some textbooks recommend using ΔY whenever Y is strongly correlated
- Using ΔY is **<u>necessary</u>** if the $\beta_1 = 1$.
- If Y_t is stationary, we call it integrated of order 0 (I(0)). If ΔY is stationary, we call it integrated of order 1 (I(1)).

Time Series Regression: Autoregressive Distributed Lag (ADL) Model

- We can add regressors (X) to AR models:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \ldots + \delta_r X_{t-r} + u_t$$

- This is an ***ADL model*** with p lags of Y and r lags of X … ***ADL(p,r)*.**

- Does it make sense to have lagged Xs as well as lagged Ys?

The ADL(p,r) Model assumptions

1. $$E\left(\varepsilon_t \mid Y_{t-1},\ldots,Y_{t-p},X_{t-1},\ldots,X_{t-r}\right)=0$$

2. (a) The variables X, Y are both weakly stationary

   (b) The pairs of observations (Y_{t-j}, X_{t-j}) and (Y_t, X_t) become independent as j gets large

3. Large outliers in Y, X are rare, i.e. $E\left(Y_t^4\right)<\infty; E\left(X_t^4\right)<\infty$

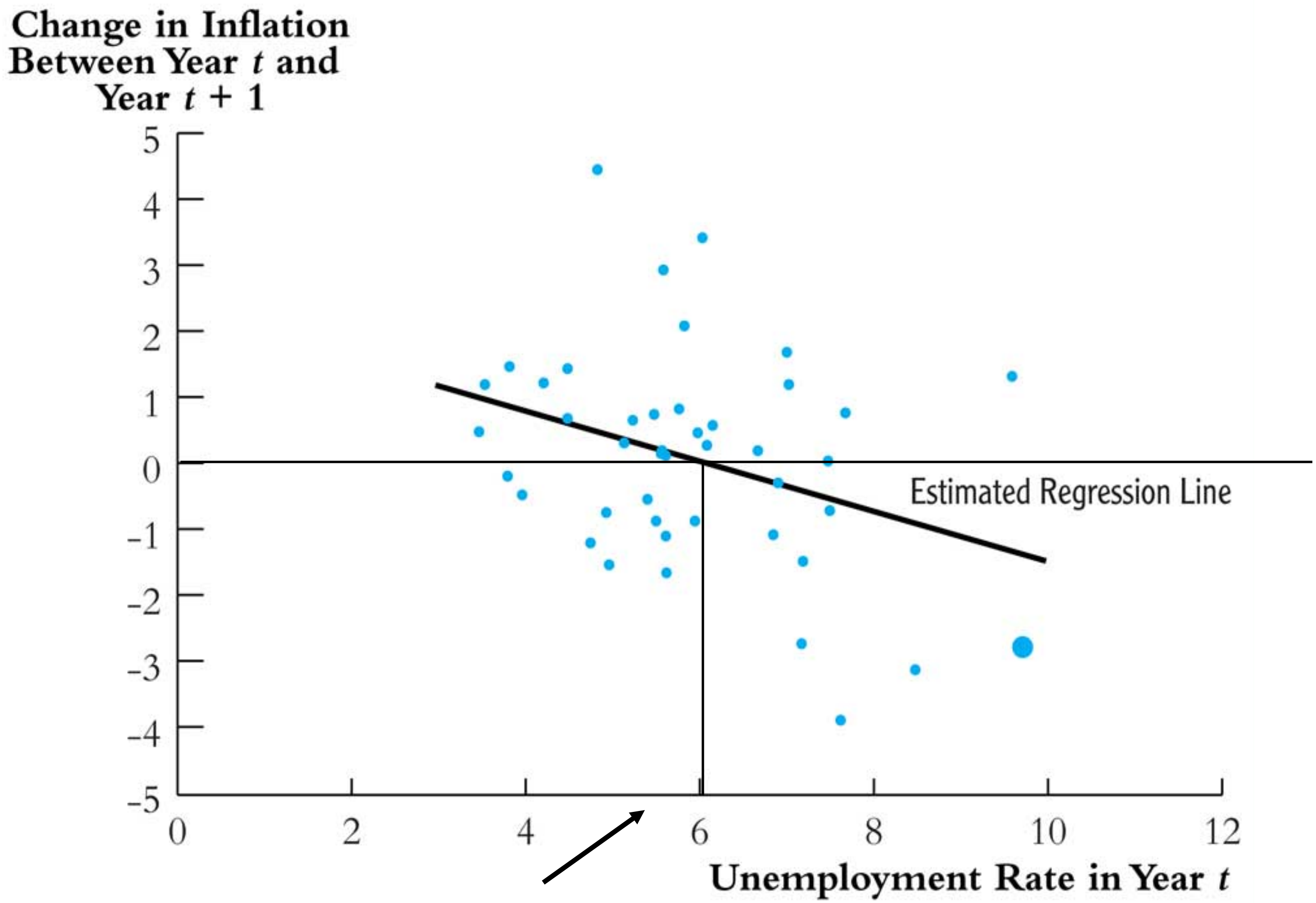
Example: inflation and unemployment

Phillips curve → unemployment above "equilibrium" rate → inflation will increase: i.e. ΔInf_t depends on U_{t-1}, …,?

- U_{t-1} at which inflation constant is often called "non-accelerating rate of inflation unemployment rate" (the NAIRU).

- Is this Phillips curve found in US economic data?
- Can it be exploited for forecasting inflation?
- Has the U.S. Phillips curve been stable over time?

The empirical U.S. “Phillips Curve,” 1962 – 2004 (annual)



One definition of the NAIRU is:

the value of *U(t-1)* for which $\Delta Inf(t) = 0$

The empirical (backwards-looking) Phillips Curve, ctd.

ADL(4,4) model of inflation (1962 – 2004):

$$\widehat{\Delta Inf}_t = \underset{(.005)}{1.24} - \underset{(.07)}{.42}\Delta Inf_{t-1} - \underset{(.08)}{.37}\Delta Inf_{t-2} + \underset{(.08)}{.02}\Delta Inf_{t-3} - \underset{(.07)}{.06}\Delta Inf_{t-4}$$

$$- \underset{(.004)}{2.33}Unem_{t-1} + \underset{(.008)}{2.74}Unem_{t-2} - \underset{(.008)}{0.86}Unem_{t-3} + \underset{(.004)}{.25}Unemp_{t-4}$$

- $\bar{R}^2 = 0.33$ – a big improvement over the AR(4), for which $\bar{R}^2 = .18$
- SER=1.39

Example: *dinf* and *unem* – STATA

```
.reg  d.acpigrowth d.l(1/4).acpigrowth l(1/4).lhur

      Source |       SS       df       MS              Number of obs =     187
-------------+------------------------------           F(8, 178)     =   12.62
       Model |  194.466872     8  24.3083589           Prob > F      =  0.0000
    Residual |   342.96794   178  1.92678618           R-squared     =  0.3618
-------------+------------------------------           Adj R-squared =  0.3332
       Total |  537.434811   186  2.88943447           Root MSE      =  1.3881

------------------------------------------------------------------------------
D.acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  acpigrowth |
         LD. |  -.4191789   .0749334    -5.59   0.000    -.5670511   -.2713066
        L2D. |  -.3711839   .0801989    -4.63   0.000     -.529447   -.2129208
        L3D. |   .0216969   .0781992     0.28   0.782    -.1326199    .1760138
        L4D. |  -.0628063   .0697722    -0.90   0.369    -.2004935     .074881
             |
        lhur |
         L1. |  -2.331154   .3948317    -5.90   0.000    -3.110307      -1.552
         L2. |   2.737917   .7625057     3.59   0.000     1.233203    4.242631
         L3. |  -.8643097   .7736796    -1.12   0.265    -2.391074    .6624548
         L4. |   .2455559   .4038164     0.61   0.544    -.5513277     1.04244
             |
       _cons |   1.241611   .4840401     2.57   0.011     .2864154    2.196806
------------------------------------------------------------------------------
```

Example: ADL(4,4) model of inflation – STATA, ctd.

```
test l1.lhur l2.lhur l3.lhur l4.lhur

     ( 1)  L.lhur = 0
     ( 2)  L2.lhur = 0
     ( 3)  L3.lhur = 0
     ( 4)  L4.lhur = 0

           F(  4,   178) =   11.40
                Prob > F =    0.0000

test l3.lhur l4.lhur

 ( 1)  L3D.lhur = 0
 ( 2)  L4D.lhur = 0

       F(  2,   178) =    0.85
            Prob > F =    0.4301
```

The null hypothesis that the coefficients on all 4 lags of the unemployment rate are all zero is rejected at the 5% significance level. However, perhaps the 3rd and 4th lags are not needed here (p-val = 0.43)

Granger Causality Test

The Granger causality statistic is the F-statistic testing the hypothesis that the coefficients on all the values of one of the regressors (e.g., $X_{1t-1}, X_{1t-2}, \dots, X_{1t-q_1}$) are zero. This null hypothesis implies that these regressors have no **predictive** content for Y_t beyond that contained in the other regressors, and the test of this null hypothesis is called the ***Granger causality test***.

"causality" is an unfortunate term here: Granger Causality simply refers to (marginal) predictive content.

Lag Length Selection Using Information Criteria

How to choose the number of lags p in an AR(p)?

- Omitted variable bias is irrelevant ***for forecasting***
- You can use sequential "downward" t- or F-tests; but the models chosen tend to be "too large" (why?)
- Another – better – way to determine lag lengths is to use an *information criterion*
- Information criteria trade off bias (too few lags) vs. variance (too many lags)
- Two *IC* are the Bayes (BIC) and Akaike (AIC)…

The Bayes Information Criterion (BIC)

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

- *First term*: always decreasing in p (larger p, better fit)
- *Second term*: always increasing in p.
  - The variance of the forecast due to estimation error increases with p – so you don't want a forecasting model with too many coefficients – but what is "too many"?
  - This term is a "penalty" for using more parameters – and thus increasing the forecast variance.
- *Minimizing BIC*(p) trades off bias and variance to determine a "best" value of p for your forecast.
  - The result is that $\hat{p}^{BIC} \xrightarrow{p} p$

Akaike Information Criterion (AIC)

Another information criterion:

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

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

The penalty term is smaller for *AIC* than *BIC* ($2 < \ln T$)

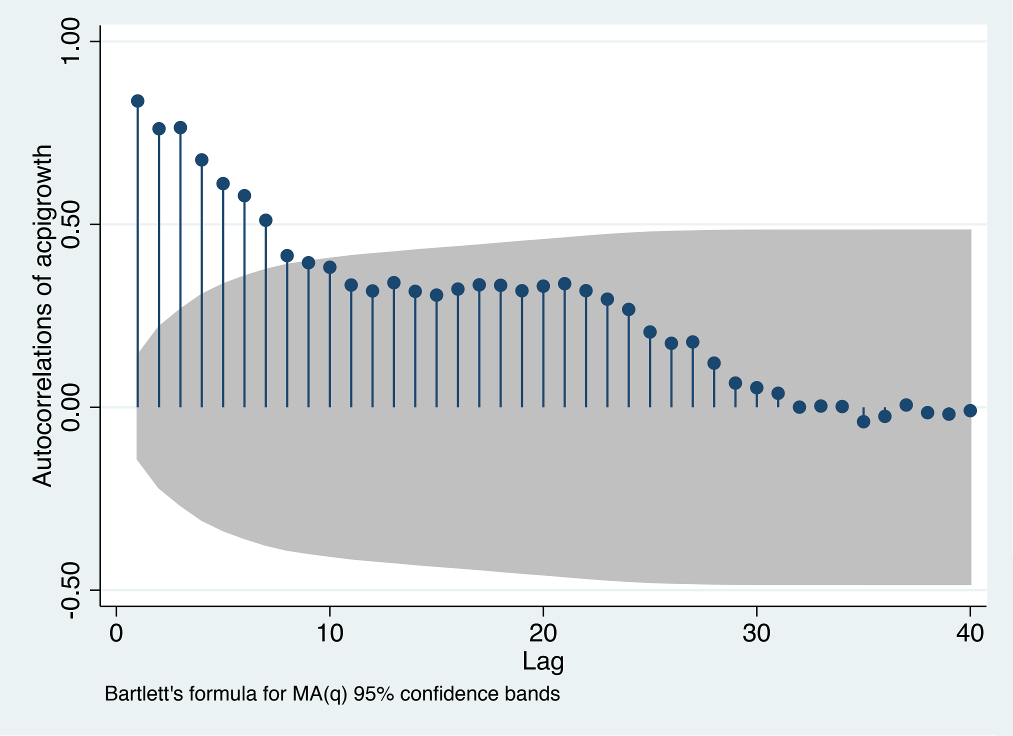
- *AIC* chooses more lags (larger p) than the *BIC*
- Desirable if you think longer lags might be important.
- However, the AIC estimator of p isn't consistent – it can overestimate p – the penalty isn't big enough

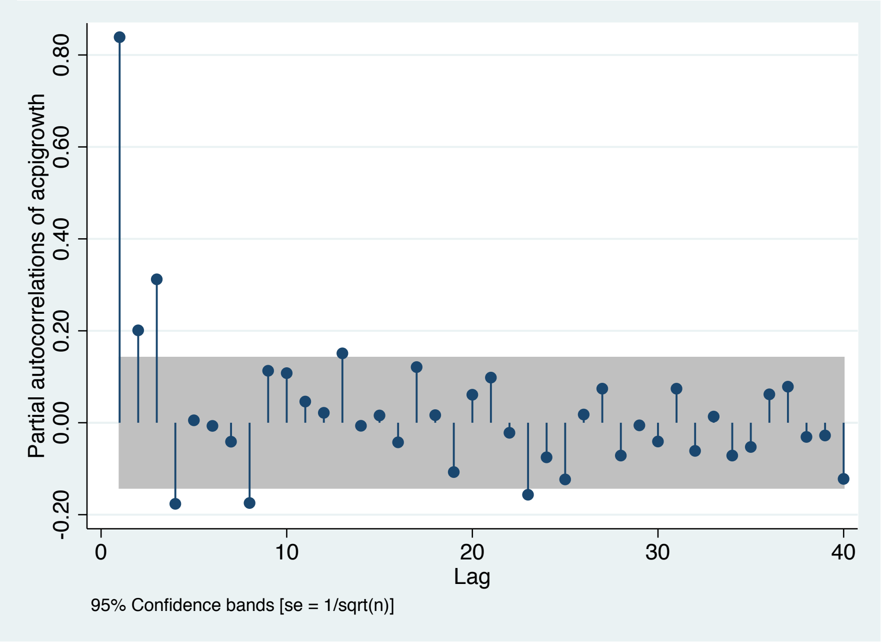
Example: AR model of inflation, lags 0–6:

# Lags	BIC	AIC	*RSS*	*Highest lag p-val*
0	-8.12	-8.13	0.055	NA
1	-8.16	-8.20	0.051	0.00
1/2	-8.267	-8.32	0.0449	0.00
1/3	**-8.272**	-8.341	0.0434	0.047
1/4	-8.254	-8.339	0.0431	0.58
1/5	-8.24	**-8.343**	0.0425	0.70
1/6	-8.22	-8.336	**0.0423**	0.95

- BIC chooses 3 lags, AIC chooses 5 lags.

Example: AR model of inflation, ACF and PACF





Generalization to ADL models

Let K = the total number of coefficients in the model (intercept, lags of Y, lags of X). The BIC is,

$$\text{BIC}(K) = \ln\left(\frac{SSR(K)}{T}\right) + K\frac{\ln T}{T}$$

- Can compute this over all possible combinations of lags of Y and lags of X (but this is a lot)!
- In practice you might choose lags of Y by BIC, and decide whether or not to include X using a Granger causality test with a fixed number of lags (number depends on the data and application)

Example: ADL model of inflation and unemployment

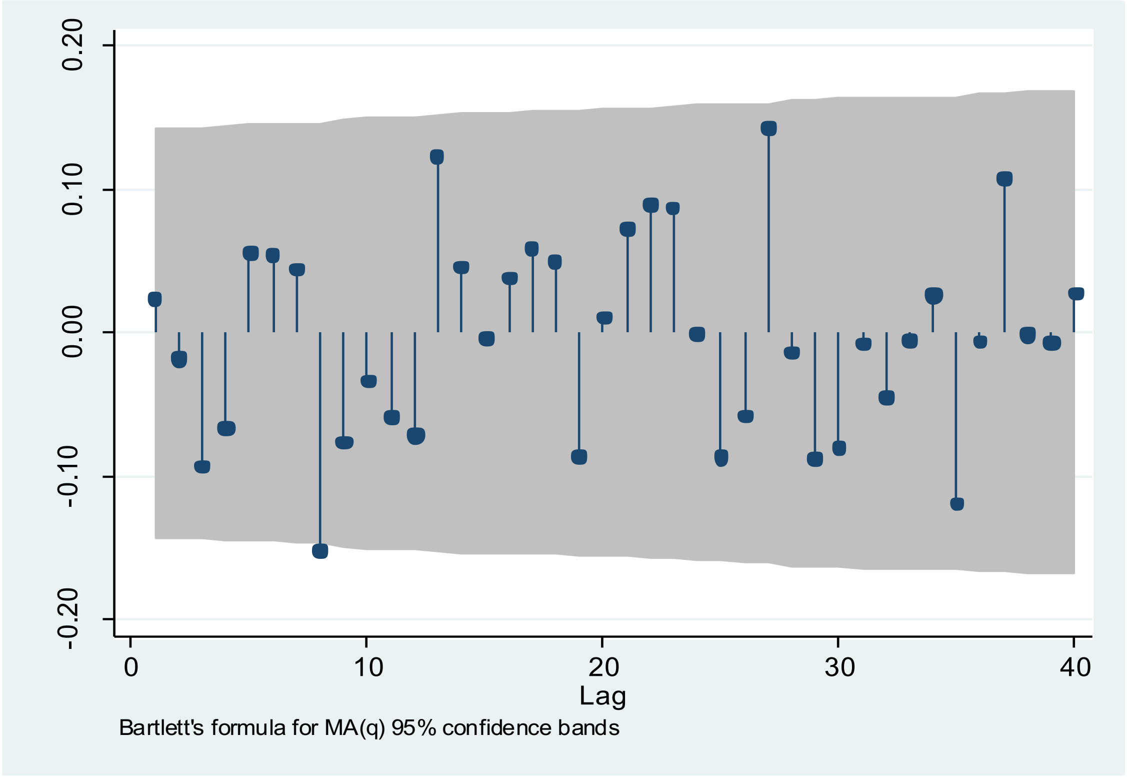
BIC	DL(r)				
AR(p)	0	1	2	3	4
0	-8.116	-8.097	-8.151	-8.185	-8.173
1	-8.161	-8.153	-8.264	-8.276	-8.256
2	-8.267	-8.292	-8.433	-8.410	-8.385
3	-8.273	-8.281	**-8.436**	-8.416	-8.391
4	-8.254	-8.268	-8.417	-8.397	-8.371

Example: *dinf* and *unem* – STATA

```
reg  d.acpigrowth d.l(1/3).acpigrowth l(1/2).lhur

      Source |       SS           df       MS      Number of obs   =       188
-------------+----------------------------------   F(5, 182)       =     20.04
       Model |  192.276636         5  38.4553272   Prob > F        =    0.0000
    Residual |  349.257612       182  1.91899787   R-squared       =    0.3551
-------------+----------------------------------   Adj R-squared   =    0.3373
       Total |  541.534248       187  2.89590507   Root MSE        =    1.3853

------------------------------------------------------------------------------
D.acpigrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  acpigrowth |
         LD. |  -.4418675   .0699779    -6.31   0.000    -.5799398   -.3037952
        L2D. |  -.3687404    .068248    -5.40   0.000    -.5033995   -.2340813
        L3D. |   .0535678   .0690586     0.78   0.439    -.0826907    .1898262
             |
        lhur |
         L1. |  -1.957739   .2986803    -6.55   0.000     -2.54706   -1.368417
         L2. |   1.770219   .2933674     6.03   0.000     1.191381    2.349058
             |
       _cons |    1.10352   .4615548     2.39   0.018     .1928336    2.014206
------------------------------------------------------------------------------
```

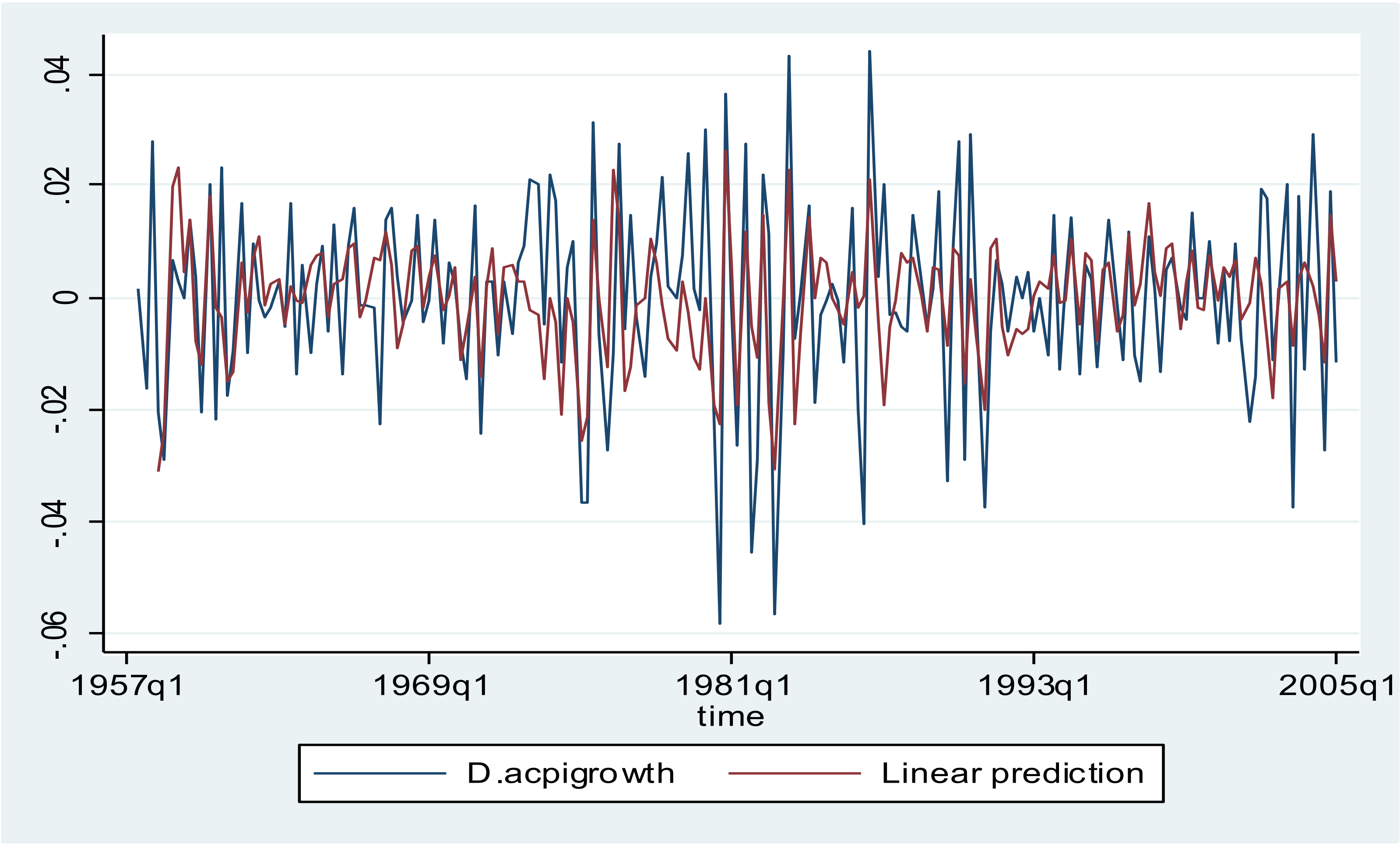
Example: *dinf* and *unem* – STATA

```
estat bgodfrey, lags(4)

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4     |          5.388               4                   0.2497
---------------------------------------------------------------------------
                        H0: no serial correlation

test l1.lhur l2.lhur

 ( 1)  L.lhur = 0
 ( 2)  L2.lhur = 0

       F(  2,   182) =   22.21
            Prob > F =    0.0000

predict res23, resid
(5 missing values generated)

ac res23
```

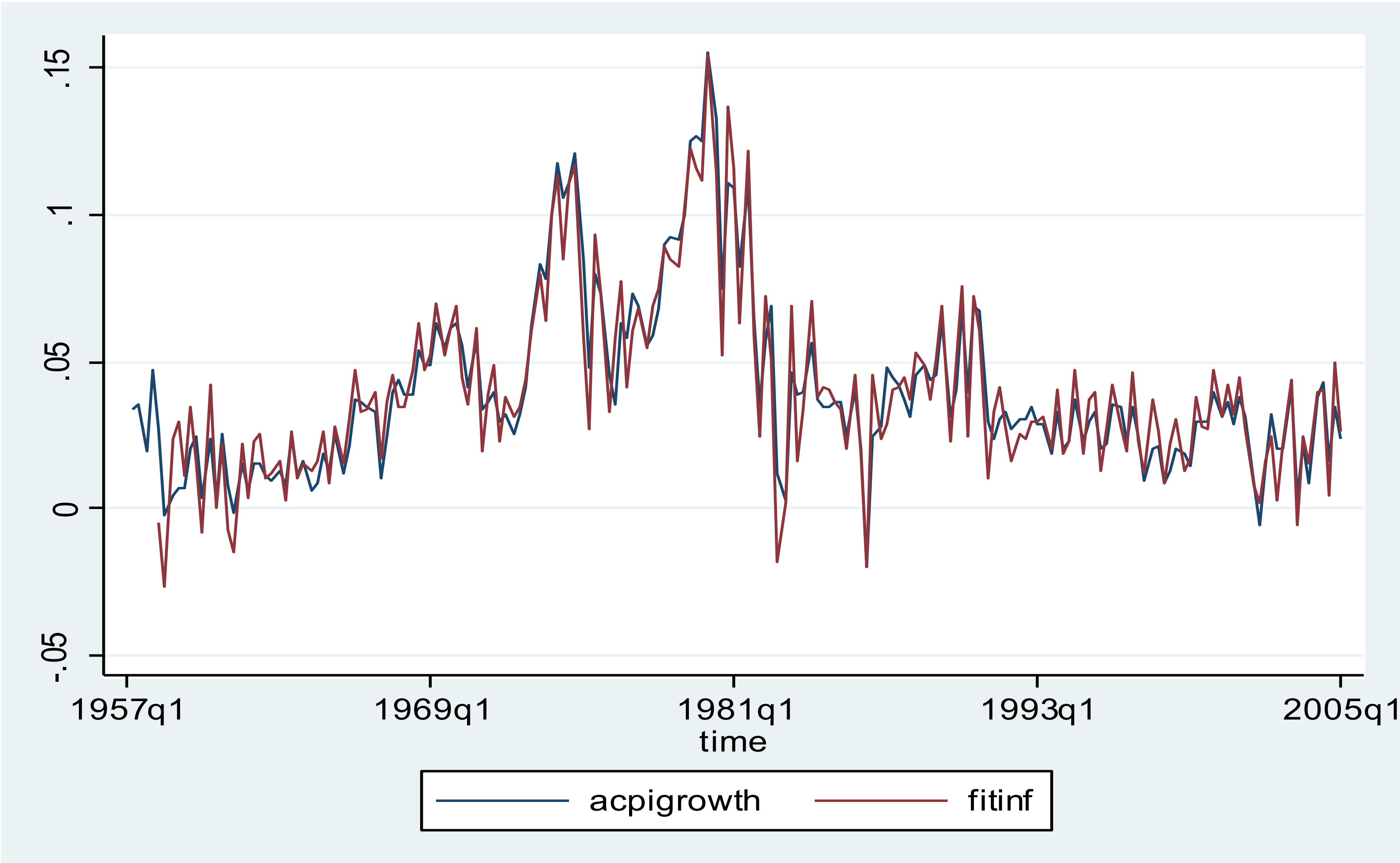


Example: *dinf* and *unem*

```
tsline d.acpigrowth  fit
gen fitinf= acpigrowth+fit
tsline  acpigrowth fitinf
```



Example*: *inf* and *unem



Vector Autoregression (VAR)

- Sims (1980) proposed to use VAR to model and forecast the multivariate time series.
- Given two times seires $\{y_{1t}, y_{2t}\}$
- A VAR(p) can be written as

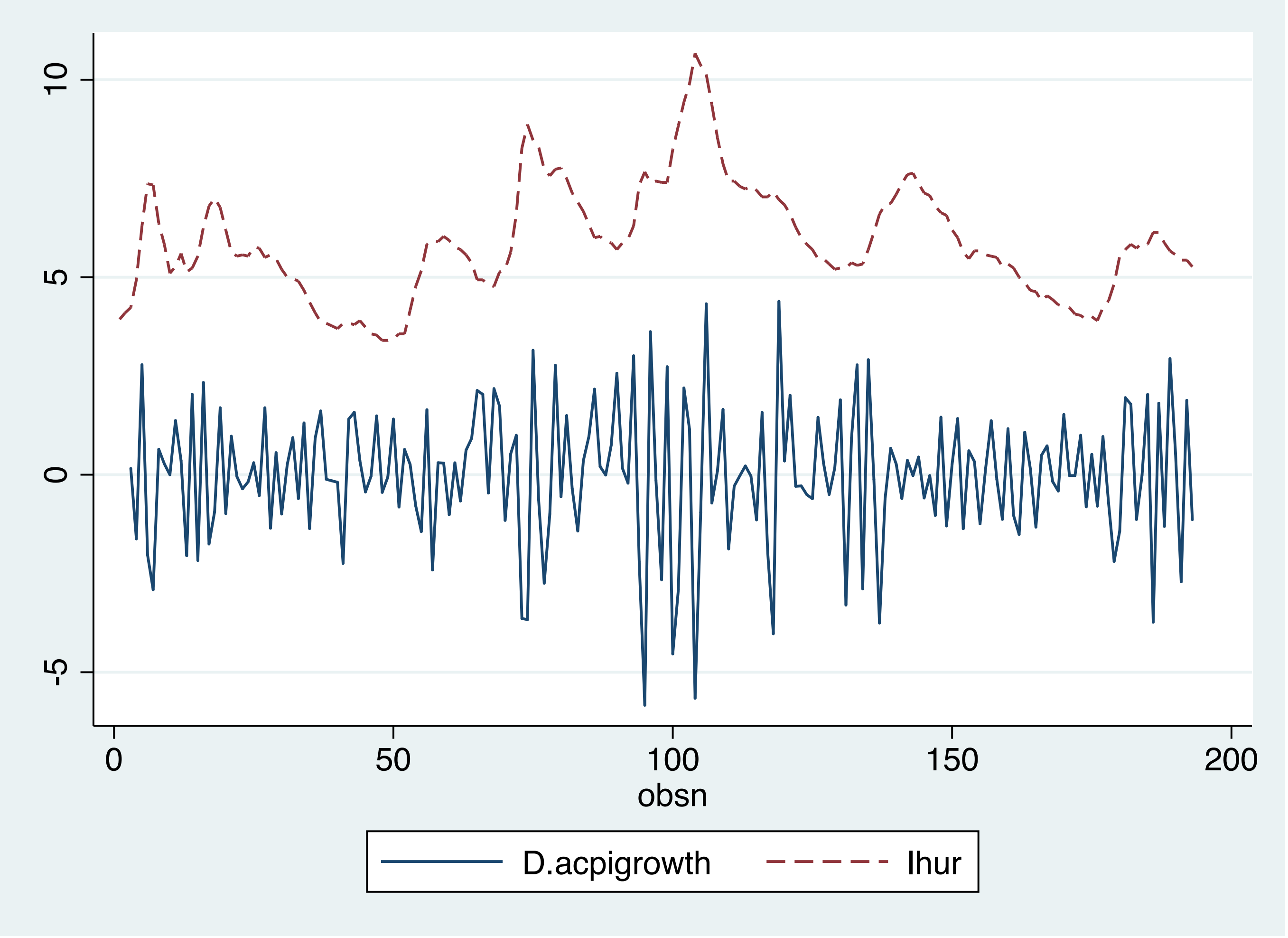
$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

where $\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$ follows a vector white noise process, or innovation process.

To choose the order of lags, there three approaches:

1. Use Information Criteria
2. Test for the significance of highest order of lag
3. Test for the autocorrelation in $\boldsymbol{\varepsilon}_t$

Example: *dinf* and *unem—bivariate VAR*



Example: *dinf* and *unem—bivariate VAR*

Determine the order of VAR using information criteria

```
varsoc d.acpigrowth lhur

   Selection-order criteria
   Sample:  7 - 193                                       Number of obs      =      187
  +---------------------------------------------------------------------------+
  |lag |    LL        LR      df    p      FPE       AIC       HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 | -694.599                        5.89733   7.45025   7.46426   7.48481  |
  |  1 | -407.517  574.16    4  0.000  .285616   4.42264   4.46465   4.52631  |
  |  2 | -336.337  142.36*   4  0.000  .139235*  3.70414*  3.77415*  3.87692* |
  |  3 | -334.441  3.7925    4  0.435   .14241   3.72664   3.82465   3.96854  |
  |  4 | -332.083   4.716    4  0.318  .144944    3.7442   3.87022   4.05521  |
  +---------------------------------------------------------------------------+
   Endogenous:  D.acpigrowth lhur
    Exogenous:  _cons
```

VAR(2) is selected.

```
var d.acpigrowth lhur,lags(1/2)

Vector autoregression

Sample:  5 - 193                              Number of obs     =       189
Log likelihood =  -347.6376                    AIC               =   3.784525
FPE            =  .1508899                    HQIC              =   3.854012
Det(Sigma_ml)  =   .1357352                    SBIC              =   3.956046

Equation           Parms     RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_acpigrowth          5     1.39879   0.3446   99.37607   0.0000
lhur                 5     .270597   0.9651   5227.896   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_acpigrowth |
  acpigrowth |
        LD. |  -.4718838   .0633887    -7.44   0.000    -.5961233   -.3476444
       L2D. |  -.3930333   .0622371    -6.32   0.000    -.5150157   -.2710508
            |
       lhur |
        L1. |  -1.918407   .2918633    -6.57   0.000    -2.490449   -1.346366
        L2. |   1.702002   .2884616     5.90   0.000     1.136628    2.267377
            |
      _cons |   1.291062   .4403916     2.93   0.003     .4279107    2.154214
-------------+----------------------------------------------------------------
lhur        |
  acpigrowth |
        LD. |   .0273254   .0122626     2.23   0.026     .0032912    .0513596
       L2D. |  -.0081393   .0120398    -0.68   0.499    -.0317369    .0154582
            |
       lhur |
        L1. |   1.609254   .0564611    28.50   0.000     1.498592    1.719915
        L2. |   -.655515    .055803   -11.75   0.000    -.7648869   -.5461431
            |
      _cons |   .2715359   .0851939     3.19   0.001     .1045589     .438513
------------------------------------------------------------------------------
```

Tests for the bivariate VAR

Most parameters are significant. Next we test the joint significance of the parameters for lagged terms.

```
varwle

   Equation: D_acpigrowth
  +-----------------------------------+
  | lag |    chi2      df  Prob > chi2 |
  |-----+-----------------------------|
  |   1 |   79.6889     2     0.000    |
  |   2 |  73.60786     2     0.000    |
  +-----------------------------------+

   Equation: lhur
  +-----------------------------------+
  | lag |    chi2      df  Prob > chi2 |
  |-----+-----------------------------|
  |   1 |  835.1099     2     0.000    |
  |   2 |   138.713     2     0.000    |
  +-----------------------------------+

   Equation: All
  +-----------------------------------+
  | lag |    chi2      df  Prob > chi2 |
  |-----+-----------------------------|
  |   1 |   908.841     4     0.000    |
  |   2 |  209.8526     4     0.000    |
  +-----------------------------------+
```

Obtain Wald lag-exclusion statistics after var or svar

The parameters of lagged terms are highly jointly significant for each individual equation as well as both equations as a system.

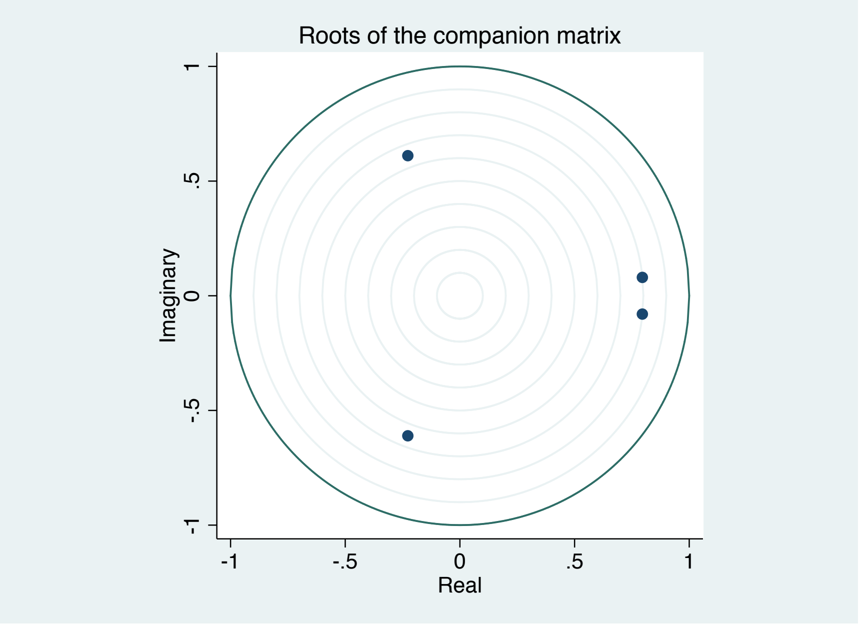
Tests for the bivariate VAR

Test whether the residuals follow the white noise process (whether autocorrelated)

```
varlmar

   Lagrange-multiplier test
  +--------------------------------------+
  | lag  |      chi2    df   Prob > chi2 |
  |------+-------------------------------|
  |   1  |    3.6683     4     0.45276   |
  |   2  |    0.6551     4     0.95674   |
  +--------------------------------------+
   H0: no autocorrelation at lag order
```

Test whether this VAR process is stable

```
varstable,graph

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue        |  Modulus   |
  |--------------------------+------------|
  |   .7956761 + .07992645i  |    .79968  |
  |   .7956761 - .07992645i  |    .79968  |
  |  -.2269912 +  .6107545i  |   .651572  |
  |  -.2269912 -  .6107545i  |   .651572  |
  +----------------------------------------+
   All the eigenvalues lie inside the unit circle.
   VAR satisfies stability condition.
```



Granger Causality between dinf and *unem*

```
vargranger

   Granger causality Wald tests
  +------------------------------------------------------------------+
  |          Equation           Excluded |   chi2     df Prob > chi2 |
  |--------------------------------------+---------------------------|
  |     D_acpigrowth                lhur |  46.786     2    0.000    |
  |     D_acpigrowth                 ALL |  46.786     2    0.000    |
  |--------------------------------------+---------------------------|
  |             lhur      D.acpigrowth |  6.8417     2    0.033    |
  |             lhur                 ALL |  6.8417     2    0.033    |
  +------------------------------------------------------------------+
```

Obviously, there is a mutual Granger causality relation between the change in inflation and the unemployment rate.

Forecast with VAR

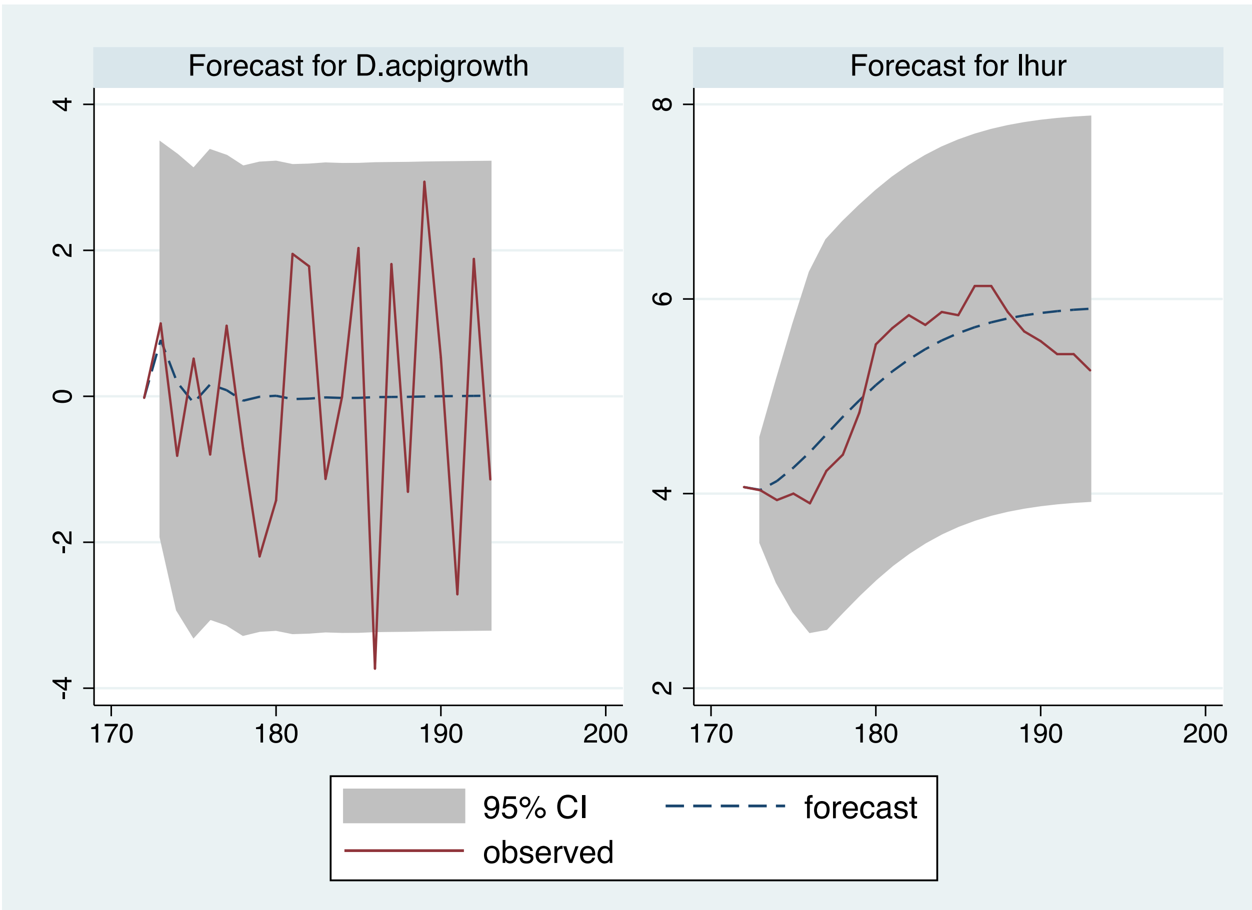
Estimate VAR using the data before 2000, the forecast for 2000Q1 to 2005Q1.

```
varbasic d.acpigrowth lhur if  time<tq(2000q1),lags(1/2) nograph

Vector autoregression

Sample:  5 - 172                             Number of obs     =       168
Log likelihood =   -309.471                   AIC              =   3.803226
FPE            =   .1537399                  HQIC             =   3.878694
Det(Sigma_ml)  =   .1364802                   SBIC             =   3.989176

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_acpigrowth          5     1.37836   0.3628   95.66408   0.0000
lhur                  5     .276246   0.9656   4715.912   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_acpigrowth |
  acpigrowth |
         LD. |  -.4534261   .0659461    -6.88   0.000    -.5826781   -.3241742
        L2D. |  -.4045939   .0646435    -6.26   0.000    -.5312928    -.277895
             |
        lhur |
         L1. |  -1.907187   .2935326    -6.50   0.000      -2.4825   -1.331874
         L2. |   1.665076   .2911991     5.72   0.000     1.094336    2.235815
             |
       _cons |   1.460851   .4605582     3.17   0.002     .5581739    2.363529
-------------+----------------------------------------------------------------
lhur         |
  acpigrowth |
         LD. |   .0336412   .0132167     2.55   0.011     .0077369    .0595455
        L2D. |  -.0132056   .0129556    -1.02   0.308    -.0385982     .012187
             |
        lhur |
         L1. |   1.625494   .0588289    27.63   0.000     1.510191    1.740797
         L2. |  -.6706413   .0583612   -11.49   0.000    -.7850272   -.5562554
             |
       _cons |   .2666984   .0923036     2.89   0.004     .0857866    .4476102
------------------------------------------------------------------------------
```

Forecast with VAR

Generate forecasts for the acpigrowth and lhur and denoted as f_ acpigrowth and f_lhur, compare with true observations.

```
fcast compute f_,step(21)

. fcast graph f_D_acpigrowth f_lhur,observed  lpattern(dash)
```



$RMSFE_{dinf}$
$= 1.733$

$RMSFE_{unem}$
$= 0.358$