# Chapter 14:
# Time Series: Regression & Forecasting

# Outline

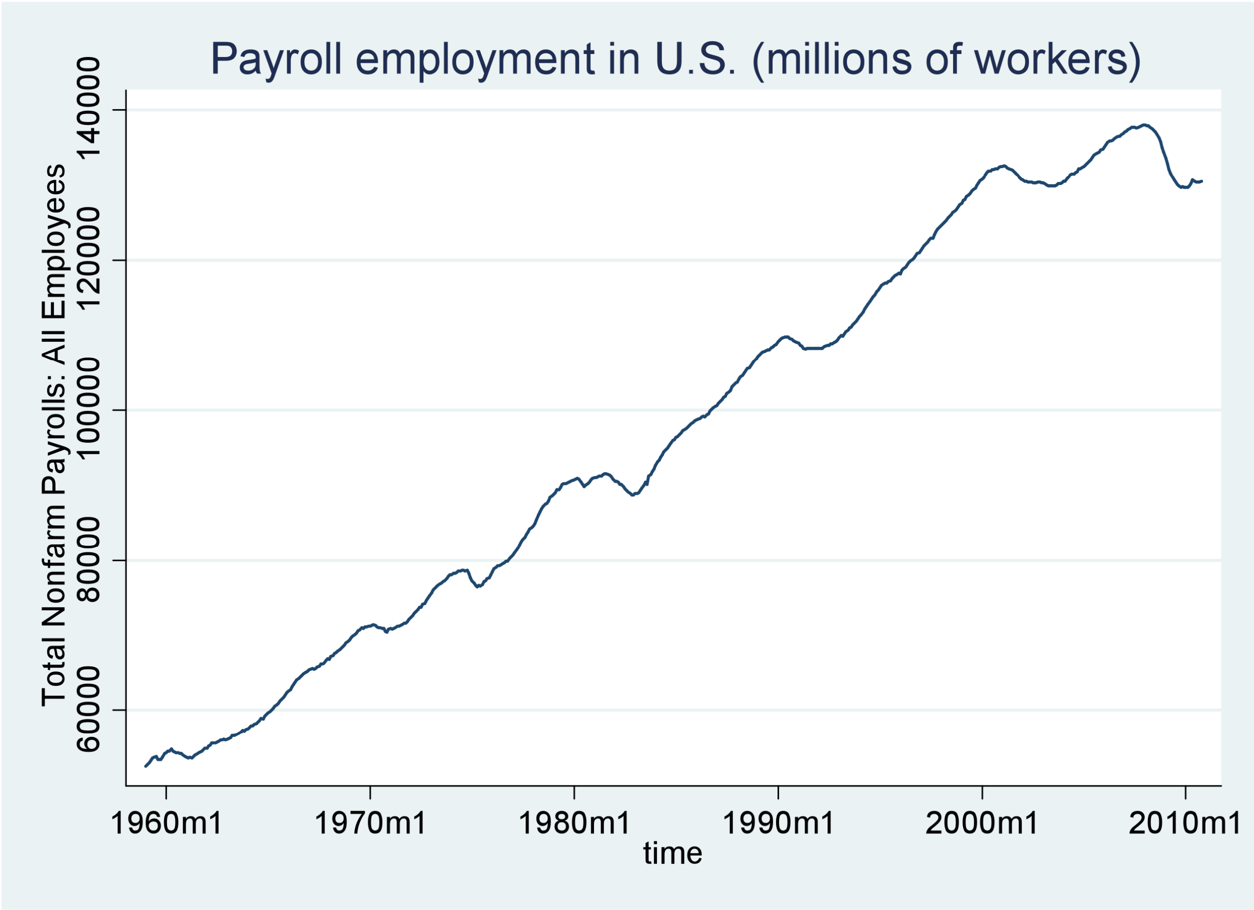
1. Time Series Data: What's Different?
2. Using Regression Models for Forecasting
3. Lags, Differences, Autocorrelation, & Stationarity
4. Autoregressions
5. The Autoregressive – Distributed Lag (ADL) Model
6. Forecast Uncertainty and Forecast Intervals
7. Lag Length Selection: Information Criteria
8. Nonstationarity I: Trends
9. Nonstationarity II: Breaks
10. Summary

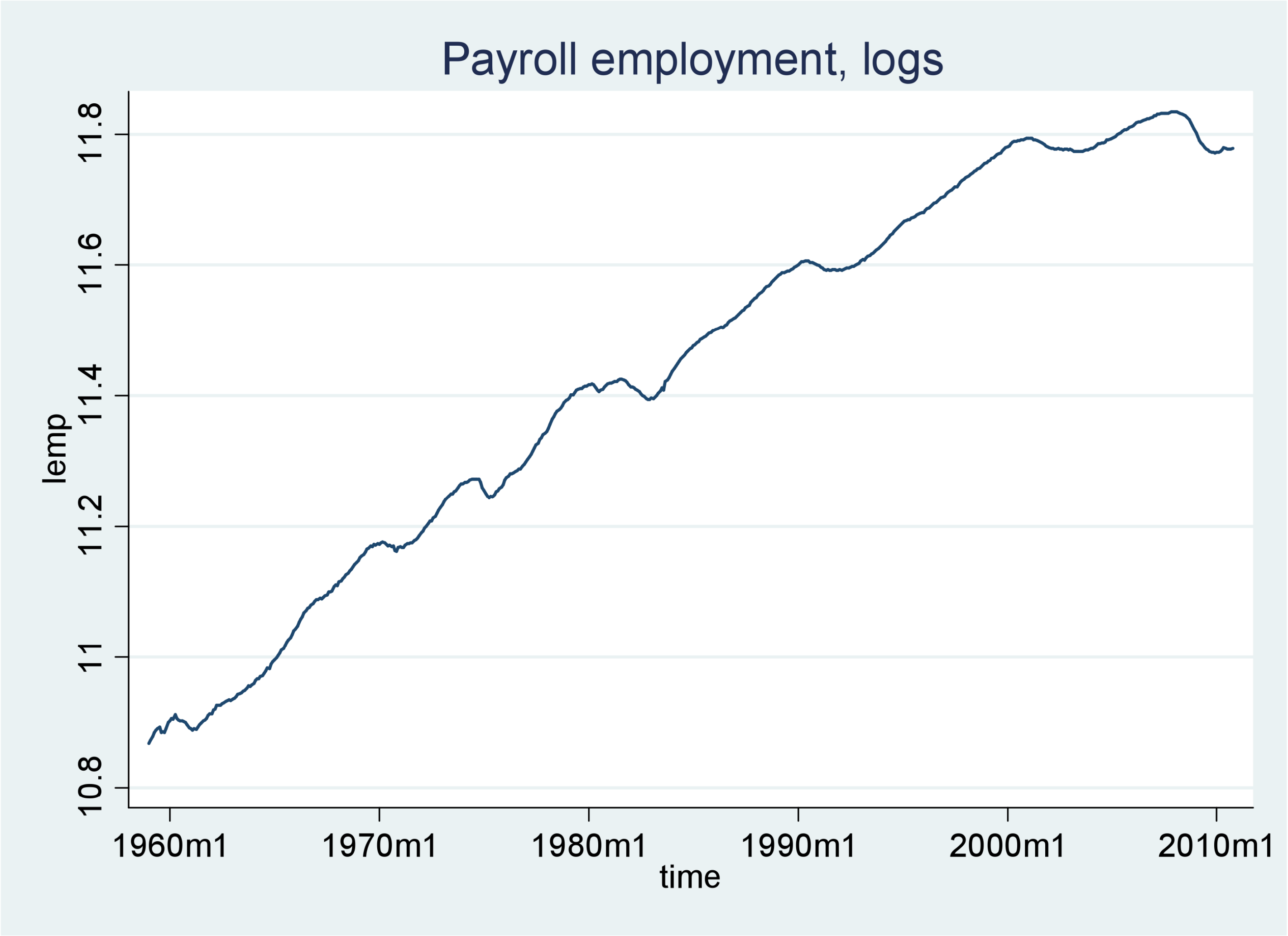
# 1. Time Series Data: What's Different?

**Time series data** are data collected on the same observational unit at multiple time periods

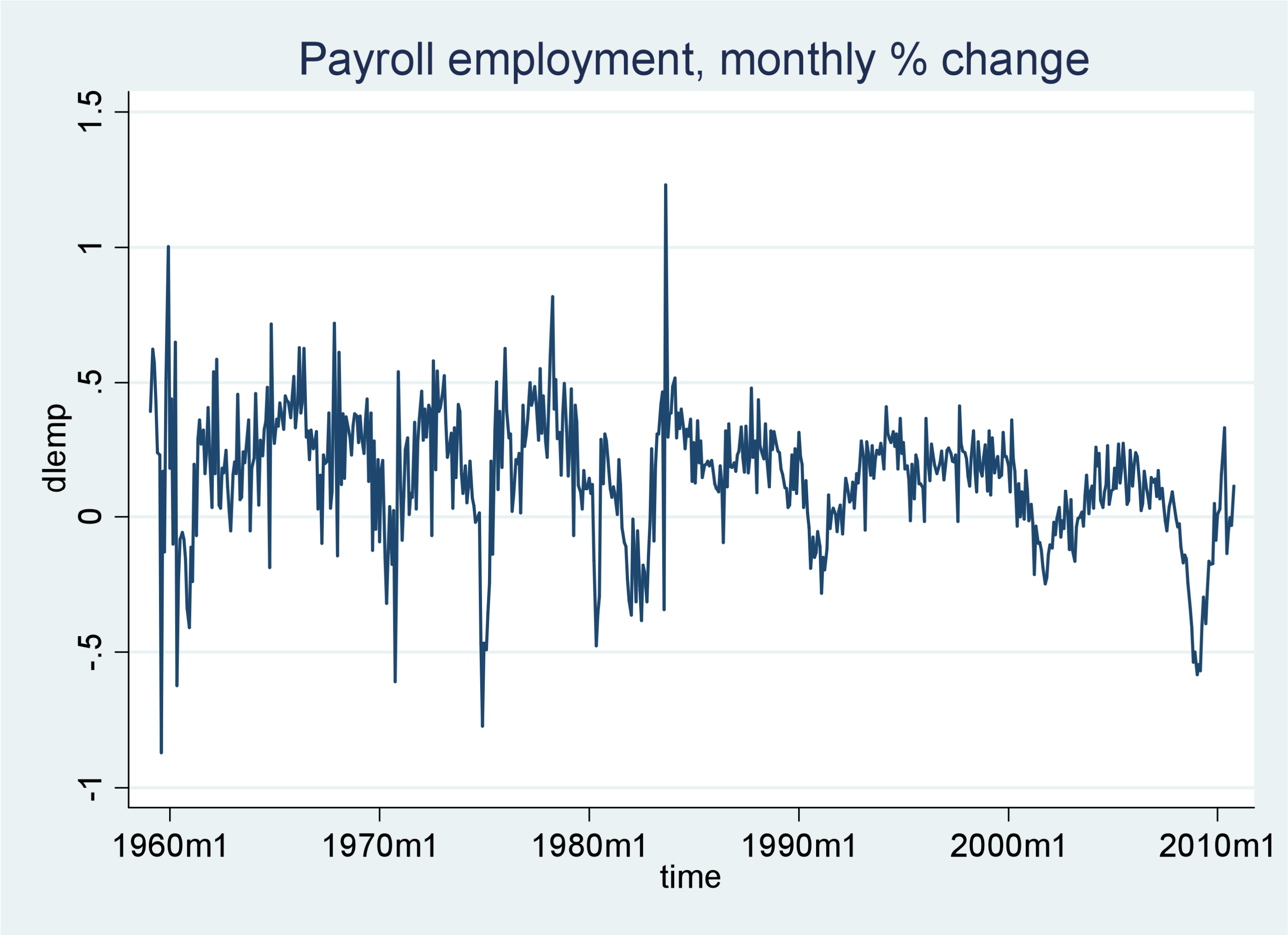
- A country's GDP (for example, 20 years of quarterly observations = 80 observations)
- Yen/$ exchange rate (minute data for 1 week)
- Cigarette consumption per capita in CA, by year

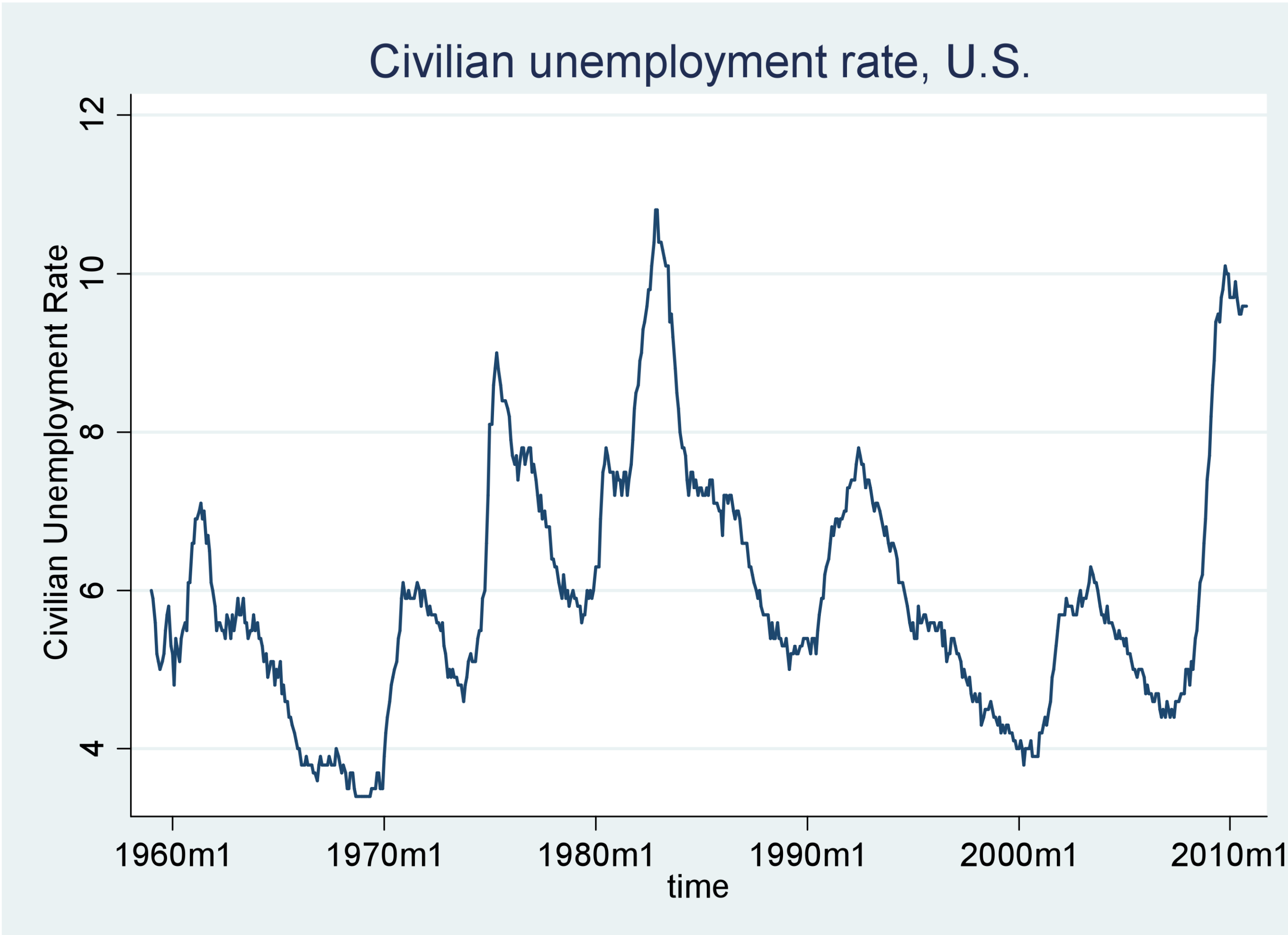
# Monthly US macro & financial time series



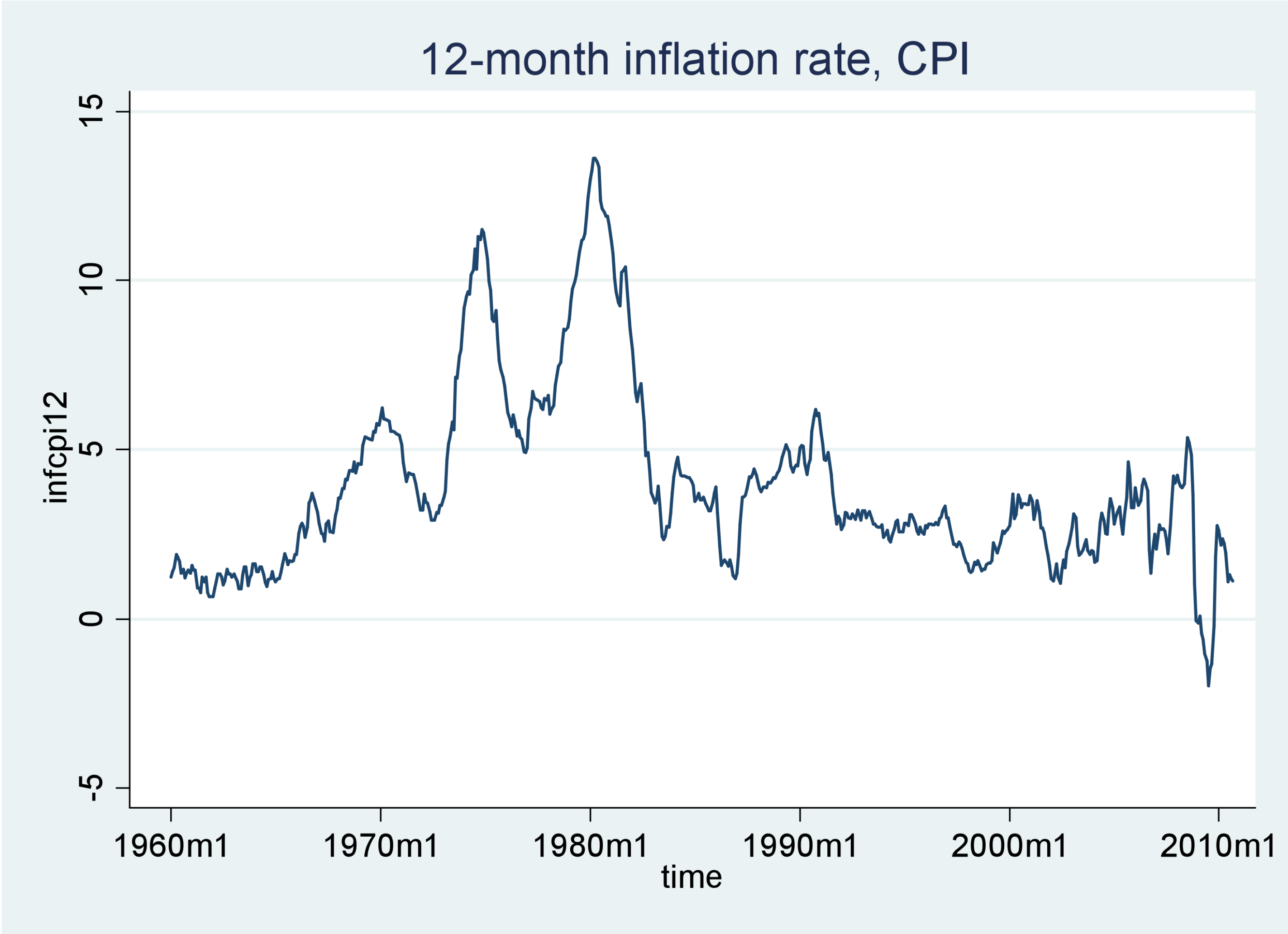


Payroll employment, logs
lemp
11.8
11.6
11.4
11.2
11
10.8
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time

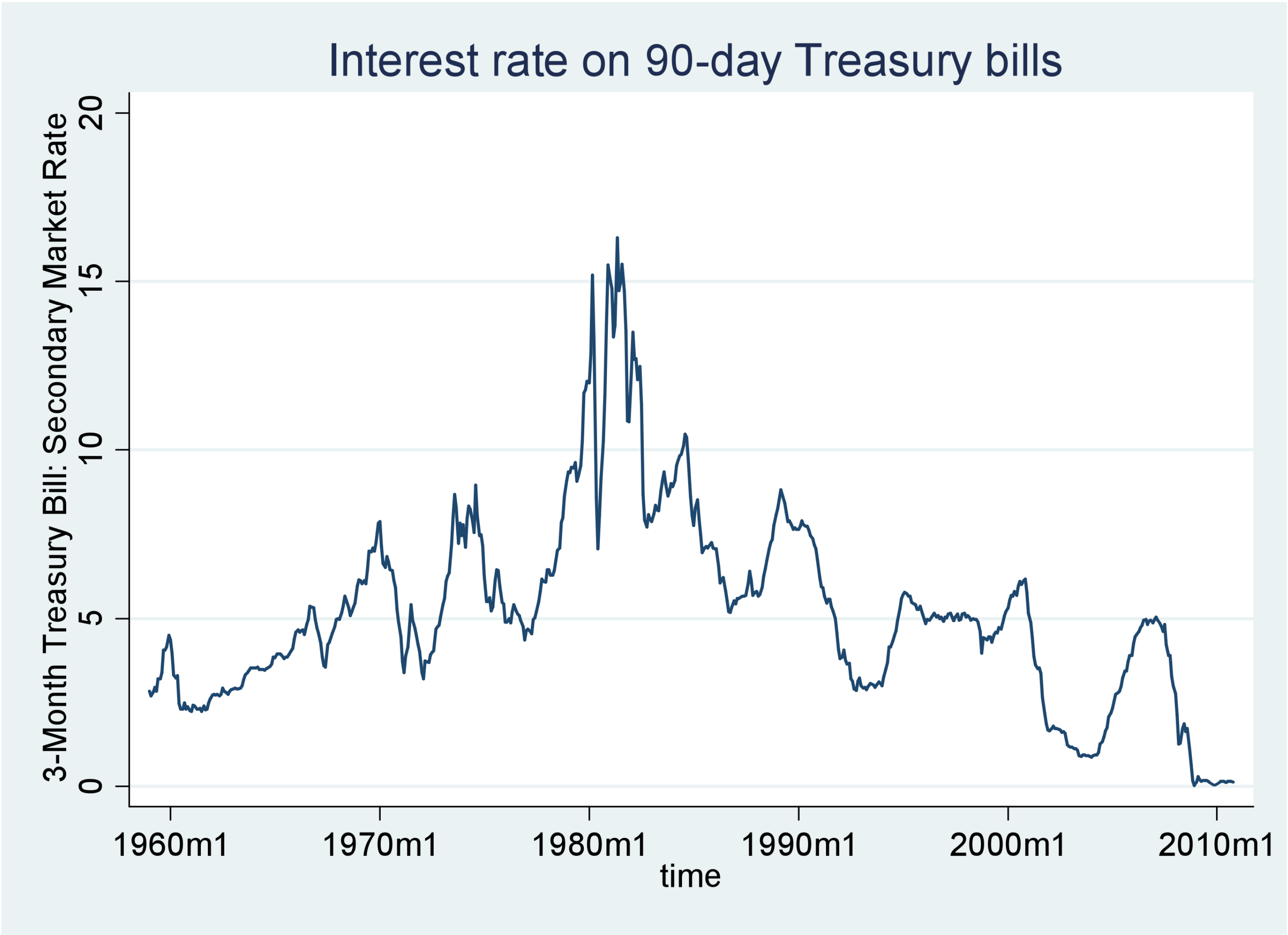

Payroll employment, monthly % change
dlemp
1.5
1
.5
0
-.5
-1
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time



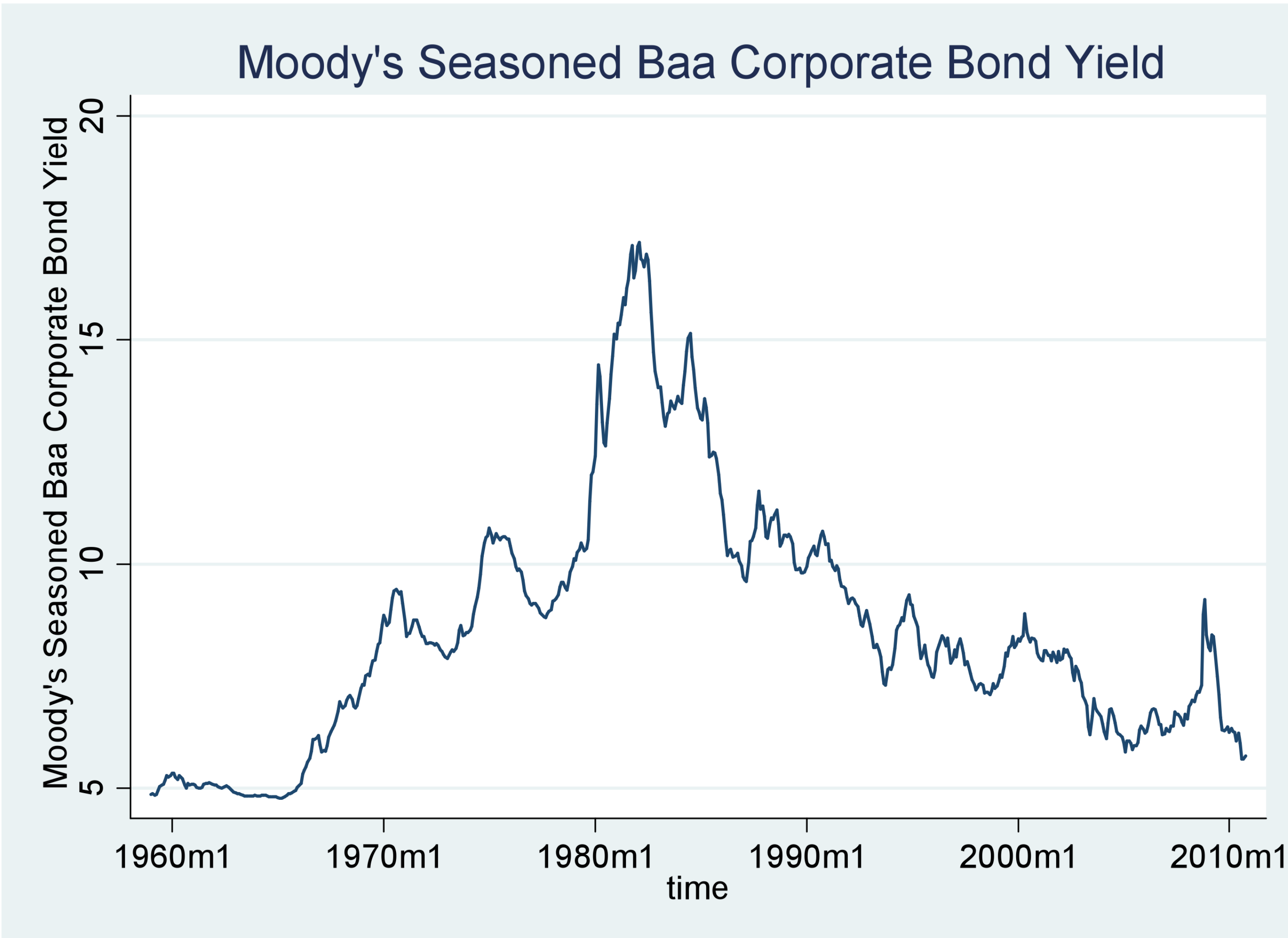
Civilian unemployment rate, U.S.
Civilian Unemployment Rate
4
6
8
10
12
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time

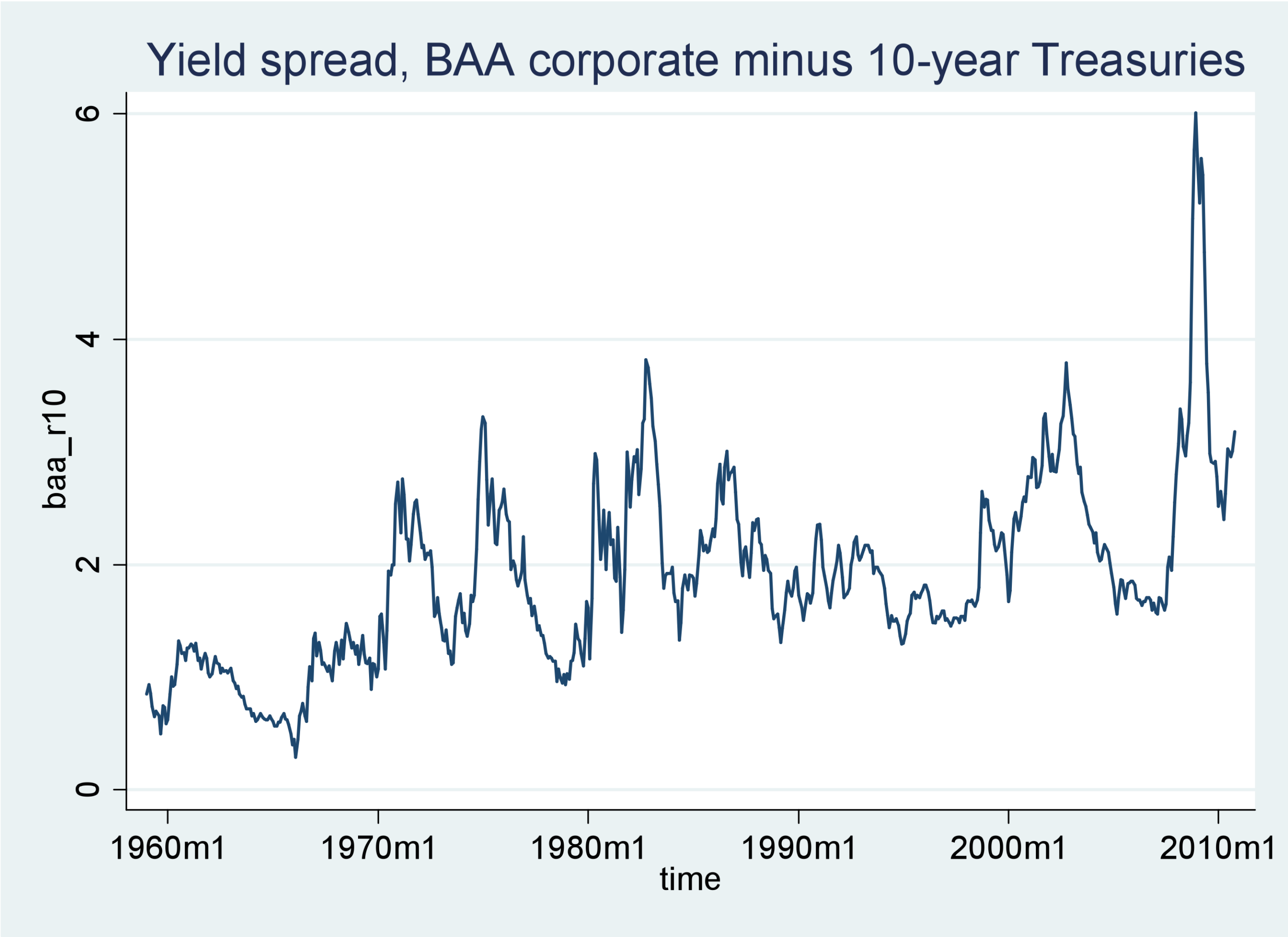


12-month inflation rate, CPI
infcpi12
15
10
5
0
-5
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time

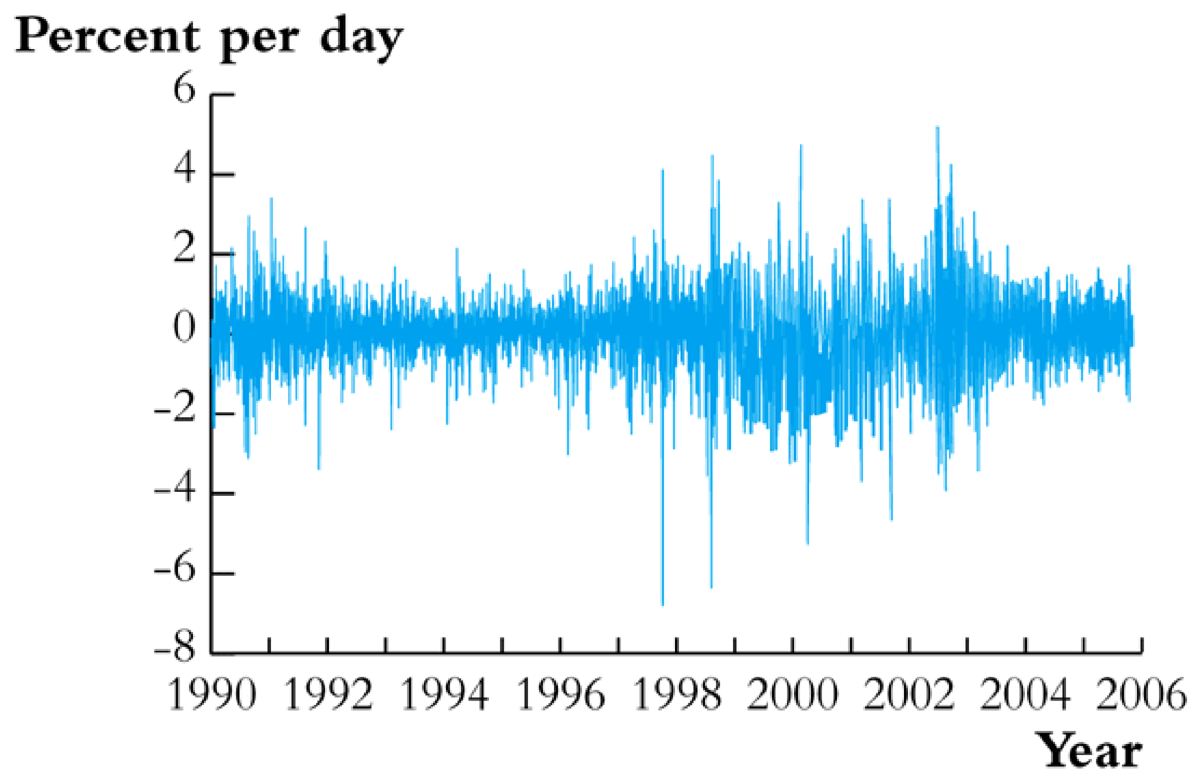

Interest rate on 90-day Treasury bills
3-Month Treasury Bill: Secondary Market Rate
0
5
10
15
20
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time


Yield on 10-year Treasury bonds
10-Year Treasury Constant Maturity Rate
0
5
10
15
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time


Moody's Seasoned Baa Corporate Bond Yield
Moody's Seasoned Baa Corporate Bond Yield
5
10
15
20
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time


Yield spread, BAA corporate minus 10-year Treasuries
baa_r10
0
2
4
6
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time

# A daily U.S. financial time series:




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

# Some uses of time series data

- Forecasting (SW Ch. 14)
- Estimation of dynamic causal effects (SW Ch. 15)
  - If the Fed increases the Federal Funds rate now, what is the effect on inflation and unemployment in 3 months?
  - What is the effect over time on cigarette consumption of a hike in the cigarette tax?
- Modeling risks in financial markets (modeling changing variances and "volatility clustering" … SW Ch. 16)
- Applications outside of economics include environmental and climate modeling, engineering (system dynamics), computer science (network dynamics),…

# Time series data: new technical issues

- Time lags
- Correlation over time (*serial correlation*, aka *autocorrelation* – as in panel data)
- Calculation of standard errors when the errors are serially correlated

A great source for U.S. macro time series data, and some international data, is the Federal Reserve Bank of St. Louis's FRED database. A good way to learn about time series data is to investigate it yourself.

# Using Regression Models for Forecasting

- Forecasting and estimation of causal effects are quite different objectives.
- For forecasting,
  - $\bar{R}^2$ matters (a lot!)
  - Omitted variable bias isn't a problem!
  - External validity is paramount: the model estimated using historical data must hold into the (near) future
  - No interest in interpreting coefficients or causality– forecasting does not require causality

    Example: If everyone exiting the building is opening umbrellas, I can forecast that it is raining … without worrying whether opening umbrellas cause rain

# Introduction to Time Series Data and Serial Correlation

Time series basics:

A. Notation
B. Lags, first differences, and growth rates
C. Autocorrelation (serial correlation)
D. Stationarity

# A. Notation

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

# B. Lags, first differences, and growth rates

## Lags, First Differences, Logarithms, and Growth Rates

- The first lag of a time series $Y_t$ is $Y_{t-1}$; its $j^{\text{th}}$ lag is $Y_{t-j}$.
- The first difference of a series, $\Delta Y_t$, 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.

***Example*: Quarterly rate of inflation at an annual rate (U.S.)**
***CPI = Consumer Price Index (Bureau of Labor Statistics)***

- CPI in the first quarter of 2004 (2004:I) = 186.57
- CPI in the second quarter of 2004 (2004:II) = 188.60
- Percentage change in CPI, 2004:I to 2004:II

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

- Percentage change in CPI, 2004:I to 2004:II, *at an annual rate* = 4×1.088 = **4.359**% ≈ **4.4**% (percent per year)
- Interest and inflation rates are reported as annual rates
- Using the logarithmic approximation to percent changes yields 4×100× [log(188.60) – log(186.57)] = **4.329**%

# *Example*: US CPI inflation – its first lag and its change

**TABLE 14.1** Inflation in the United States in 2004 and the First Quarter of 2005

| Quarter | U.S. CPI | Rate of Inflation at an Annual Rate ($Inf_t$) | First Lag ($Inf_{t-1}$) | Change in Inflation ($\Delta Inf_t$) |
|---|---|---|---|---|
| 2004:I | 186.57 | 3.8 | 0.9 | 2.9 |
| 2004:II | 188.60 | 4.4 | 3.8 | 0.6 |
| 2004:III | 189.37 | 1.6 | 4.4 | −2.8 |
| 2004:IV | 191.03 | 3.5 | 1.6 | 1.9 |
| 2005:I | 192.17 | 2.4 | 3.5 | −1.1 |

The annualized rate of inflation is the percentage change in the CPI from the previous quarter to the current quarter, multiplied by four. The first lag of inflation is its value in the previous quarter, and the change in inflation is the current inflation rate minus its first lag. All entries are rounded to the nearest decimal.

# C. Autocorrelation (serial correlation)

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

- The first **autocovariance** of $Y_t$ is $\text{cov}(Y_t, Y_{t-1})$
- The first **autocorrelation** of $Y_t$ is $\text{corr}(Y_t, Y_{t-1})$
- 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 – relating to $(Y_t, Y_{t-1})$'s population joint distribution

## Autocorrelation (Serial Correlation) and Autocovariance

The $j^{th}$ autocovariance of a series $Y_t$ is the covariance between $Y_t$ and its $j^{th}$ lag, $Y_{t-j}$, and the $j^{th}$ autocorrelation coefficient is the correlation between $Y_t$ and $Y_{t-j}$. That is,

$$j^{th} \text{ autocovariance} = \text{cov}(Y_t, Y_{t-j}) \tag{14.3}$$

$$j^{th} \text{ autocorrelation} = \rho_j = \text{corr}(Y_t, Y_{t-j}) = \frac{\text{cov}(Y_t, Y_{t-j})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-j})}}. \tag{14.4}$$

The $j^{th}$ autocorrelation coefficient is sometimes called the $j^{th}$ serial correlation coefficient.

# Sample autocorrelations

The $j^{th}$ **sample autocorrelation** is an estimate of the $j^{th}$ population autocorrelation:

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

where

$$\widehat{\mathrm{cov}(Y_t, Y_{t-j})} := \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \overline{Y}_{j+1,T})(Y_{t-j} - \overline{Y}_{1,T-j})$$

Where $\overline{Y}_{j+1,T}$ is the sample average of $Y_t$ for t = j+1,…,T.

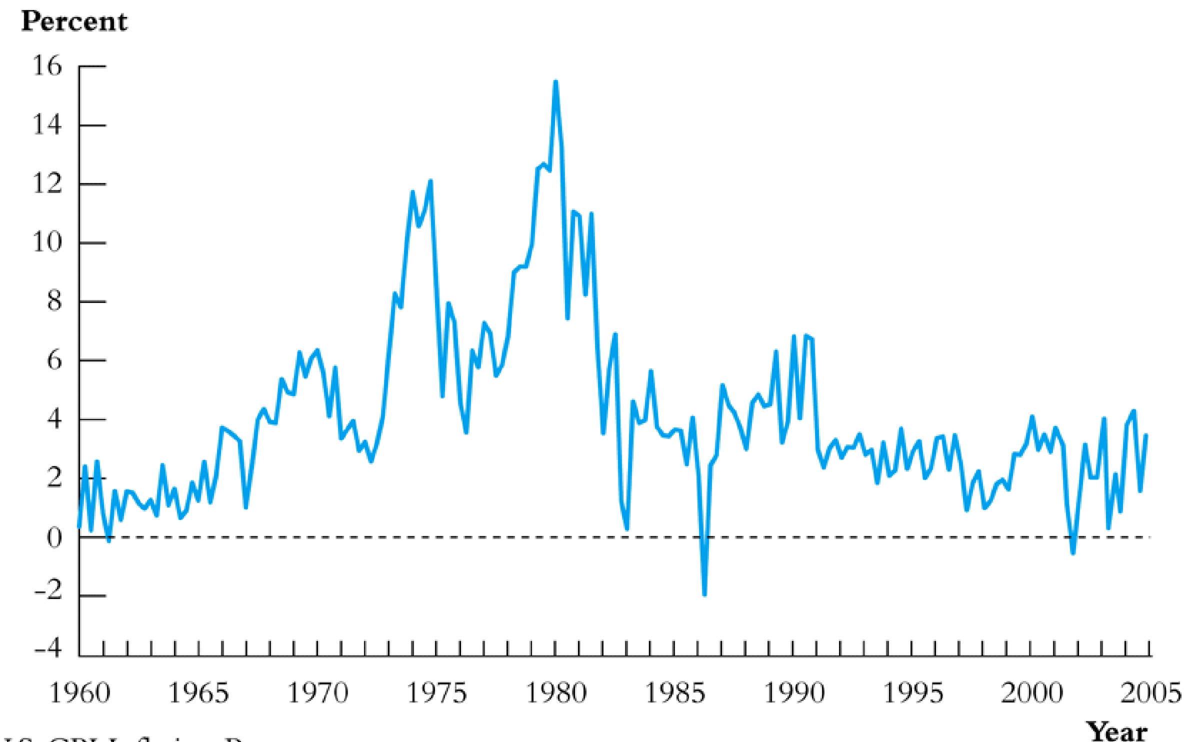
NOTE:

- The summation is over t=j+1 to T
- The divisor is T, not T – j (a time series convention)

## *Example*: Autocorrelations of:
## (1) the quarterly rate of U.S. inflation
## (2) the quarter-to-quarter change in the quarterly rate of inflation

**TABLE 14.2** First Four Sample Autocorrelations of the U.S. Inflation Rate and Its Change, 1960:I–2004:IV

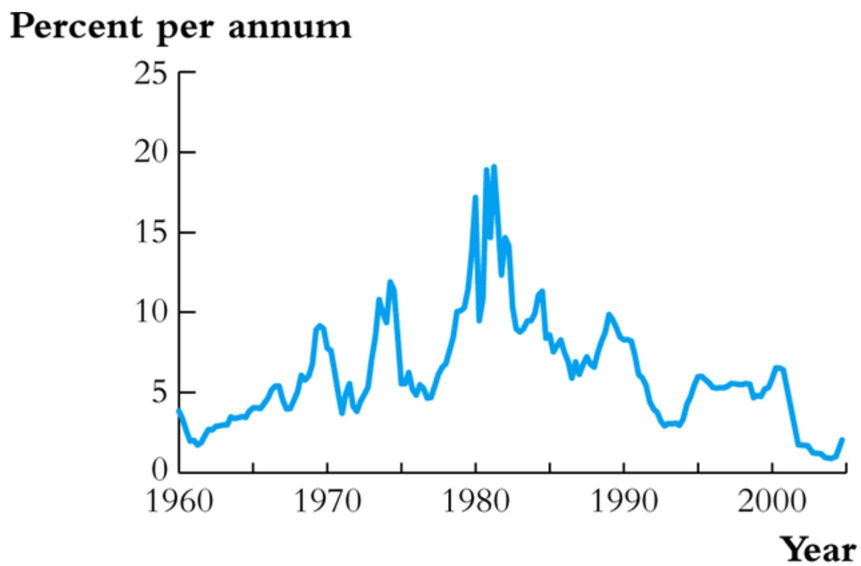
| | Autocorrelation of: | |
|---|---|---|
| **Lag** | **Inflation Rate ($Inf_t$)** | **Change of Inflation Rate ($\Delta Inf_t$)** |
| 1 | 0.84 | −0.26 |
| 2 | 0.76 | −0.25 |
| 3 | 0.76 | 0.29 |
| 4 | 0.67 | −0.06 |



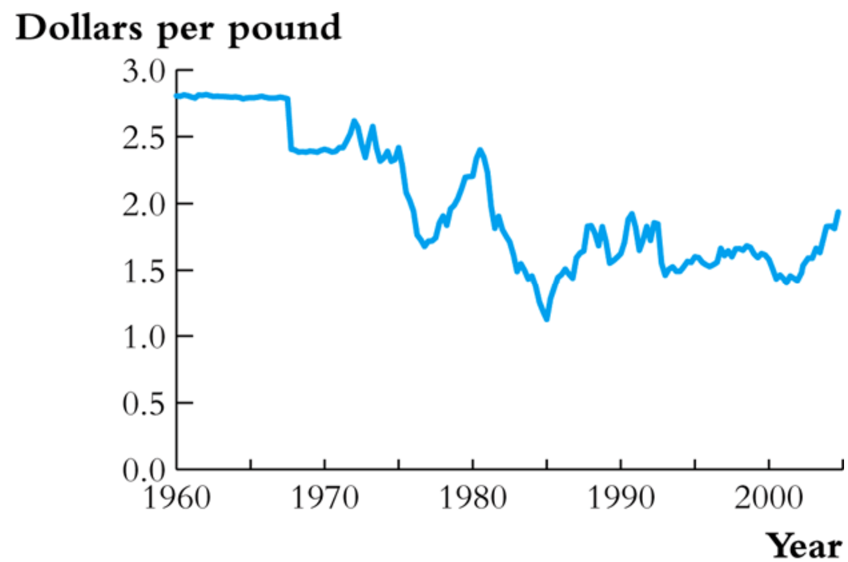
(a) U.S. CPI Inflation Rate

- The inflation rate is highly serially correlated ($\rho_1$ = .84)
- Last quarter's inflation rate contains much information about this quarter's inflation rate
- The plot is dominated by multiyear swings, with surprise movements.

# *Other economic time series*

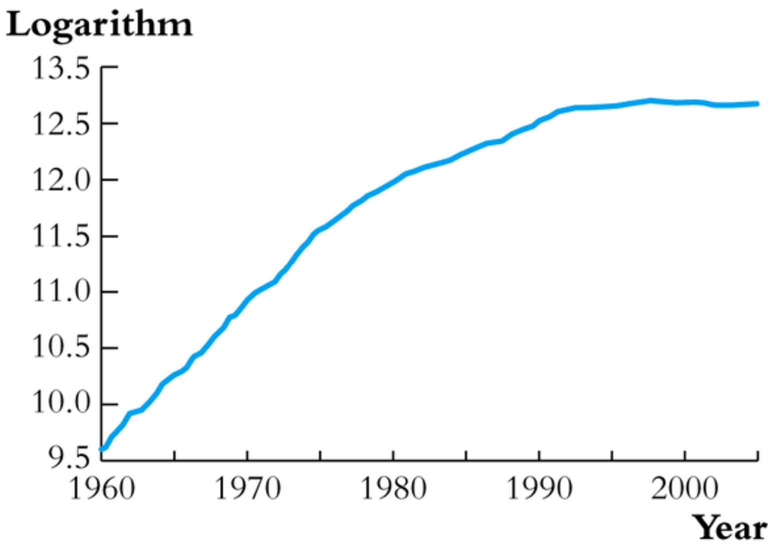



**(a)** Federal Funds Interest Rate

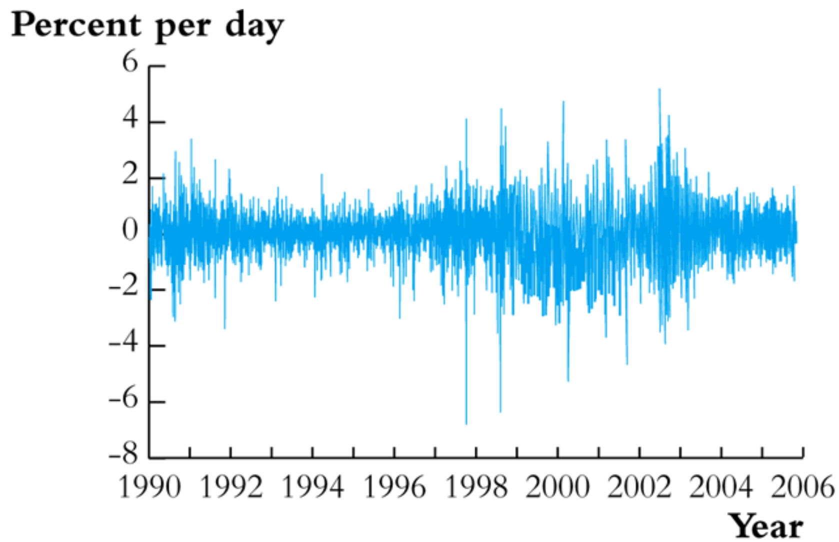



**(b)** U.S. Dollar/British Pound Exchange Rate

# *Other economic time series, ctd:*




**(c)** Logarithm of GDP in Japan




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

# D. Stationarity

Stationarity says that history is relevant. Stationarity is a key requirement for external validity of time series regression.

## Stationarity

A time series $Y_t$ is *stationary* if its probability distribution does not change over time, that is, if the joint distribution of $(Y_{s+1}, Y_{s+2}, \ldots, Y_{s+T})$ does not depend on $s$ regardless of the value of $T$; otherwise, $Y_t$ is said to be *nonstationary*. A pair of time series, $X_t$ and $Y_t$, are said to be *jointly stationary* if the joint distribution of $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \ldots, X_{s+T}, Y_{s+T})$ does not depend on $s$ regardless of the value of $T$. Stationarity requires the future to be like the past, at least in a probabilistic sense.

For now, assume that $Y_t$ is stationary (we return to this later).

# Autoregressions

- A natural starting point for a forecasting model is to use past values of Y (that is, $Y_{t-1}$, $Y_{t-2}$,…) to forecast $Y_t$.
- An **autoregression** is a regression model in which $Y_t$ is regressed against its own lagged values.
- The number of lags used as regressors is called the **order** of the autoregression.
  - **First order autoregression**: $Y_t$ is regressed on $Y_{t-1}$
  - **$p^{th}$ order autoregression**: $Y_t$ is regressed on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$.

# AR(1): First Order Autoregressive Model

The population AR(1) model is

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

- $\beta_0$ and $\beta_1$ *do not* have causal interpretations
- AR(1) model can be estimated by OLS of $Y_t$ on $Y_{t-1}$
- if $\beta_1 = 0$, $Y_{t-1}$ is not useful for forecasting $Y_t$
- Testing whether $\beta_1 = 0$ vs. $\beta_1 \neq 0$ is tantamount to testing whether $Y_{t-1}$ is not useful for forecasting $Y_t$

# Example: AR(1) of the change in inflation

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

$$\widehat{\Delta Inf_t} = 0.017 - 0.238\Delta Inf_{t-1} \quad \bar{R}^2 = 0.05$$
$$\qquad (0.126) \quad (0.096)$$

The $\bar{R}^2$ is low. Is the lagged change in inflation a useful predictor of the current change in inflation? Yes, it is:

- $t = -.238/.096 = -2.47 < -1.96$
- Reject $H_0$: $\beta_1 = 0$ at the 5% significance level
- Next, how to command Stata to find this

# Stata: AR(1) model of inflation – STATA

First, tell STATA that you are using time series data

```
generate time=q(1959q1)+_n-1;
```

*_n is the observation no. So this command creates a new variable time that has a special quarterly date format*

```
format time %tq;
```

*Specify the quarterly date format*

```
sort time;
```

*Sort by time*

```
tsset time;
```

*Let STATA know that the variable time is the variable you want to indicate*

# *Example*: AR(1) model of inflation – STATA, ctd.

```
. gen lcpi = log(cpi);                    variable cpi is already in memory

. gen inf = 400*(lcpi[_n]-lcpi[_n-1]);    quarterly rate of inflation at an
                                          annual rate
```

*This creates a new variable, inf, whose "nth" observation is 400 times the first difference of lcpi*

*compute first 8 sample autocorrelations*

```
. corrgram inf if tin(1960q1,2004q4), noplot lags(8);
```

| LAG | AC | PAC | Q | Prob>Q |
|---|---|---|---|---|
| 1 | 0.8359 | 0.8362 | 127.89 | 0.0000 |
| 2 | 0.7575 | 0.1937 | 233.5 | 0.0000 |
| 3 | 0.7598 | 0.3206 | 340.34 | 0.0000 |
| 4 | 0.6699 | -0.1881 | 423.87 | 0.0000 |
| 5 | 0.5964 | -0.0013 | 490.45 | 0.0000 |
| 6 | 0.5592 | -0.0234 | 549.32 | 0.0000 |
| 7 | 0.4889 | -0.0480 | 594.59 | 0.0000 |
| 8 | 0.3898 | -0.1686 | 623.53 | 0.0000 |

`if tin(1962q1,2004q4)` *is STATA time series syntax for using only observations between 1962q1 and 1999q4 (inclusive). The "tin(.,.)" option requires defining the time scale first, as we did above*

# *Example*: AR(1) model of inflation – STATA, ctd

```
. gen dinf = inf[_n]-inf[_n-1];
. reg dinf L.dinf if tin(1962q1,2004q4), r;     L.dinf is the first lag of dinf


Linear regression                                      Number of obs =     172
                                                       F(  1,   170) =    6.08
                                                       Prob > F      =  0.0146
                                                       R-squared     =  0.0564
                                                       Root MSE      =  1.6639

------------------------------------------------------------------------------
             |               Robust
        dinf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dinf |
         L1. |  -.2380348   .0965034    -2.47   0.015    -.4285342   -.0475354
       _cons |   .0171013   .1268831     0.13   0.893    -.2333681    .2675707
------------------------------------------------------------------------------

.  dis "Adjusted Rsquared = " _result(8);
Adjusted Rsquared = .05082278
```

# Forecasts: terminology and notation

- Predicted values are "in-sample" (usual definition)
- Forecasts are "out-of-sample" (in future)
- Notation:
  - $Y_{T+1|T}$ = forecast of $Y_{T+1}$ based on $Y_T, Y_{T-1}, \ldots$, using the population (true unknown) coefficients
  - $\hat{Y}_{T+1|T}$ = forecast of $Y_{T+1}$ based on $Y_T, Y_{T-1}, \ldots$, using the estimated coefficients from data through period T.
  - For an AR(1):
    - $Y_{T+1|T} = \beta_0 + \beta_1 Y_T$
    - $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using data through period T.

# Forecast errors

The one-period ahead forecast error is,

forecast error = $Y_{T+1} - \hat{Y}_{T+1|T}$

The distinction between a forecast error and a residual:

- residual is "in-sample"
- forecast error is "out-of-sample" – the value of $Y_{T+1}$ is not used in the estimation of the regression coefficients

# *Example*: forecasting inflation by AR(1)

AR(1) estimated using data from 1962:I – 2004:IV:

$$\widehat{\Delta Inf_t} = 0.017 - 0.238\Delta Inf_{t-1}$$

$Inf_{2004:\text{III}}$ = 1.6 (units are percent, at an annual rate)
$Inf_{2004:\text{IV}}$ = 3.5
$\Delta Inf_{2004:\text{IV}}$ = 3.5 – 1.6 = 1.9
The forecast of $\Delta Inf_{2005:\text{I}}$ is:

$$\hat{\Delta} Inf_{2005:I|2004:IV} = 0.017 - 0.238 \times 1.9 = -0.44 \approx -0.4$$

so

$$\hat{I}nf_{2005\,I|2004:IV} = Inf_{2004:\text{IV}} + \widehat{\Delta Inf}_{2005:I|2000:IV} = 3.5 - 0.4 = 3.1\%$$

# AR(p) : Forecasting from multiple lags

The $p^{\text{th}}$ 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
- The AR(1) model is a special case, p=1
- The coefficients do not have a causal interpretation
- To test the hypothesis that $Y_{t-2},\ldots,Y_{t-p}$ do not further help forecast $Y_t$, beyond $Y_{t-1}$, use an $F$-test
- Use $t$- or $F$-tests to determine the lag order $p$
- Or, better, determine $p$ using an "information criterion" (*later*)

# *Example*: AR(4) model of inflation

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

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

- *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, above and beyond the first lag, in two senses:
  - statistical: statistically significant
  - substantive: $\bar{R}^2$ increases substantially
- Next, how to command Stata to find this

# *Example*: AR(4) model of inflation – STATA

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

Linear regression Number of obs =      172
        F(  4,   167) =     7.93
        Prob > F      =   0.0000
        R-squared     =   0.2038
        Root MSE      =   1.5421
 ------------------------------------------------------------------------------
             |               Robust
        dinf |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dinf |
         L1. |  -.2579205    .0925955    -2.79   0.006    -.4407291   -.0751119
         L2. |  -.3220302    .0805456    -4.00   0.000     -.481049   -.1630113
         L3. |   .1576116    .0841023     1.87   0.063    -.0084292    .3236523
         L4. |  -.0302685    .0930452    -0.33   0.745    -.2139649    .1534278
       _cons |   .0224294    .1176329     0.19   0.849    -.2098098    .2546685
------------------------------------------------------------------------------
```

***NOTES***

- L(1/4).dinf *is A convenient way to say "use lags 1–4 of dinf as regressors"*
- *L1,…,L4 refer to the first, second, third, fourth lags of dinf*

# *Example*: AR(4) model of inflation – STATA, ctd.

```
.  dis "Adjusted Rsquared = " _result(8);
Adjusted Rsquared = .18474733
```

*result(8) is the rbar-squared of the most recently run regression*

```
.  test L2.dinf L3.dinf L4.dinf;

 ( 1)  L2.dinf = 0.0
 ( 2)  L3.dinf = 0.0
 ( 3)  L4.dinf = 0.0

       F(  3,   147) =    6.71
            Prob > F =    0.0003
```

*L2.dinf* is the second lag of dinf, etc.

# **Digression:** why $\Delta Inf$, *not Inf, in the AR's?*

The AR(1) model of $\Delta Inf_{t-1}$ *is an AR(2) model of* $Inf_t$:

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

or

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

or

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

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

They seem equivalent, upon transforming coefficients.

AR(1) model of $\Delta Inf$: $\Delta Inf_t = \beta_0 + \beta_1 \Delta Inf_{t-1} + u_t$

AR(2) model of $Inf$: $Inf_t = \gamma_0 + \gamma_1 Inf_t + \gamma_2 Inf_{t-1} + v_t$

- If equivalent, why use $\Delta Inf_t$, not $Inf_t$?

- When $Y_t$ is strongly serially correlated, the OLS estimator of the AR coefficient is biased towards zero.
- ~~In the extreme case that the AR coefficient = 1, $Y_t$ isn't stationary: the $u_t$'s accumulate and $Y_t$ blows up.~~
- If $Y_t$ isn't stationary, the regression output can be unreliable ($t$-stats need not have normal distributions, more later)
- Here, $Inf_t$ is strongly serially correlated – so we use $\Delta Inf$

# Additional Predictors: Model of Autoregressive Distributed Lag (ADL)

- So far, predictor variables have been only past values of *Y*
- Other variables might be useful predictors of *Y* in conjunction:

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

- **ADL(p,r):**
  **autoregressive distributed lags** with p lags of Y, r lags of X

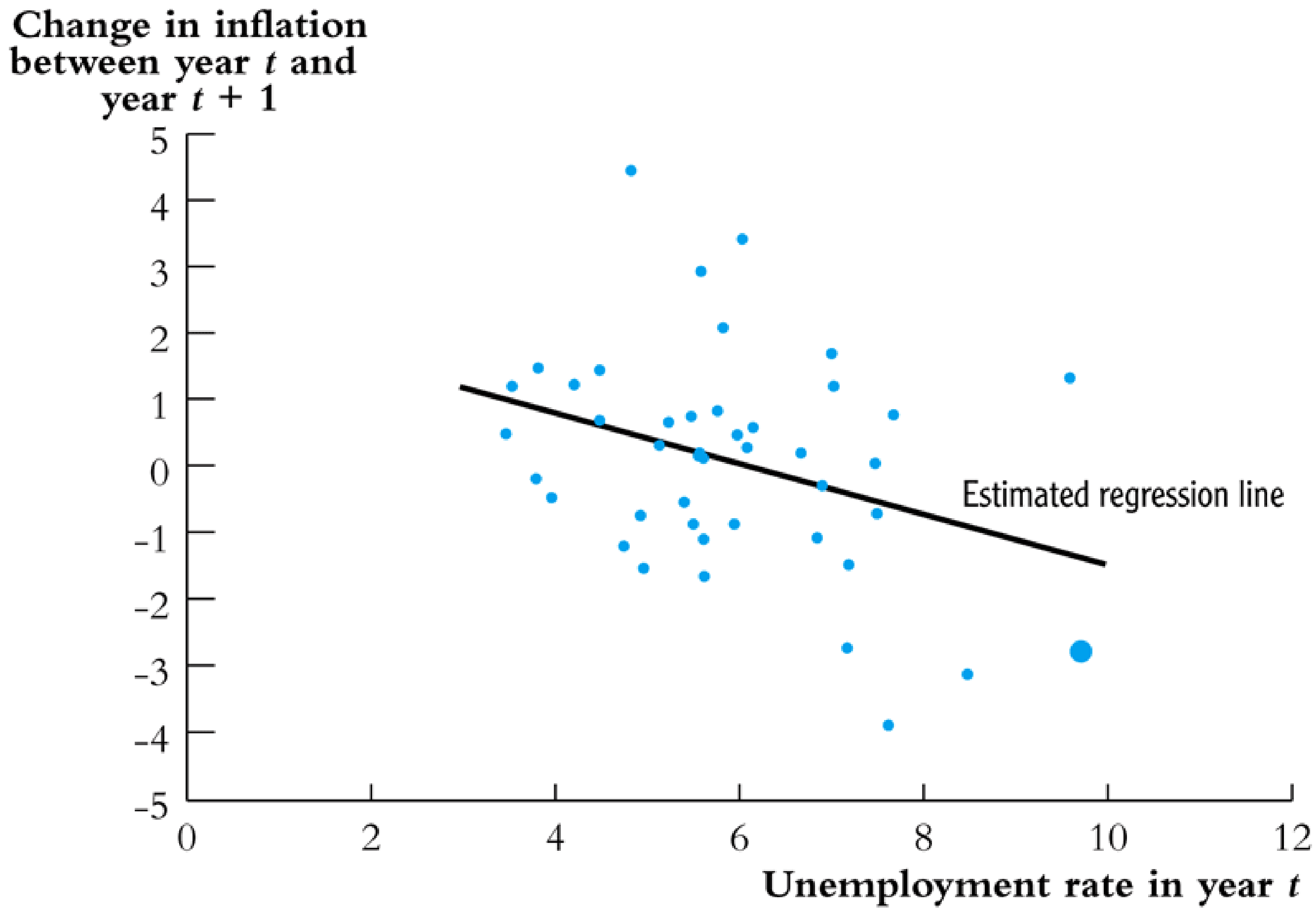
# *Example*: Inflation and Unemployment

According to the “Phillips curve,” if unemployment is above its equilibrium/natural rate, then the rate of inflation will decrease. That is, $\Delta Inf_t$ is related to lagged values of the unemployment rate, with a negative coefficient

Questions:

- Is the Phillips curve found in US economic data?
- Can it forecast inflation?
- Has the U.S. Phillips curve been stable over time?

# U.S. "Phillips Curve," annual, 1962–2004



# The Empirical Phillips Curve, ctd.

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

$$\widehat{\Delta Inf_t} = \underset{(.44)}{1.30} - \underset{(.08)}{.42}\Delta Inf_{t-1} - \underset{(.09)}{.37}\Delta Inf_{t-2} + \underset{(.08)}{.06}\Delta Inf_{t-3} - \underset{(.08)}{.04}\Delta Inf_{t-4}$$

$$- \underset{(.46)}{\mathbf{2.64}} Unem_{t-1} + \underset{(.86)}{3.04} Unem_{t-2} - \underset{(.89)}{0.38} Unem_{t-3} + \underset{(.45)}{.25} Unemp_{t-4}$$

$\bar{R}^2$ = 0.34 – a big improvement over the AR(4), for which

$\bar{R}^2$ = .18

# *Example*: *dinf* and *unem* – STATA

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

Linear regression                                  Number of obs =     172
                                                   F(  8,   163) =    8.95
                                                   Prob > F      =  0.0000
                                                   R-squared     =  0.3663
                                                   Root MSE      =  1.3926

------------------------------------------------------------------------------
             |               Robust
        dinf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dinf |
         L1. |  -.4198002   .0886973    -4.73   0.000    -.5949441   -.2446564
         L2. |  -.3666267   .0940369    -3.90   0.000    -.5523143   -.1809391
         L3. |   .0565723   .0847966     0.67   0.506    -.1108691    .2240138
         L4. |  -.0364739   .0835277    -0.44   0.663    -.2014098     .128462
        unem |
         L1. |  -2.635548   .4748106    -5.55   0.000    -3.573121   -1.697975
         L2. |   3.043123   .8797389     3.46   0.001     1.305969    4.780277
         L3. |  -.3774696   .9116437    -0.41   0.679    -2.177624    1.422685
         L4. |  -.2483774   .4605021    -0.54   0.590    -1.157696    .6609413
       _cons |   1.304271   .4515941     2.89   0.004     .4125424       2.196
------------------------------------------------------------------------------
```

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

```
.  dis "Adjusted Rsquared = " _result(8);
Adjusted Rsquared = .33516905


.  test L1.unem L2.unem L3.unem L4.unem;

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

       F(  4,   163) =    8.44
            Prob > F =    0.0000
```

*The lags of unem are significant*

*The null hypothesis that the coefficients on the lags of the unemployment rate are all zero is rejected at the 1% significance level using the F-statistic*

**The test of the joint hypothesis that none of the X's is a useful predictor, above and beyond lagged values of *Y*, is called a *Granger causality test***

## Granger Causality Tests (Tests of Predictive Content)

The Granger causality statistic is the $F$-statistic testing the hypothesis that the coefficients on all the values of one of the variables in Equation (14.20) (for example, the coefficients on $X_{1t-1}, X_{1t-2}, \ldots, X_{1t-q_1}$) 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 misnomer: Granger Causality only about marginal predictive content.*

# Forecast uncertainty & forecast intervals

Why do you need a measure of forecast uncertainty?

- To construct forecast intervals
- To 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:

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

Two components: randomness of model, randomness of sample

# Mean squared forecast error (MSFE) is:

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

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

- MSFE = var($u_{T+1}$) + uncertainty from estimation
- 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 its root:

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

# The root mean squared forecast error (RMSFE)

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

- The RMSFE measures the spread of the forecast error distribution.
- The RMSFE is like the standard deviation of $u_t$, except that it focuses on estimated coefficients, not true coefficients
- The RMSFE measures the average magnitude of the forecasting mistakes

# Three ways to estimate the RMSFE

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

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

   Usually, this isn't practical – it requires having an historical record of actual forecasts from your model

3. Use a <u>simulated</u> 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 $t$ (ignore data beyond 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 forecast error, $Y_{t+1} - \hat{Y}_{t+1|t}$
- Plug this forecast error into the MSFE formula,

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

# Using the RMSFE to construct forecast intervals

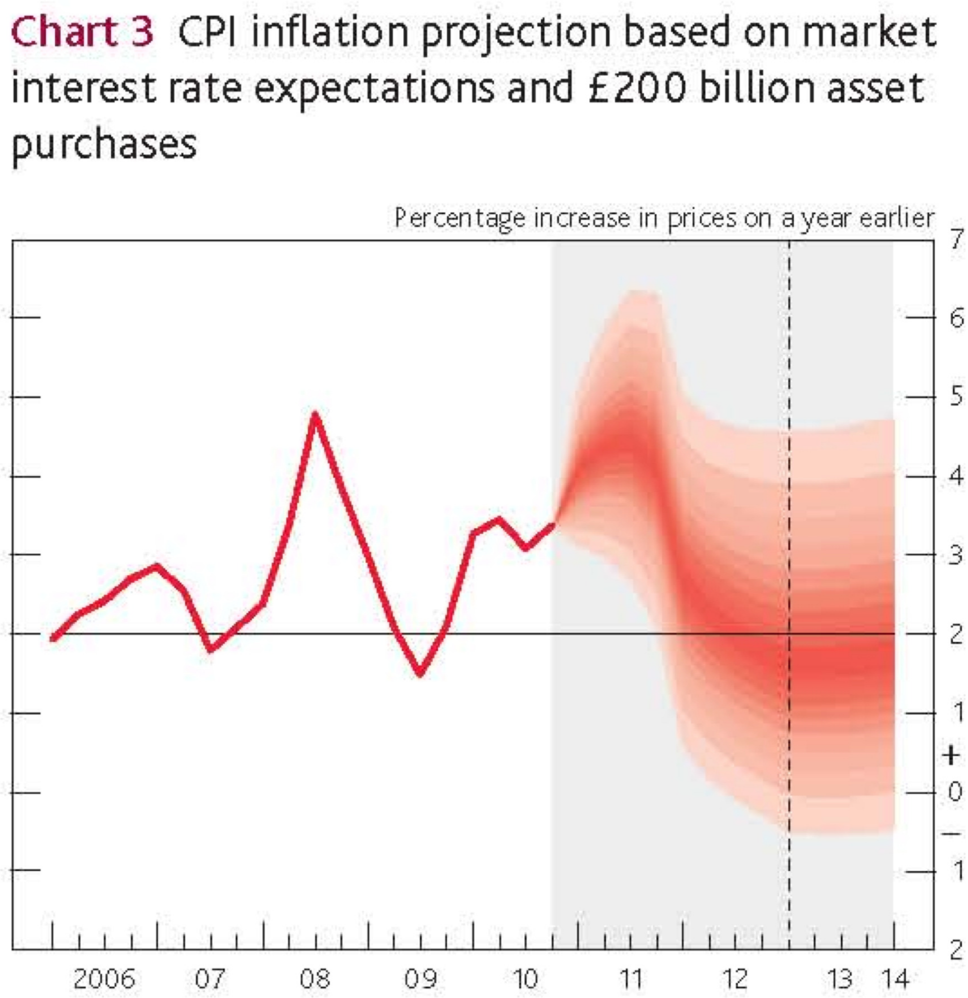
If $u_{T+1}$ is normally distributed and homoskedastic, then a 95% forecast interval can be constructed as

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

*Note*:

- A 95% forecast interval is not a confidence interval ($Y_{T+1}$ isn't an estimable coefficient, it is given random variable!)
- This interval is only valid if $u_{T+1}$ is normal, homoskedastic – but still might be a reasonable approximation and is a commonly used measure of forecast uncertainty
- Often "67%" forecast intervals are used: $\pm \widehat{RMSFE}$
- If not normal or heteroskedastic, one must model the error term prior to applying CLT … more delicate

# Example #1: The Bank of England "Fan Chart", Feb. 2011




- http://www.bankofengland.co.uk/publications/inflationreport/ir11feb.pdf

# Example #2: *Monthly Bulletin* of the European Central Bank, March 2011, staff macroeconomic projections

**Table A Macroeconomic projections for the euro area**

(average annual percentage changes) [1), 2)]

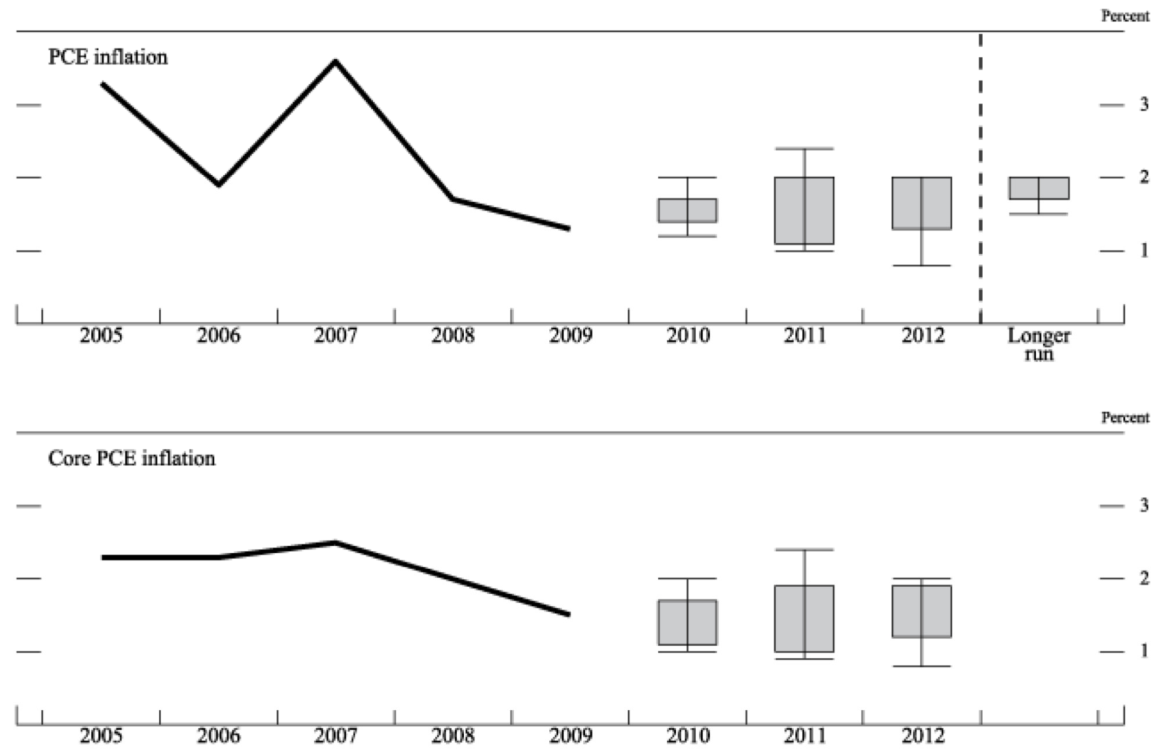
| | 2010 | 2011 | 2012 |
|---|---|---|---|
| HICP | 1.6 | 2.0 - 2.6 | 1.0 - 2.4 |
| Real GDP | 1.7 | 1.3 - 2.1 | 0.8 - 2.8 |
| Private consumption | 0.7 | 0.6 - 1.4 | 0.4 - 2.2 |
| Government consumption | 0.8 | -0.3 - 0.5 | -0.5 - 0.9 |
| Gross fixed capital formation | -0.8 | 0.4 - 3.4 | 0.7 - 5.5 |
| Exports (goods and services) | 10.9 | 4.9 - 9.5 | 3.0 - 9.2 |
| Imports (goods and services) | 9.0 | 3.5 - 7.7 | 2.8 - 8.4 |

1) The projections for real GDP and its components are based on working day-adjusted data. The projections for imports and exports include intra-euro area trade.
2) The data refer to the euro area including Estonia, except for the HICP data in 2010. The average annual percentage change for the HICP in 2011 is based on a euro area composition in 2010 that already includes Estonia.

*How did they compute these intervals*?

http://www.ecb.int/pub/mb/html/index.en.html

# Example #3: Fed, Semiannual Report to Congress, Feb. 2010: Figure 1. Central Tendencies and Ranges of Economic Projections, 2010-12 and over the Longer Run




*How did they compute these intervals*?

http://www.federalreserve.gov/boarddocs/rptcongress/annual09/default.htm

# 7. Lag Length Selection Using Information Criteria

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

- You can use sequential "downward" *t*- or *F*-tests: Start with many lags, run a test on final lags having zero coeff's. If fail to reject, re-estimate model with one fewer lag, run test again … repeat until test is rejected.
- Intuitive logic, but problematic: Chances of false rejection are high! (False rejections occur 5% of time, for each test … there are many tests!)
- Alternative to determine lags is an *information criterion*
- Information criteria trade off bias (too few lags) vs. variance of estimate (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 – avoid 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.

- How is penalty term chosen? Why this?
  - The result is that $\hat{p}^{BIC} \xrightarrow{p}$ p! (SW, App. 14.5)

# Akaike Information Criterion (AIC)

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

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

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

- *AIC* estimates more lags (larger *p*) than the *BIC*
- This might be 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 | $R^2$ |
|---|---|---|---|
| 0 | 1.095 | 1.076 | 0.000 |
| 1 | 1.067 | 1.030 | 0.056 |
| 2 | **0.955** | 0.900 | 0.181 |
| 3 | 0.957 | **0.884** | 0.203 |
| 4 | 0.986 | 0.895 | 0.204 |
| 5 | 1.016 | 0.906 | 0.204 |
| 6 | 1.046 | 0.918 | **0.204** |

- BIC chooses 2 lags, AIC chooses 3 lags.
- If you used the $R^2$ to enough digits, you would (always) select the largest possible number of lags.

# Generalization of BIC to Multivariate (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 $X$ by a Granger causality test with a fixed number of lags (number depends on the data and application)

# 8. Nonstationarity I: Trends

So far, we have assumed that the data are stationary, that is, the distribution of ($Y_{s+1}$,…, $Y_{s+T}$) doesn't depend on $s$.

Otherwise, the time series is so-called "nonstationary."

Two important types of nonstationarity are:

- Trends (SW Section 14.6)
- Structural breaks (model instability) (SW Section 14.7)

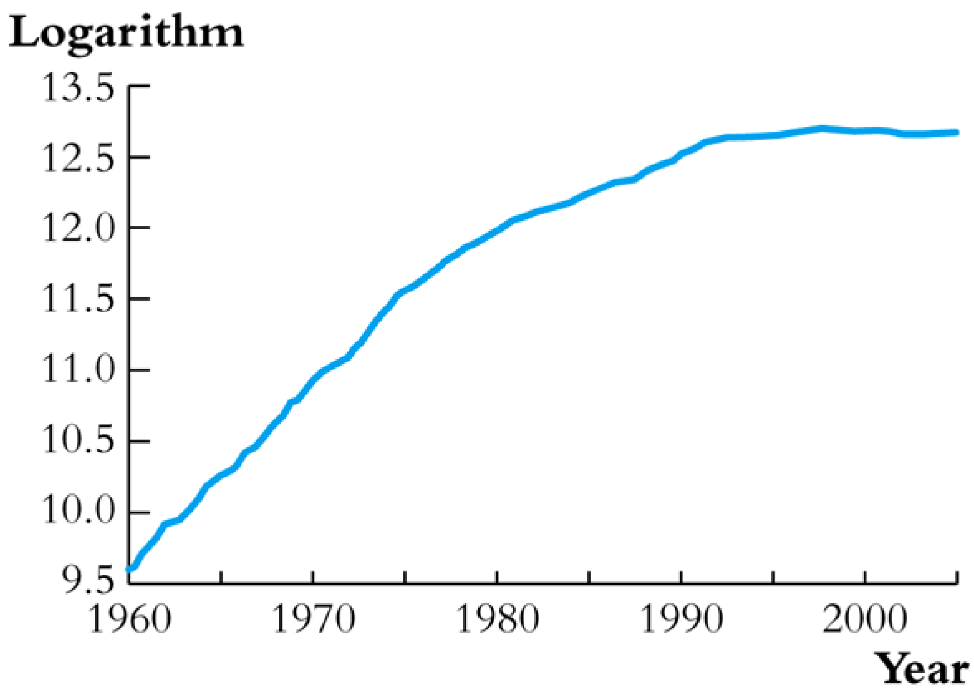
# Outline of discussion of trends in time series data:

A. What is a trend?
B. Deterministic and stochastic trends
C. What problems are caused by trends?
D. How do you detect stochastic trends (statistical tests)?
E. How to address/mitigate problems raised by trends

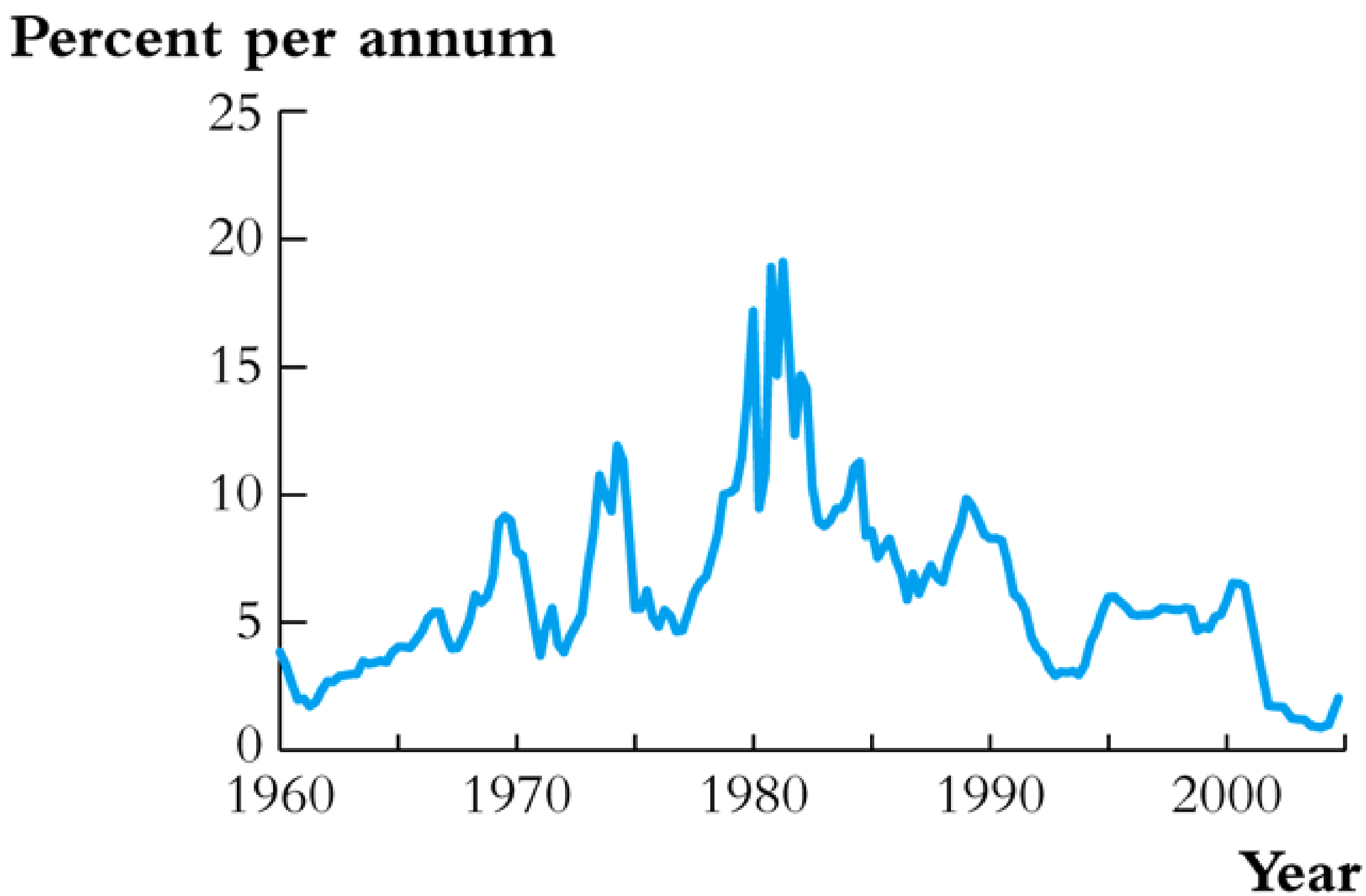
# A. What is a trend?

A trend is a persistent, long-term movement or tendency in the data. Trends need not be just a straight line!

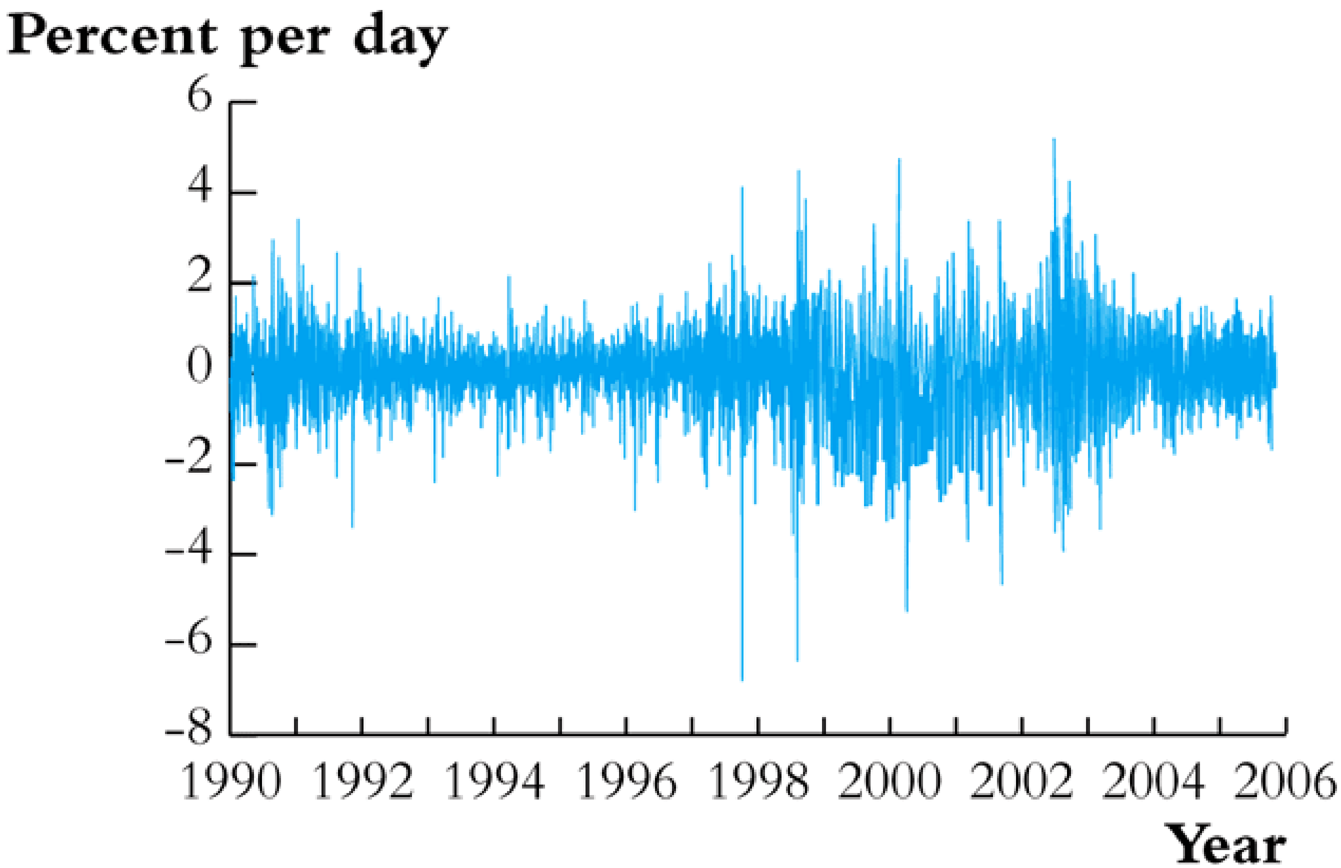
*Which of these series has a trend?*




**(c)** Logarithm of GDP in Japan



(a) Federal Funds Interest Rate




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

## *What is a trend, ctd.*

The three series:

- Log Japan GDP clearly has a long-run trend – not a straight line, but a slowly decreasing trend – fast growth during the 1960s and 1970s, slower during the 1980s, stagnating during the 1990s/2000s.
- Inflation has long-term ~~swings, periods in~~ which it is persistently high for many years ('70s/early '80s) and periods in which it is persistently low. Maybe it has a trend – hard to tell.
- NYSE daily changes has no apparent trend. There are periods of persistently high volatility – but this isn't a trend.

# B. Deterministic and stochastic trends

A trend is a long-term movement or tendency in the data.

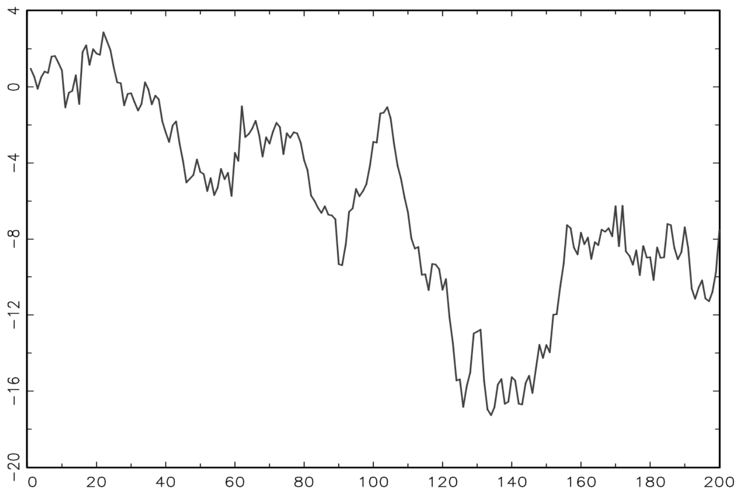
- A **deterministic trend** is a nonrandom function of time (e.g. $y_t = t$, or $y_t = t^2$).
- A **stochastic trend** is random and varies over time

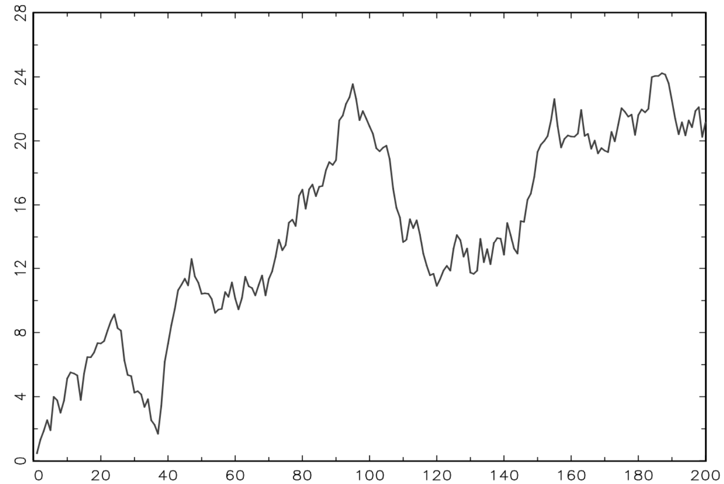
  Example: the **random walk**

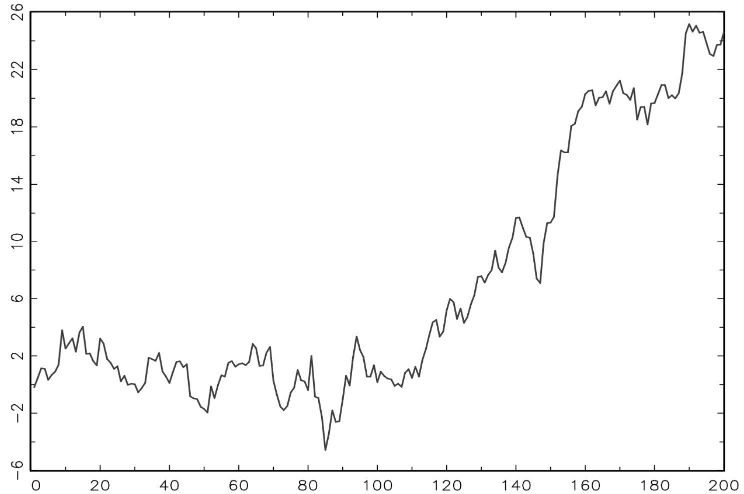
$$Y_t = Y_{t-1} + u_t, \text{ where } u_t \text{ is serially uncorrelated}$$

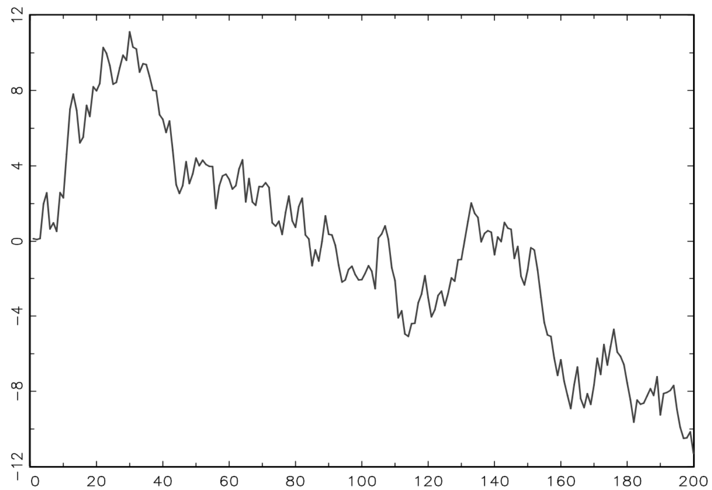
If $Y_t$ follows a random walk, then the value of $Y$ tomorrow is the value of $Y$ today, plus an unpredictable disturbance (that persists forever)

# Four artificially generated random walks, $T$ = 200:



# 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) Suppose $Y_0 = 0$. Then var($Y_t$) = $t\sigma_u^2$ .

- This ~~variance depends on *t*~~ (increases linearly with $t$), so $Y_t$ isn't stationary (recall the definition of stationarity).

# 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 $\beta_0$: 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.

# Deterministic and stochastic trends, ctd.

We will conclude this practical advice:

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

Upcoming specifics that lead to this advice:

- Relation between the random walk model and AR(1), AR(2), AR($p$) ("unit autoregressive root")
- The Dickey-Fuller test for whether a $Y_t$ has a random walk trend

# Stochastic trends & 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 from Ch 7.3:

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

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.

Thus the AR(2) model can be rearranged as,

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

where $\delta = \beta_1 + \beta_2 - 1$ and $\gamma_1 = -\beta_{2.}$

<u>Claim</u>: if $1 - \beta_1 z - \beta_2 z^2 = 0$ has a unit root, then $\beta_1 + \beta_2 = 1$

$$z = \frac{\beta_1 \pm \sqrt{\beta_1^2 + 4\beta_2}}{-2\beta_2} \quad z = 1 \Rightarrow -2\beta_2 - \beta_1 = \sqrt{\phantom{x}} \Rightarrow 4\beta_2^2 + 4\beta_1\beta_2 + \beta_1^2 = \beta_1^2 + 4\beta_2 \Rightarrow \beta_2 + \beta_1 = 1$$

So 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.**

# Unit roots in the AR($p$) model

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,

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

where

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

$\gamma_1 = -(\beta_2 + \ldots + \beta_p)$

$\gamma_2 = -(\beta_3 + \ldots + \beta_p)$

…

$\gamma_{p-1} = -\beta_p$

# Unit roots in the AR(p) model, ctd.

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

Claim: **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:**

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

# C. What problems are caused by trends?

1. AR coefficients are strongly biased towards zero. This leads to poor forecasts.

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…

# Log Japan gdp (smooth line) and US inflation (both rescaled), 1965-1981

*Seemingly a positive relationship!*

# Log Japan gdp (smooth line) and US inflation (both rescaled), 1982-1999

…relationship gone, once differences taken (just a trend in common)

# D. How do you detect stochastic trends?

1. Plot the data – are there persistent long-run movements?
2. Use a regression-based test for a random walk: the Dickey-Fuller test for a unit root.

The Dickey-Fuller test in an AR(1)

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

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.

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$$
$$H_0: \delta = 0 \text{ (that is, } \beta_1 = 1\text{) v. } H_1: \delta < 0$$

DF test: compute the *t*-statistic testing $\delta = 0$

- Under $H_0$, this $t$ statistic does **not** have a normal distribution! (Our distribution theory applies to stationary variables and $Y_t$ is nonstationary – this matters!)
- You need to use the table of Dickey-Fuller critical values. There are two cases, which have different critical values:

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

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

# Table of DF Critical Values

(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 and time trend)

**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 |

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 of the null hypothesis of a unit root (random walk trend) vs. the alternative that the autoregression is stationary.

# The Dickey-Fuller Test in an AR($p$)

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

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

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, $\delta < 1$.

The DF test in an AR($p$) (intercept only):

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

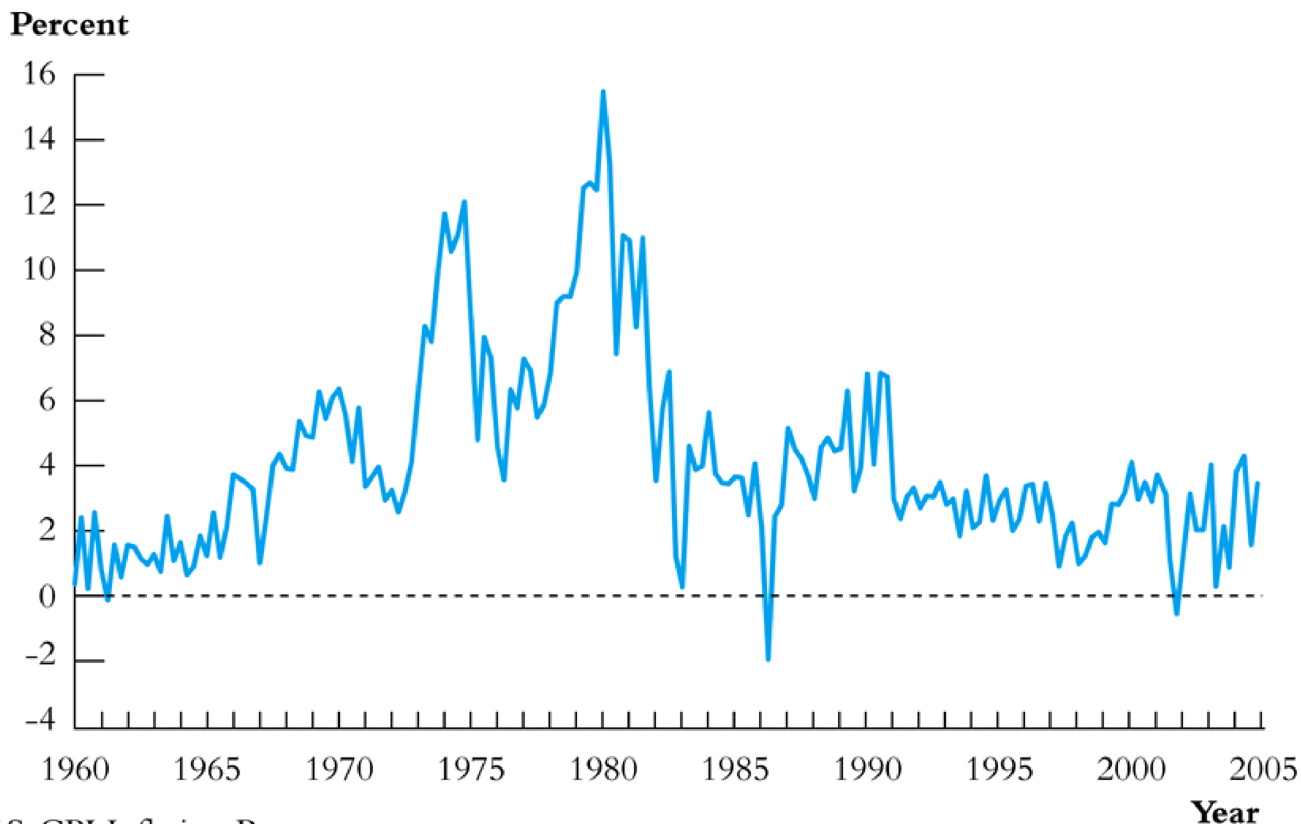
Modification for time trend: include $t$ as a regressor in (*)

# 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 – no long-term growth in the series
- In the intercept & trend specification, the alternative is that *Y* is stationary around a linear time trend – the series has long-term growth.

# *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

## **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*-statistic = –2.69**

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

# DF $t$-statistic = –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 |

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

- Some evidence of a unit root – not clear cut.
- Whether the inflation rate has a unit root is hotly debated among empirical monetary economists. *What does it mean for inflation to have a unit root?*
- We will model inflation as having a unit root.

*Note*: you can choose the lag length in the DF regression by BIC or AIC (for inflation, both reject at 10%, not 5% level)

# E. How to address/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 $Y_t$ ($\Delta Y_t$)
- 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 $\Delta Inf_t$

# 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 $Y_t$. 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 $\Delta Y_t$ for regression analysis and forecasting. If no unit root, use $Y_t$.

# 9. Nonstationarity II: Breaks

The second type of nonstationarity we consider is that the coefficients of the model are not constant over the full sample.

Clearly, it is a problem for forecasting if the model describing the historical data differs from the current model (Problem of external validity.)

So we will:

- Review two ways to detect changes in coefficients: tests for a break, and pseudo out-of-sample forecast analysis
- Work through an example: the U.S. Phillips curve

## A. Tests for a break (change) in regression coefficients

**Case I: The break date is known**

Suppose the break is known to have occurred at date $\tau$. Stability of the coefficients can be tested by estimating a fully interacted regression model. In the ADL(1,1) case:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1}$$

$$+ \gamma_0 D_t(\tau) + \gamma_1[D_t(\tau)\times Y_{t-1}] + \gamma_2[D_t(\tau)\times X_{t-1}] + u_t$$

where $D_t(\tau) = 1$ if $t \geq \tau$, and $= 0$ otherwise.

If $\gamma_0 = \gamma_1 = \gamma_2 = 0$, coefficients are constant over the sample.
If at least one of $\gamma_0$, $\gamma_1$, or $\gamma_2$ is nonzero, the regression function changes at date $\tau$.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1}$$

$$+ \gamma_0 D_t(\tau) + \gamma_1[D_t(\tau)\times Y_{t-1}] + \gamma_2[D_t(\tau)\times X_{t-1}] + u_t$$

where $D_t(\tau) = 1$ if $t \geq \tau$, and $= 0$ otherwise

The **Chow test statistic** for a break at date $\tau$ is the (heteroskedasticity-robust) $F$-statistic that tests:

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

vs. $H_1$: at least one of $\gamma_0$, $\gamma_1$, or $\gamma_2$ is nonzero

- Note that you can apply this to a subset of the coefficients, e.g. only the coefficient on $X_{t-1}$.
- Limitation: test presumes known/suspected break date … what if $\tau$ is unknown?

# Case II: The break date is unknown

Why consider this case?

- You might suspect there is a break, but not know when
- You might want to test the null hypothesis of coefficient stability against the general alternative that there has been a break sometime.
- Even if you think you know the break date, if that "knowledge" is based on prior inspection of the series then you have in effect "estimated" the break date. This invalidates the Chow test critical values!

# **The Quandt Likelihood Ratio (QLR) Statistic** (also called the "sup-Wald" statistic)

The QLR statistic = the maximum Chow statistic

- Let $F(\tau)$ = the Chow test statistic testing the hypothesis of no break at date $\tau$.
- The **QLR test statistic** is the maximum of all the Chow $F$-statistics on a range of $\tau$, $\tau_0 \leq \tau \leq \tau_1$:

$$QLR = \max[F(\tau_0), F(\tau_0+1), \ldots, F(\tau_1-1), F(\tau_1)]$$

- A conventional choice for $\tau_0$ and $\tau_1$ are the inner 70% of the sample (exclude the first and last 15%).
- Should you use the usual $F_{q,\infty}$ critical values?

# *The QLR test, ctd.*

$QLR = \max[F(\tau_0), F(\tau_0+1), \ldots, F(\tau_1-1), F(\tau_1)]$

- The large-sample null distribution of $F(\tau)$ for a given (fixed, not estimated) $\tau$ is $F_{q,\infty}$
- However, you may be estimating $\tau$ multiple times. The critical value must be larger than $F_{q,\infty}$
- Critical values are tabulated in SW Table 14.6…

- ***Get this***: in large samples, *QLR* has the distribution,

$$\max_{a \le s \le 1-a}\left(\frac{1}{q}\sum_{i=1}^{q}\frac{B_i'(s)^2}{s(1-s)}\right)$$

where $\{B_i\}$, $i$ =1,…,$n$, are independent continuous-time "Brownian Bridges" on $0 \le s \le 1$ (a Brownian Bridge is a Brownian motion deviated from its mean; a Brownian motion is a Gaussian [normally-distributed] random walk in continuous time), and where $a$ = .15 (exclude first and last 15% of the sample)

- Critical values are tabulated in SW Table 14.6…

**TABLE 14.6** Critical Values of the QLR Statistic with 15% Trimming

| Number of Restrictions ($q$) | 10% | 5% | 1% |
|---|---|---|---|
| 1 | 7.12 | 8.68 | 12.16 |
| 2 | 5.00 | 5.86 | 7.78 |
| 3 | 4.09 | 4.71 | 6.02 |
| 4 | 3.59 | 4.09 | 5.12 |
| 5 | 3.26 | 3.66 | 4.53 |
| 6 | 3.02 | 3.37 | 4.12 |
| 7 | 2.84 | 3.15 | 3.82 |
| 8 | 2.69 | 2.98 | 3.57 |
| 9 | 2.58 | 2.84 | 3.38 |
| 10 | 2.48 | 2.71 | 3.23 |
| 11 | 2.40 | 2.62 | 3.09 |
| 12 | 2.33 | 2.54 | 2.97 |
| 13 | 2.27 | 2.46 | 2.87 |
| 14 | 2.21 | 2.40 | 2.78 |
| 15 | 2.16 | 2.34 | 2.71 |
| 16 | 2.12 | 2.29 | 2.64 |
| 17 | 2.08 | 2.25 | 2.58 |
| 18 | 2.05 | 2.20 | 2.53 |
| 19 | 2.01 | 2.17 | 2.48 |
| 20 | 1.99 | 2.13 | 2.43 |

These critical values apply when $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$ (rounded to the nearest integer), so the $F$-statistic is computed for all potential break dates in the central 70% of the sample. The number of restrictions $q$ is the number of restrictions tested by each individual $F$-statistic. Critical values for other trimming percentages are given in Andrews (2003).

Note that these critical values are larger than the $F_{q,\infty}$ critical values – for example, $F_{1,\infty}$ 5% critical value is 3.84.

# *Example*: Has the postwar U.S. Phillips Curve been stable?

Recall the ADL(4,4) model of $\Delta Inf_t$ and $Unemp_t$ – the empirical backwards-looking Phillips curve, estimated over (1962 – 2004):

$$\widehat{\Delta Inf_t} = \underset{(.44)}{1.30} - \underset{(.08)}{.42}\Delta Inf_{t-1} - \underset{(.09)}{.37}\Delta Inf_{t-2} + \underset{(.08)}{.06}\Delta Inf_{t-3} - \underset{(.08)}{.04}\Delta Inf_{t-4}$$

$$- \underset{(.46)}{2.64} Unem_{t-1} + \underset{(.86)}{3.04} Unem_{t-2} - \underset{(.89)}{0.38} Unem_{t-3} + \underset{(.45)}{.25} Unemp_{t-4}$$

Has this model been stable over the full period 1962-2004?

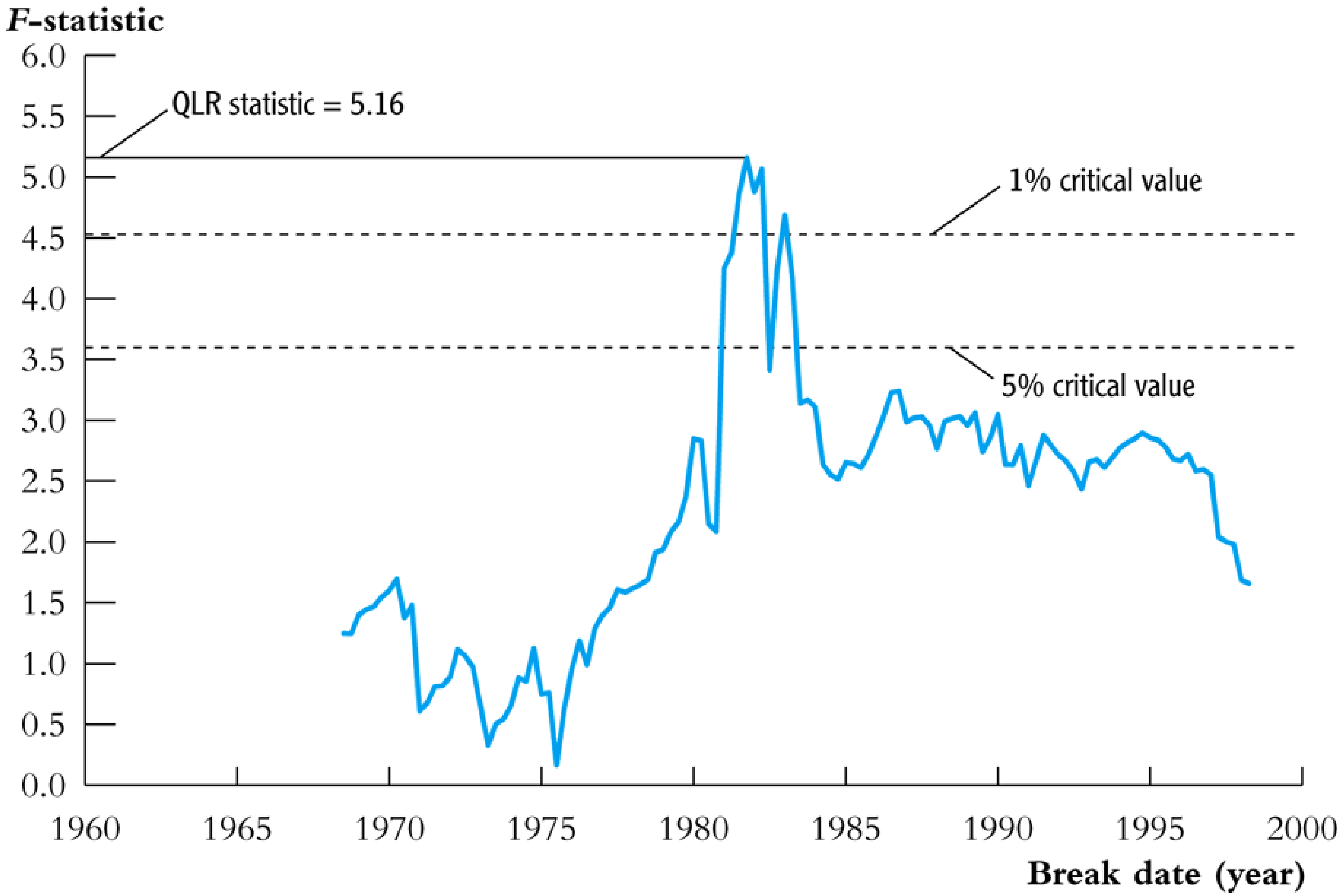
# QLR tests of the stability of the U.S. Phillips curve.

dependent variable: $\Delta Inf_t$

regressors: intercept, $\Delta Inf_{t-1}$,…, $\Delta Inf_{t-4}$, $Unemp_{t-1}$,…, $Unemp_{t-4}$

- test for constancy of intercept only (other coefficients are assumed constant): $QLR$ = 2.865 ($q$ = 1).
  - 10% critical value = 7.12 → don't reject at 10% level
- test for constancy of intercept and coefficients on $Unemp_t$,…, $Unemp_{t-3}$ (coefficients on $\Delta Inf_{t-1}$,…, $\Delta Inf_{t-4}$ are constant): $QLR$ = 5.158 ($q$ = 5)
  - 1% critical value = 4.53 → reject at 1% level
  - Estimate break date: maximal $F$ occurs in 1981:IV
- Conclude that there is a break in the inflation – unemployment relation, with estimated date of 1981:IV

# ***Figure 14.5*** *F-Statistics Testing for a Break in Equation (14.17) at Different Dates*



# B. Assessing Model Stability using Pseudo Out-of-Sample Forecasts

- The QLR test does not work well towards the very end of the sample – but this is usually the most interesting part!
- One way to check whether the model is working at the end of the sample is to see whether the pseudo out-of-sample (*poos*) forecasts are "on track" in the most recent observations. This is an informal diagnostic (not a formal test) which complements formal testing using the QLR.

# Application to the U.S. Phillips Curve:

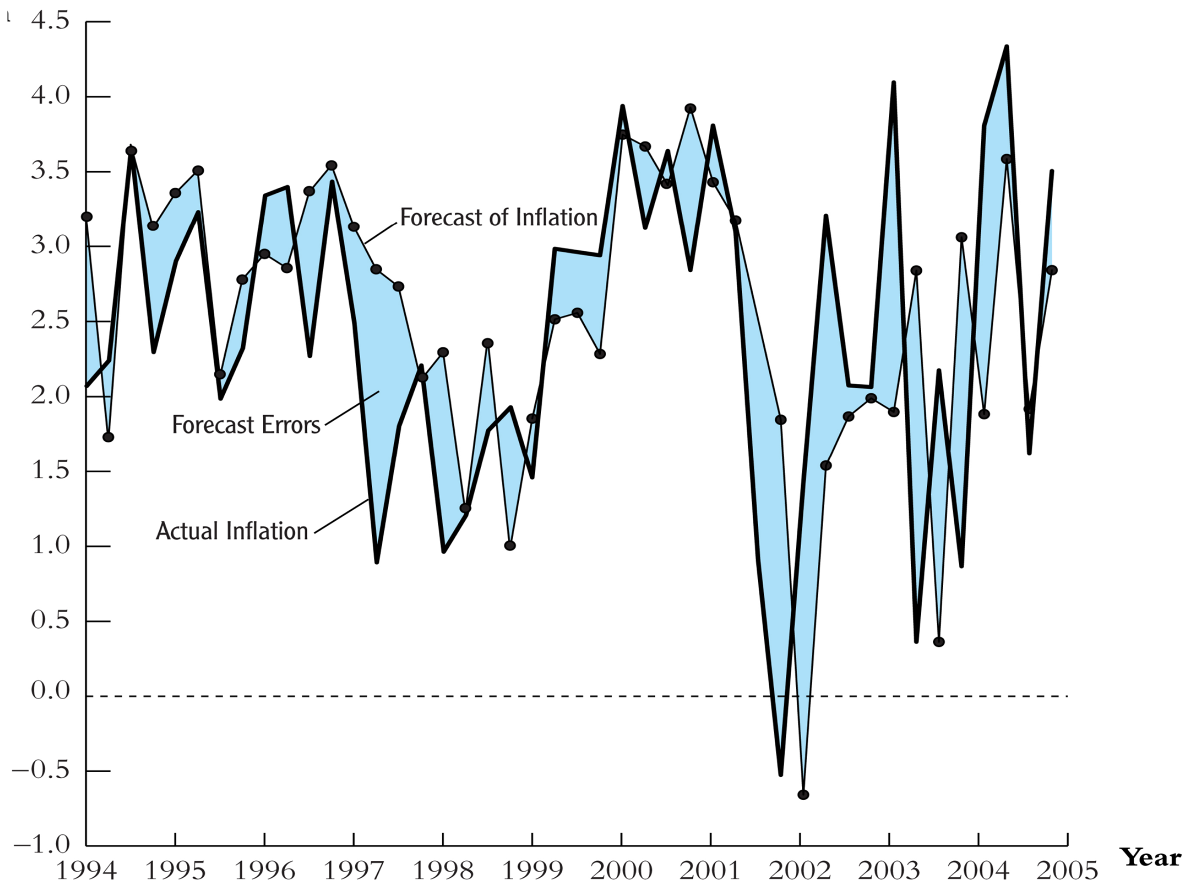
*Was the U.S. Phillips Curve stable toward the end of the sample?*

- We found a break in 1981:IV – so for this analysis, we only consider regressions that start in 1982:I – ignore the earlier data from the "old" model ("regime").
- <u>Regression model:</u>

  dependent variable: $\Delta Inf_t$

  regressors: 1, $\Delta Inf_{t-1}$,…, $\Delta Inf_{t-4}$, $Unemp_{t-1}$,…, $Unemp_{t-4}$
- Pseudo out-of-sample forecasts:
  - Compute regression over $t$ = 1982:I,…, $P$
  - Compute *poos* forecast, $\widehat{\Delta Inf}_{P+1|P}$, and forecast error
  - Repeat for $P$ = 1994:I,…, 2005:I

## *POOS* forecasts of $\Delta$*Inf* using ADL(4,4) model with *Unemp*




There are some big forecast errors (in 2001) but they do not appear to be getting bigger – the model isn't deteriorating

# *poos forecasts using the Phillips curve, ctd.*

Some summary statistics:

- Mean forecast error, 1999:I – 2004:IV = 0.11 (*SE* = 0.27)
  - No evidence that the forecasts are systematically to high or too low
- *poos RMSFE*, 1999:I – 2004:IV: 1.32
- *SER*, model fit 1982:I – 1998:IV: 1.30
  - The *poos RMSFE* ≈ the in-sample *SER* – another indication that forecasts are not doing any worse (or better) out of sample than in-sample

This analysis suggests that there was not a substantial change in the forecasts produced by the ADL(4,4) model towards the end of the sample

# 10. Conclusion: Time Series Forecasting Models (SW Section 14.8)

- For forecasting purposes, it isn't important to have coefficients with a causal interpretation!

- The tools of regression can be used to construct reliable forecasting models – even though there is no causal interpretation of the coefficients:
  - AR($p$) – common "benchmark" models
  - ADL($p$,$q$) – add $q$ lags of $X$ (another predictor)
  - Granger causality tests – test whether a variable $X$ and its lags are useful for predicting $Y$ given lags of $Y$.

# Conclusion, ctd.

- New ideas and tools:
  - stationarity
  - forecast intervals using the RMSFE
  - pseudo out-of-sample forecasting
  - BIC for model selection
  - Ways to check/test for nonstationarity:
    - Dickey-Fuller test for a unit root (stochastic trend)
    - Test for a break in regression coefficients:
      - Chow test at a known date
      - QLR test at an unknown date
    - *poos* analysis for end-of-sample forecast breakdown