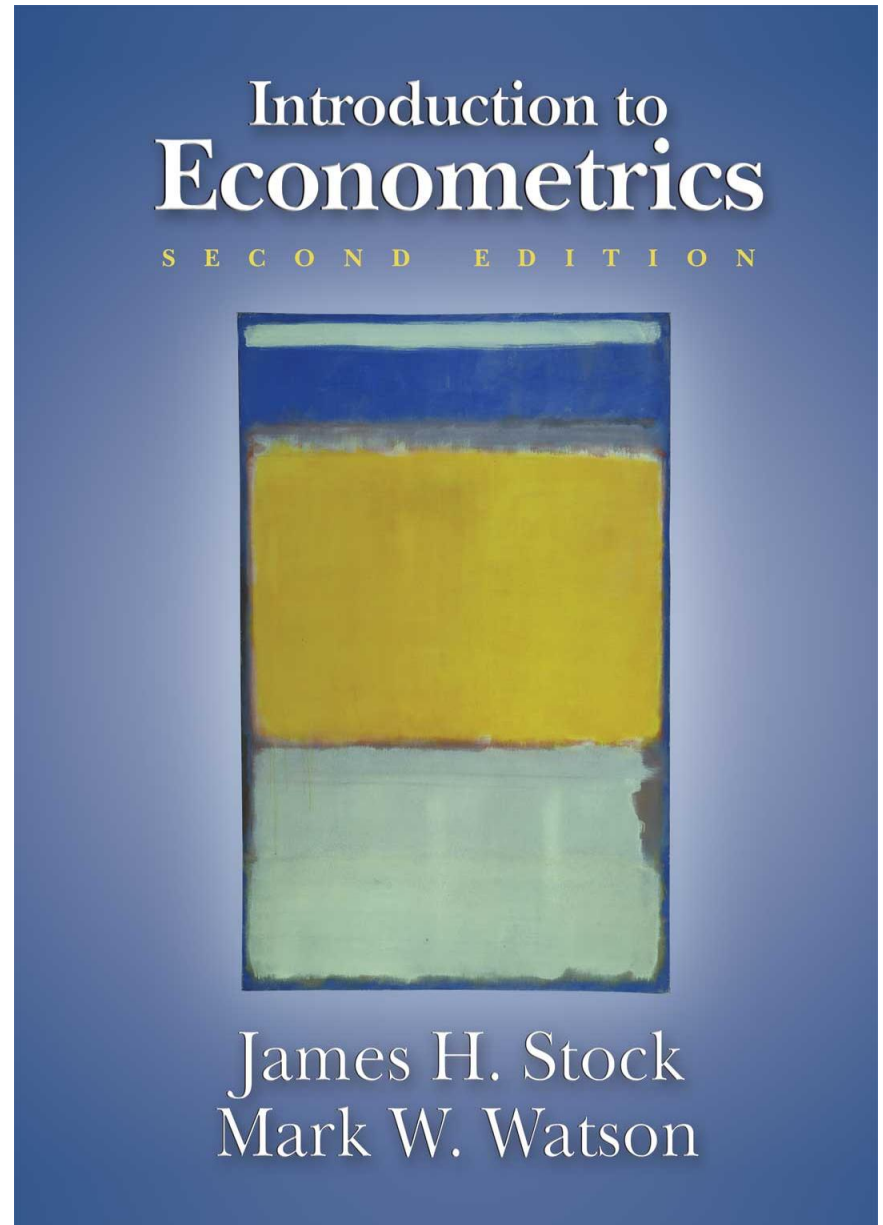


Chapter 14

Introduction to Time Series Regression and Forecasting



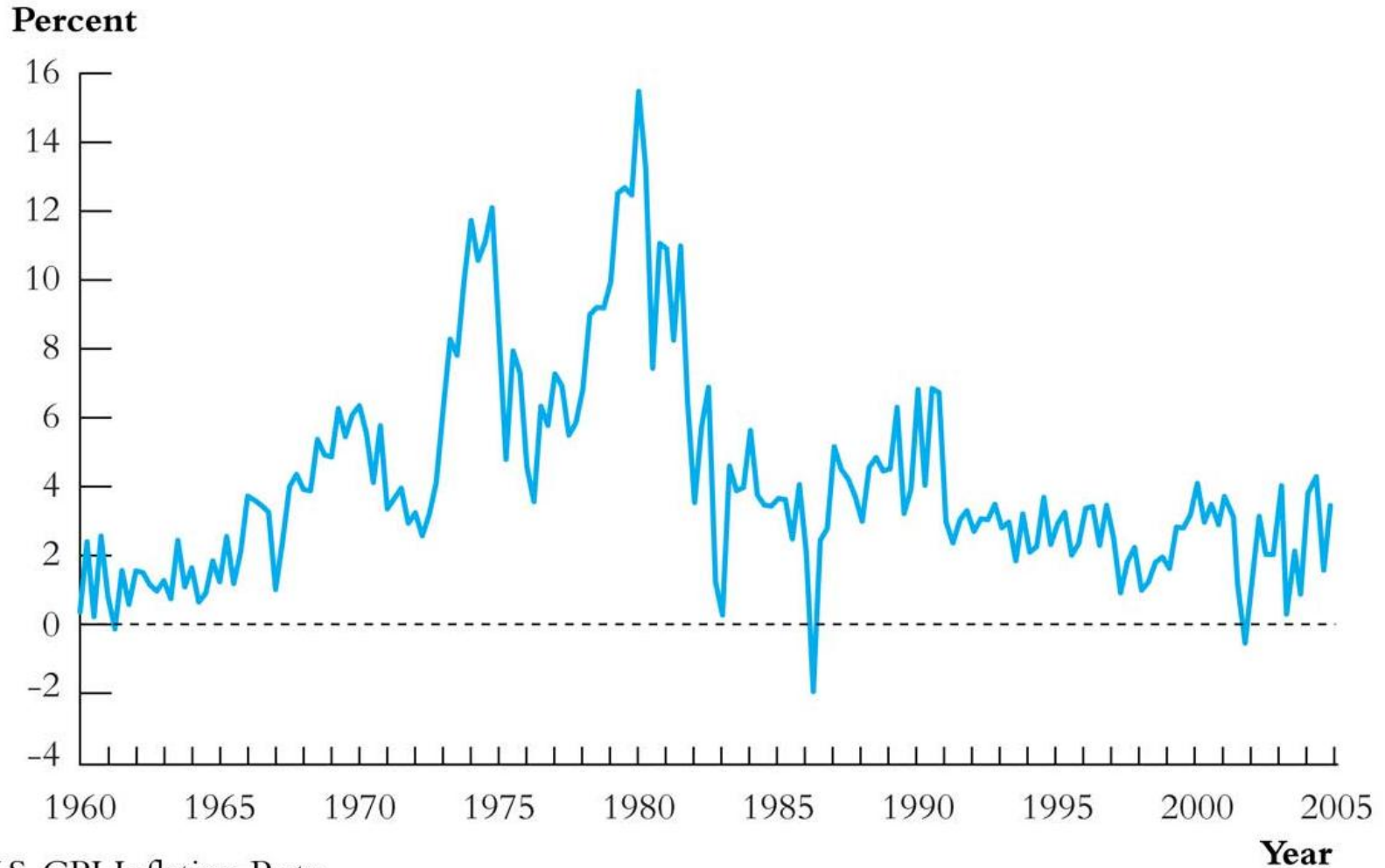
Introduction to Time Series Regression and Forecasting

(SW Chapter 14)

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

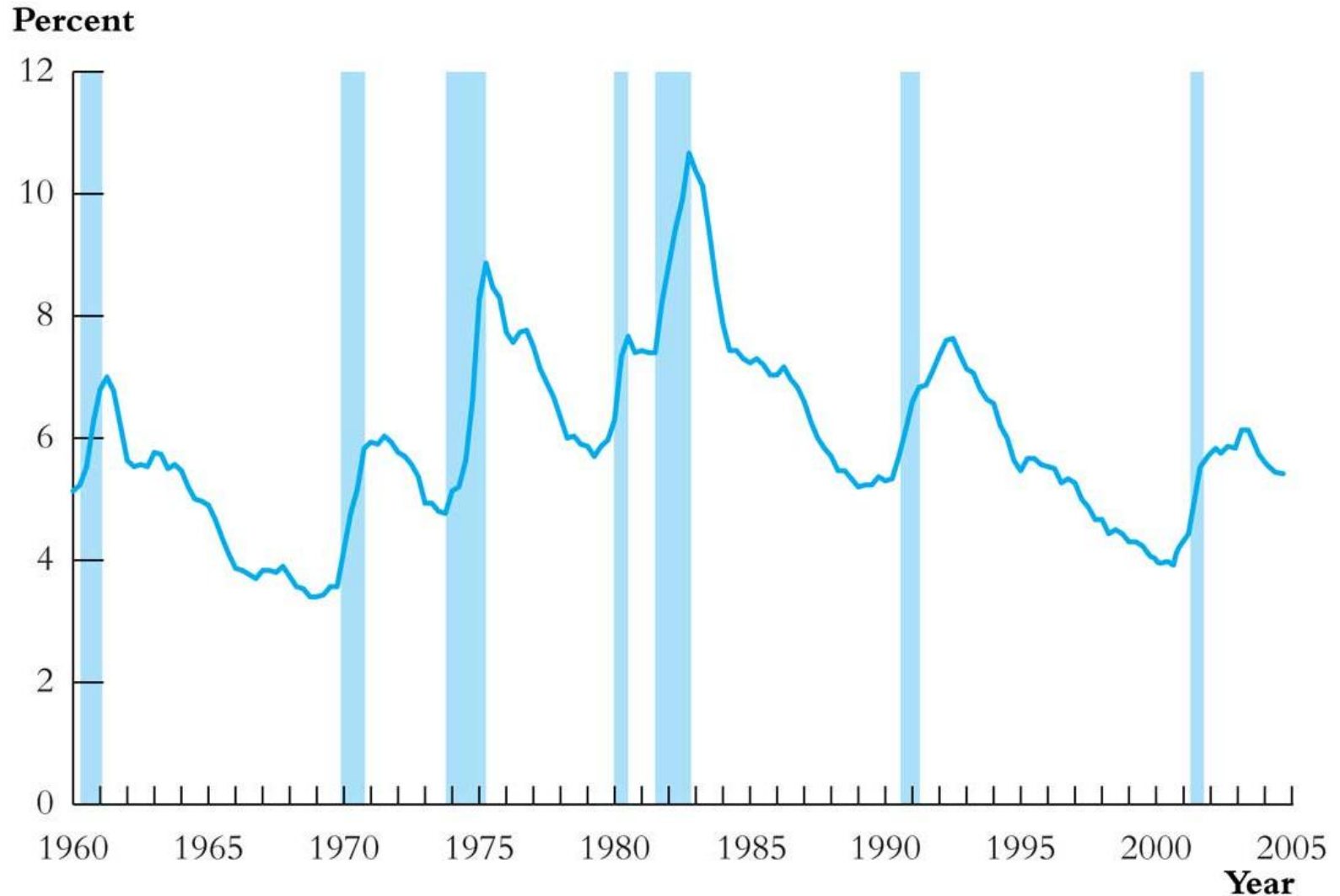
- Aggregate consumption and GDP for a country (for example, 20 years of quarterly observations = 80 observations)
- Yen/\$, pound/\$ and Euro/\$ exchange rates (daily data for 1 year = 365 observations)
- Cigarette consumption per capita in a state, by year

Example #1 of time series data: US rate of price inflation, as measured by the quarterly percentage change in the Consumer Price Index (CPI), at an annual rate



(a) U.S. CPI Inflation Rate

Example #2: US rate of unemployment



(b) U.S. Unemployment Rate

Why use time series data?

- To develop forecasting models
 - What will the rate of inflation be next year?
- To estimate dynamic causal effects
 - If the Fed increases the Federal Funds rate now, what will be the effect on the rates of inflation and unemployment in 3 months? in 12 months?
 - What is the effect over time on cigarette consumption of a hike in the cigarette tax?
- Or, because that is your only option ...
 - Rates of inflation and unemployment in the US can be observed only over time!

Time series data raises new technical issues

- Time lags
- Correlation over time (*serial correlation*, a.k.a. *autocorrelation*)
- Forecasting models built on regression methods:
 - *autoregressive* (AR) models
 - *autoregressive distributed lag* (ADL) models
 - need not (typically do not) have a causal interpretation
- Conditions under which dynamic effects can be estimated, and how to estimate them
- Calculation of standard errors when the errors are serially correlated

Using Regression Models for Forecasting (SW Section 14.1)

- 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!
 - We will not worry about interpreting coefficients in forecasting models
 - External validity is paramount: the model estimated using historical data must hold into the (near) future

Introduction to Time Series Data and Serial Correlation (SW Section 14.2)

First, some notation and terminology.

Notation for time series data

- Y_t = value of Y in period t .
- Data set: $Y_1, \dots, Y_T = T$ observations on the time series random variable Y
- We consider only consecutive, evenly-spaced observations (for example, monthly, 1960 to 1999, no missing months) (missing and non-evenly spaced data introduce technical complications)

We will transform time series variables using lags, first differences, logarithms, & growth rates

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.

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*
$$= 4 \times 1.088 = \mathbf{4.359\%} \approx \mathbf{4.4\%}$$
 (percent per year)
- Like interest rates, inflation rates are (as a matter of convention) reported at an annual rate.
- Using the logarithmic approximation to percent changes yields
$$4 \times 100 \times [\log(188.60) - \log(186.57)] = \mathbf{4.329\%}$$

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 (ΔInf_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, times 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.

Autocorrelation

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

- The first autocorrelation of Y_t is $\text{corr}(Y_t, Y_{t-1})$
- The first *autocovariance* of Y_t is $\text{cov}(Y_t, Y_{t-1})$
- 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 the population joint distribution of (Y_t, Y_{t-1})

AUTOCORRELATION (SERIAL CORRELATION) AND AUTOCOVARANCE

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,

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

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

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:

$$\hat{\rho}_j = \frac{\overline{\text{cov}}(Y_t, Y_{t-j})}{\sqrt{\text{var}}(Y_t)}$$

where

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

where $\bar{Y}_{j+1,T}$ is the sample average of Y_t computed over observations $t = j+1, \dots, T$. *NOTE:*

- the summation is over $t=j+1$ to T (*why?*)
- The divisor is T , not $T - j$ (this is the conventional definition used for time series data)

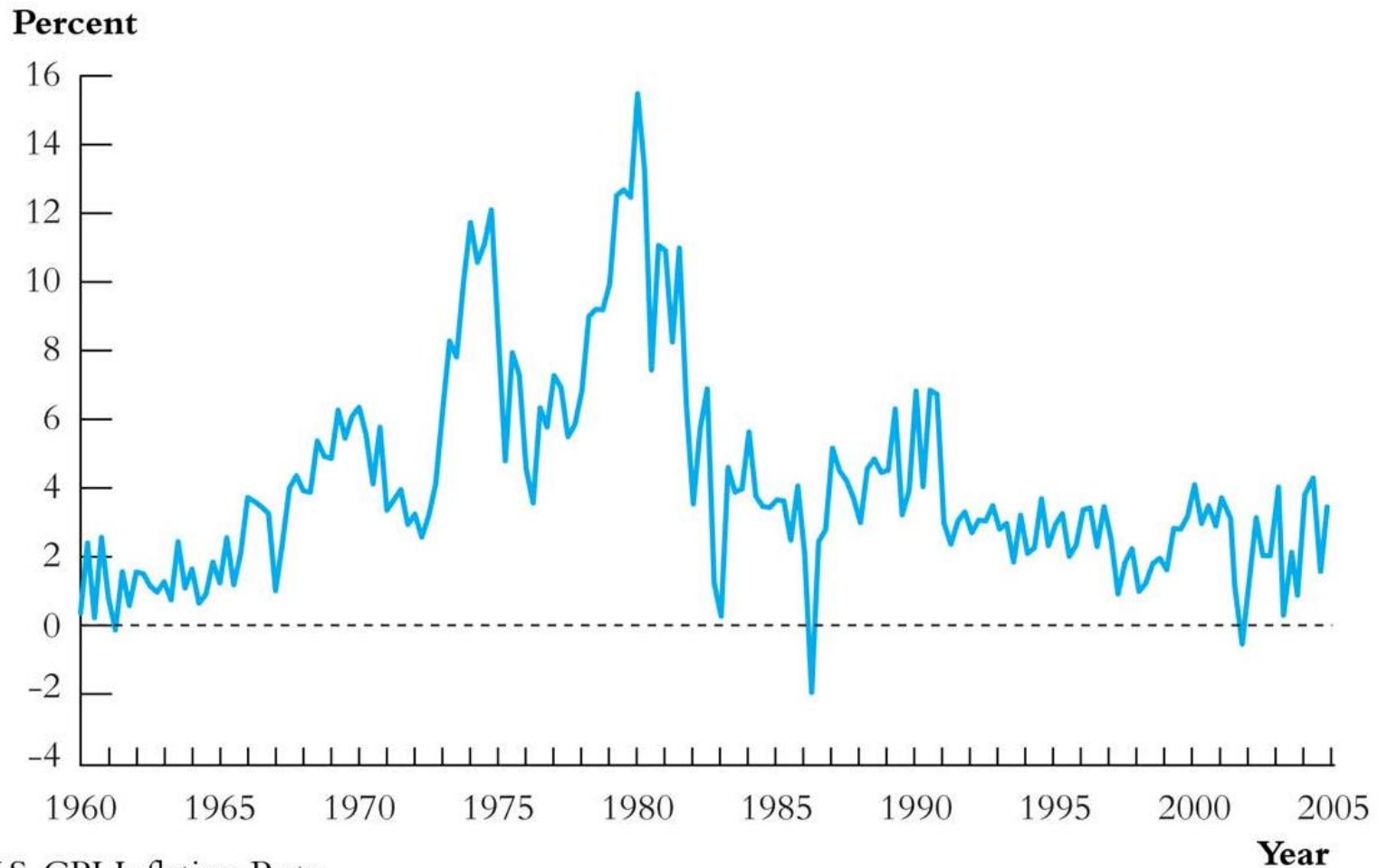
Example:

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

Lag	Autocorrelation of:	
	Inflation Rate (Inf_t)	Change of Inflation Rate (ΔInf_t)
1	0.84	−0.26
2	0.76	−0.25
3	0.76	0.29
4	0.67	−0.06

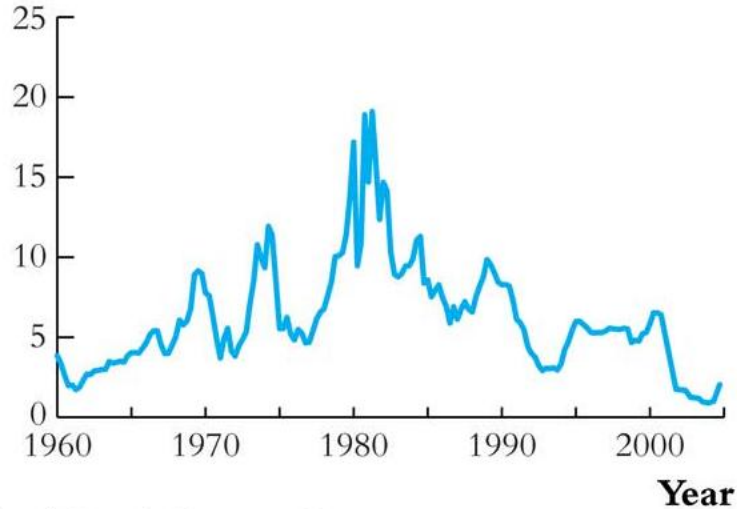


(a) U.S. CPI Inflation Rate

- The inflation rate is highly serially correlated ($\phi_1 = .84$)
- Last quarter's inflation rate contains much information about this quarter's inflation rate
- The plot is dominated by multiyear swings
- But there are still surprise movements!

Other economic time series:

Percent per Annum



(a) Federal Funds Interest Rate

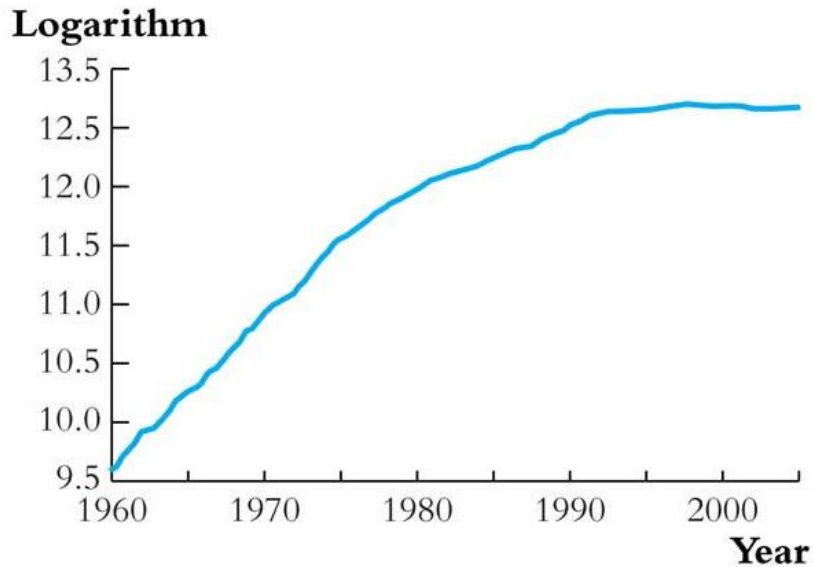
Dollars per Pound



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

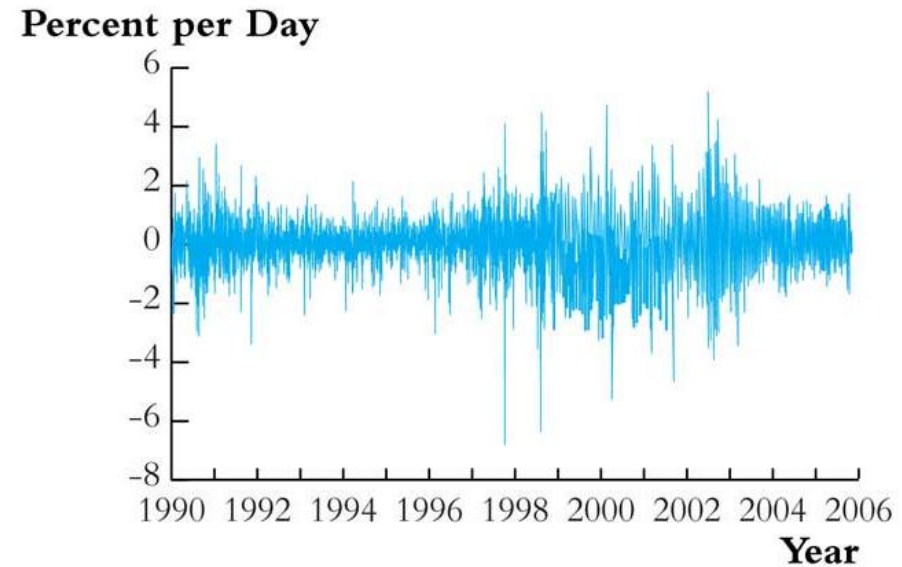
Other economic time series, ctd:

(a) Federal Funds Interest Rate



(c) Logarithm of GDP in Japan

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



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

Stationarity: a key requirement for external validity of time series regression

Stationarity says that history is relevant:

STATIONARITY

A time series Y_t is *stationary* if its probability distribution does not change over time, that is, if the joint distribution of $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ does not depend on s ; 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 $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \dots, X_{s+T}, Y_{s+T})$ does not depend on s . 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

(SW Section 14.3)

A natural starting point for a forecasting model is to use past values of Y (that is, Y_{t-1}, Y_{t-2}, \dots) 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.
 - In a *first order autoregression*, Y_t is regressed against Y_{t-1}
 - In a p^{th} *order autoregression*, Y_t is regressed against $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$.

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 regression of 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

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

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

$$\begin{aligned} \Delta Inf_t &= 0.017 - 0.238 \Delta Inf_{t-1} \quad \bar{R}^2 = 0.05 \\ &\quad (0.126) \quad (0.096) \end{aligned}$$

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

- $t = -.238/.096 = -2.47 > 1.96$ (in absolute value)
- \Rightarrow 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!

Example: AR(1) model of inflation – STATA

First, let STATA know 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 the time scale

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, the "nth" observation of which is 400 times difference between the nth observation on lcpi and the "n-1"th observation on that is, 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 bet 1962q1 and 1999q4 (inclusive). The "tin(.,.)" option requires defining the time 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
```

dinf	Coef.	Robust 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” (the usual definition)
- *Forecasts* are “out-of-sample” – in the future
- *Notation:*
 - $Y_{T+1|T}$ = forecast of Y_{T+1} based on Y_T, Y_{T-1}, \dots , using the population (true unknown) coefficients
 - $\hat{Y}_{T+1|T}$ = forecast of Y_{T+1} based on Y_T, Y_{T-1}, \dots , using the estimated coefficients, which are estimated using 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,

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

The distinction between a forecast error and a residual is the same as between a forecast and a predicted value:

- a *residual* is “in-sample”
- a *forecast error* is “out-of-sample” – the value of Y_{T+1} isn’t used in the estimation of the regression coefficients

Example: forecasting inflation using an AR(1)

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

$$\overline{\Delta Inf}_t = 0.017 - 0.238\Delta Inf_{t-1}$$

$Inf_{2004:III} = 1.6$ (units are percent, at an annual rate)

$Inf_{2004:IV} = 3.5$

$\Delta Inf_{2004:IV} = 3.5 - 1.6 = 1.9$

The forecast of $\Delta Inf_{2005:I}$ is:

$$\overline{\Delta Inf}_{2005:I|2000:IV} = 0.017 - 0.238 \times 1.9 = -0.44 \approx -0.4$$

so

$$\overline{Inf}_{2005:I|2000:IV} = Inf_{2004:IV} + \overline{\Delta Inf}_{2005:I|2000:IV} = 3.5 - 0.4 = 3.1\%$$

The AR(p) model: using multiple lags for forecasting

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

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

- The AR(p) model uses p lags of Y as regressors
- The AR(1) model is a special case
- The coefficients do not have a causal interpretation
- To test the hypothesis that Y_{t-2}, \dots, 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” (*more on this later...*)

Example: AR(4) model of inflation

$$\overline{\Delta Inf}_t = .02 - .26\Delta Inf_{t-1} - .32\Delta Inf_{t-2} + .16\Delta Inf_{t-3} - .03\Delta Inf_{t-4},$$

(.12) (.09) (.08) (.08) (.09)

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

- F -statistic testing lags 2, 3, 4 is 6.91 (p -value $< .001$)
- \overline{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 – both in a statistical sense (are statistically significant) and in a substantive sense (substantial increase in the \overline{R}^2)

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
```

dinf	Coef.	Robust 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,... 4th lags of `dinf`

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

```
. dis "Adjusted Rsquared = " _result(8); result(8) is the rbar-squared  
Adjusted Rsquared = .18474733 of the most recently run regression
```



```
. test L2.dinf L3.dinf L4.dinf; L2.dinf is the second lag of dinf, etc
```



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



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


Digression: we used ΔInf , not Inf , in the AR's. Why?

The AR(1) model of 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

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

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

So why use ΔInf_t , not Inf_t ?

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

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

- 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, our regression theory are working with here breaks down
- Here, Inf_t is strongly serially correlated – so to keep ourselves in a framework we understand, the regressions are specified using ΔInf
- More on this later...

Time Series Regression with Additional Predictors and the Autoregressive Distributed Lag (ADL) Model (SW Section 14.4)

- So far we have considered forecasting models that use only past values of Y
- It makes sense to add other variables (X) that might be useful predictors of Y , above and beyond the predictive value of lagged values of Y :

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

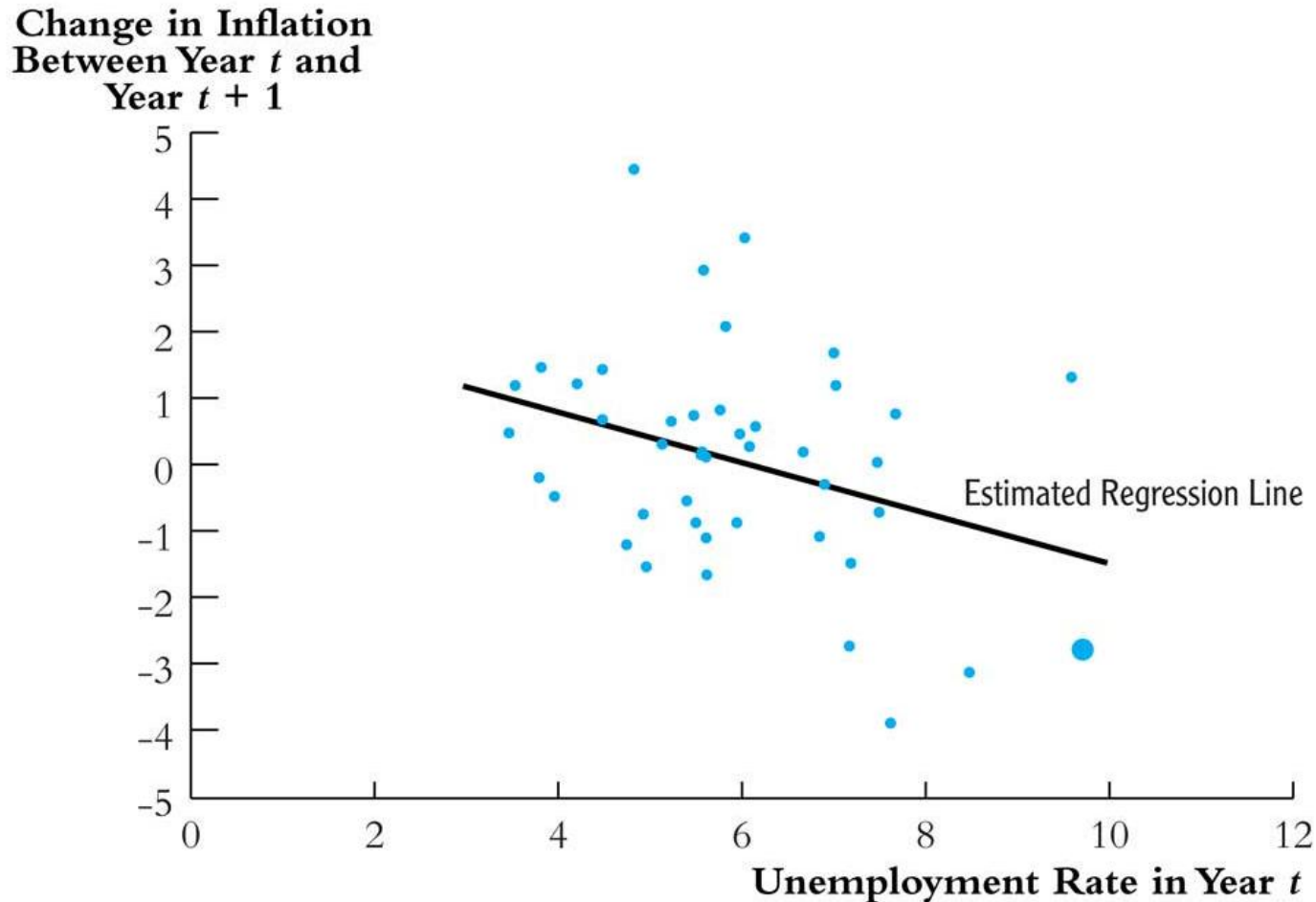
- This is an *autoregressive distributed lag model* with p lags of Y and r lags of X ... ***ADL(p,r)***.

Example: inflation and unemployment

According to the “Phillips curve,” if unemployment is above its equilibrium, or “natural,” rate, then the rate of inflation will increase. That is, ΔInf_t is related to lagged values of the unemployment rate, with a negative coefficient

- The rate of unemployment at which inflation neither increases nor decreases is often called the “non-accelerating rate of inflation” unemployment rate (the NAIRU).
- Is the 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 that it is the value of u for which $\Delta Inf = 0$ – the x intercept of the regression line.

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

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

$$\overline{\Delta Inf}_t = 1.30 - .42\Delta Inf_{t-1} - .37\Delta Inf_{t-2} + .06\Delta Inf_{t-3} - .04\Delta Inf_{t-4}$$

(.44) (.08) (.09) (.08) (.08)

$$- 2.64Unem_{t-1} + 3.04Unem_{t-2} - 0.38Unem_{t-3} + .25Unemp_{t-4}$$

(.46) (.86) (.89) (.45)

- $\bar{R}^2 = 0.34$ – 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
```

dinf	Coef.	Robust 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 The lags of unem are significant  
Prob > F = 0.0000
```

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 $X_{1t-1}, X_{1t-2}, \dots, X_{1t-q_1}$) 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

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

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\beta}_2 X_T$$

The forecast error is:

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_2 - \beta_2)X_T]$$

The *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 = $\text{var}(u_{T+1})$ + uncertainty arising because of estimation error
- If the sample size is large, the part from the estimation error is (much) smaller than $\text{var}(u_{T+1})$, 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 $\text{RMSFE} \approx \sigma_u$, so estimate the RMSFE by the SER.
2. Use an actual forecast history for $t = t_1, \dots, T$, then estimate by

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

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, $t = t_1 - 1, \dots, T - 1$
- 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,

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

Why the term “pseudo out-of-sample forecasts”?

Using the RMSFE to construct forecast intervals

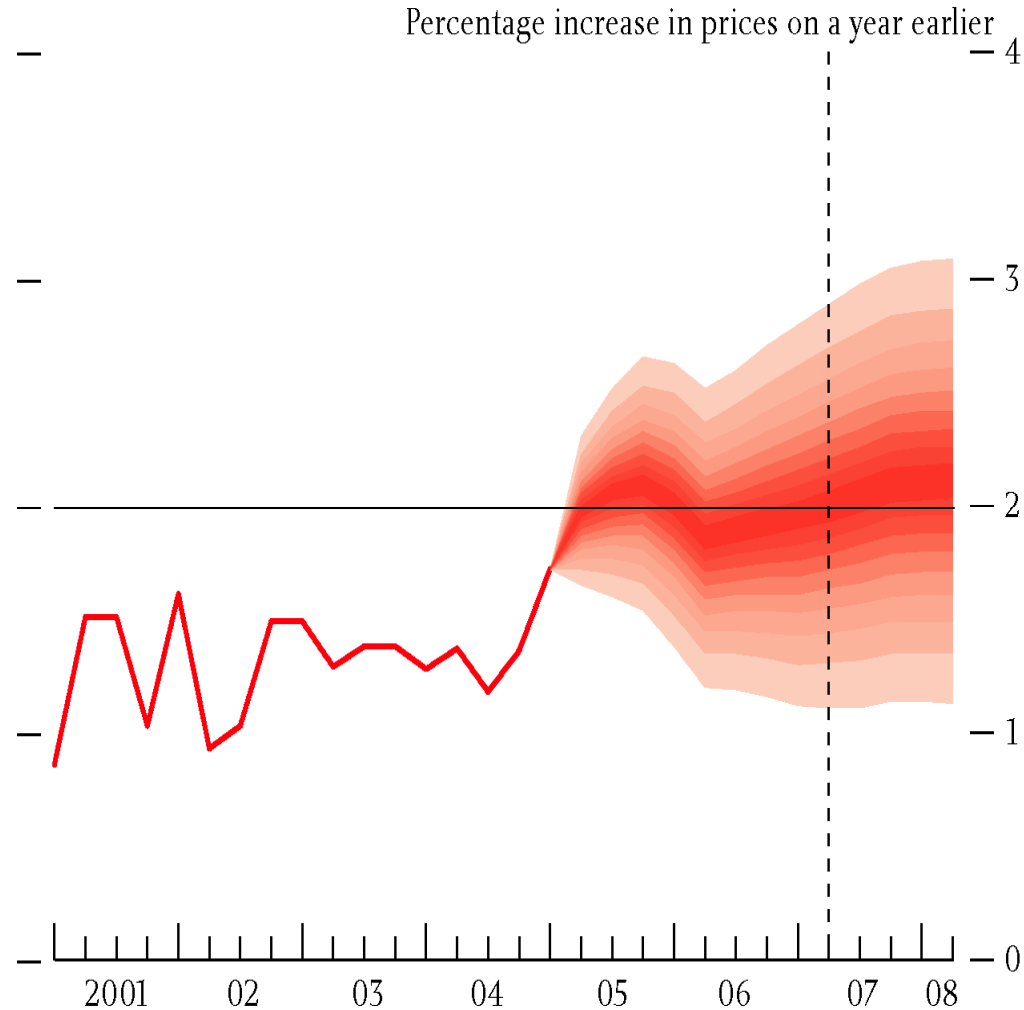
If u_{T+1} is normally distributed, then a 95% forecast interval can be constructed as

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

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: $\pm \sqrt{RMSFE}$

Example #1: the Bank of England “Fan Chart”, 11/05



<http://www.bankofengland.co.uk/publications/inflationreport/index.htm>

Example #2: *Monthly Bulletin of the European Central Bank, Dec. 2005, Staff macroeconomic projections*

Bank, Dec. 2005, Staff macroeconomic projections

Table 10 Macroeconomic projections

(average annual percentage changes)¹⁾

	2004	2005	2006	2007
HICP	2.1	2.1-2.3	1.6-2.6	1.4-2.6
Real GDP	1.8	1.2-1.6	1.4-2.4	1.4-2.4
Private consumption	1.4	1.1-1.5	0.9-1.9	0.6-2.2
Government consumption	1.1	0.4-1.6	1.1-2.1	0.5-1.5
Gross fixed capital formation	1.6	1.4-2.4	1.6-4.6	1.5-4.7
Exports (goods and services)	6.1	2.6-5.4	4.3-7.5	4.1-7.3
Imports (goods and services)	6.3	3.4-6.0	4.2-7.6	3.5-6.9

1) For each variable and horizon, ranges are based on the average absolute difference between the actual outcomes and past projections by euro area central banks. The projections for real GDP and its components refer to working-day-adjusted data. The projections for exports and imports include intra-euro area trade.

Precisely how, did they compute these intervals?

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

Example #3: Fed, Semiannual Report to Congress, 7/04

Economic projections for 2004 and 2005

Federal Reserve Governors and Reserve Bank presidents		
Indicator	Range	Central tendency
		2005
<i>Change, fourth quarter to fourth quarter</i>		
Nominal GDP	4-3/4 to 6-1/2	5-1/4 to 6
Real GDP	3-1/2 to 4	3-1/2 to 4
PCE price index excl food and energy	1-1/2 to 2-1/2	1-1/2 to 2
<i>Average level, fourth quarter</i>		
Civilian unemployment rate	5 to 5-1/2	5 to 5-1/4

How did they compute these intervals?

<http://www.federalreserve.gov/boarddocs/hh/>

Lag Length Selection Using Information Criteria (SW Section 14.5)

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)

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

- *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 $\text{BIC}(p)$* trades off bias and variance to determine a “best” value of p for your forecast.
 - The result is that $\hat{p}^{\text{BIC}} \xrightarrow{p} p!$ (SW, App. 14.5)

Another information criterion: *Akaike Information Criterion (AIC)*

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

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

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 using a Granger causality test with a fixed number of lags (number depends on the data and application)

Nonstationarity I: Trends

(SW Section 14.6)

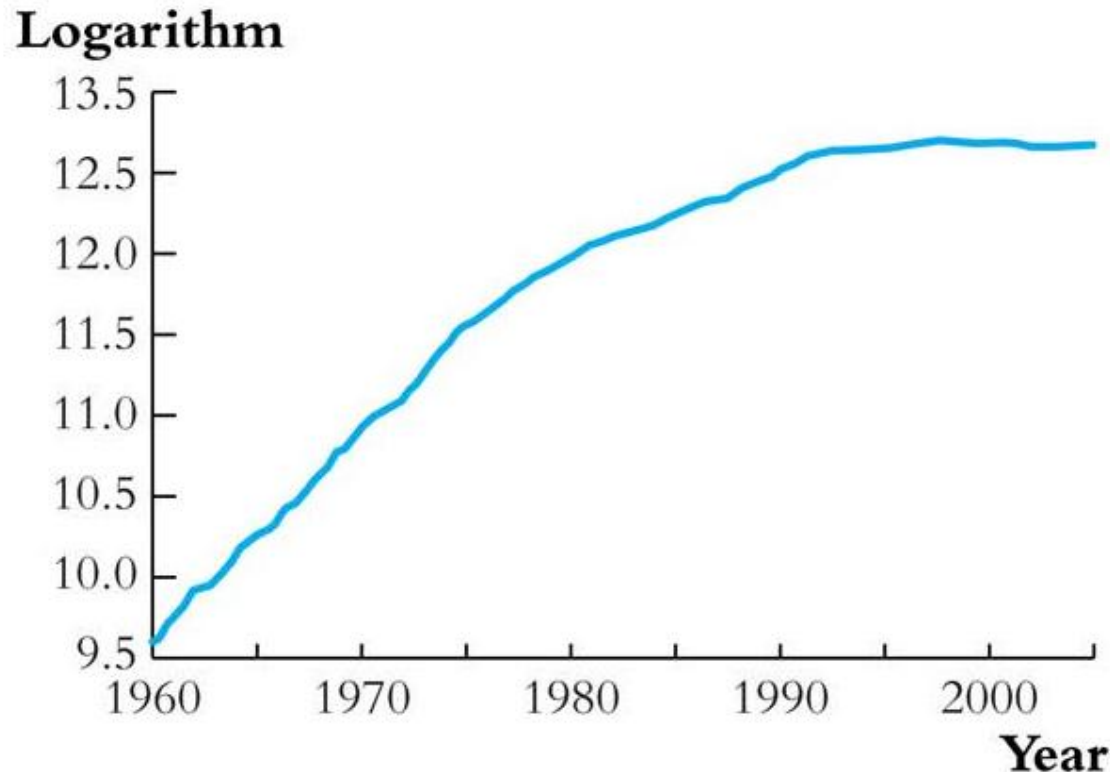
- So far, we have assumed that the data are well-behaved – technically, that the data are stationary.
- Now we will discuss two of the most important ways that, in practice, data can be nonstationary (that is, deviate from stationarity). You need to be able to recognize/detect nonstationarity, and to deal with it when it occurs.
- Two important types of nonstationarity are:
 - Trends (SW Section 14.6)
 - Structural breaks (model instability) (SW Section 14.7)
- Up now: trends

Outline of discussion of trends in time series data:

1. What is a trend?
2. What problems are caused by trends?
3. How do you detect trends (statistical tests)?
4. How to address/mitigate problems raised by trends

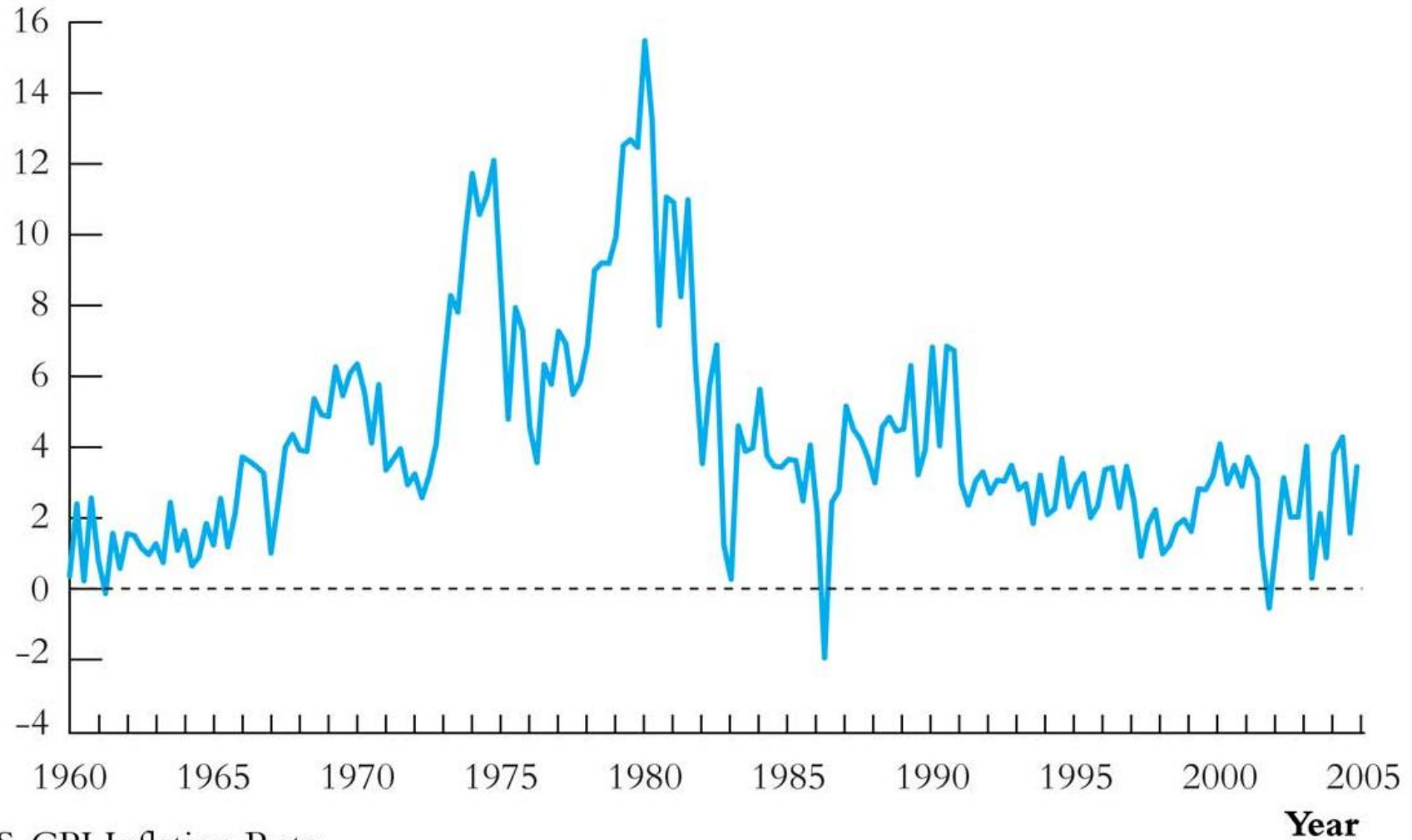
1. What is a trend?

- A trend is a long-term movement or tendency in the data.
- Trends need not be just a straight line!
- Which of these series has a trend?



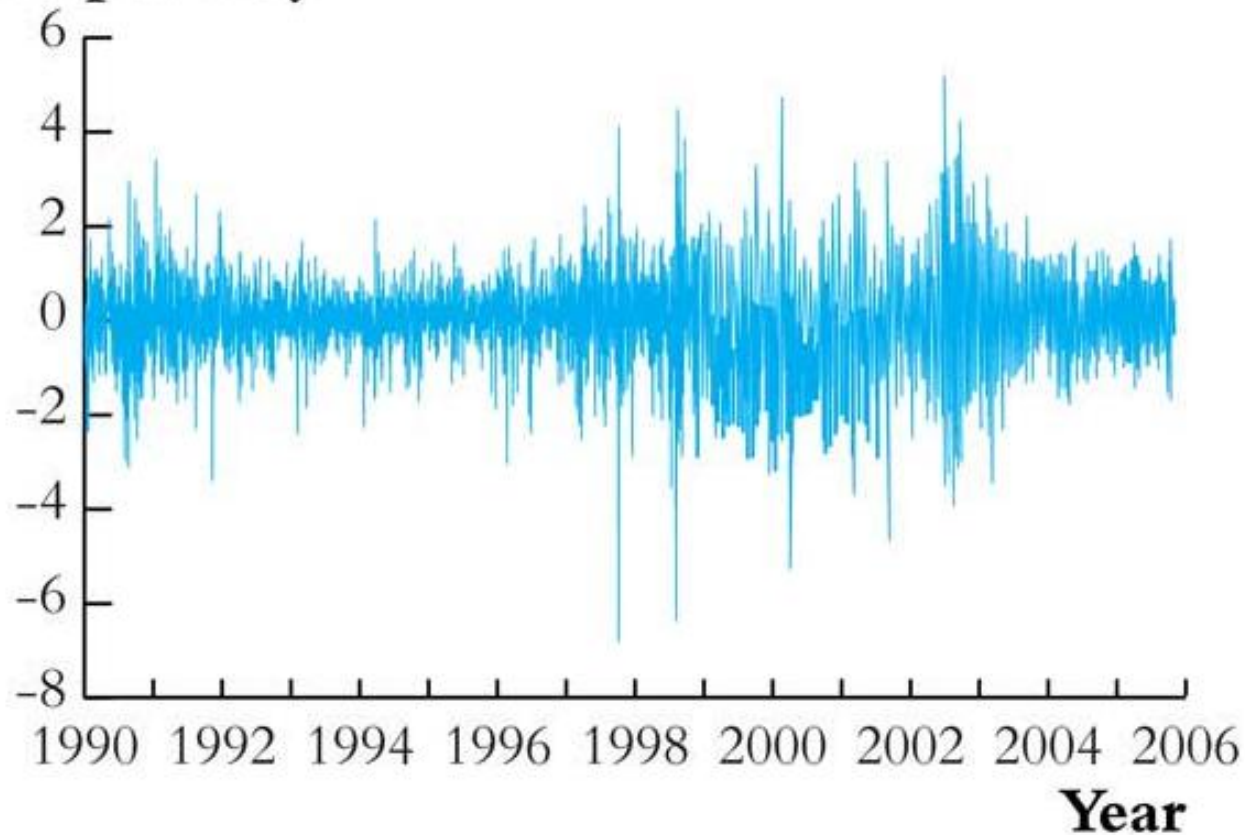
(c) Logarithm of GDP in Japan

Percent



(a) U.S. CPI Inflation Rate

Percent per Day



(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.

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
- 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.

Deterministic and stochastic trends, ctd.

Where we are headed is the following practical advice:

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 .

Upcoming specifics that lead to this advice:

- Relation between the random walk model and AR(1), AR(2), AR(p) (“unit autoregressive root”)
- A regression test for detecting a random walk trend arises naturally from this development

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)

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

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 $\gamma_1 = -\beta_2$.

Claim: 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.

Unit roots in the AR(p) model

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

This regression can be rearranged as,

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

where

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

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

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

...

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

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

The AR(p) model can be written as,

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

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

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:

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

2. What problems are caused by trends?

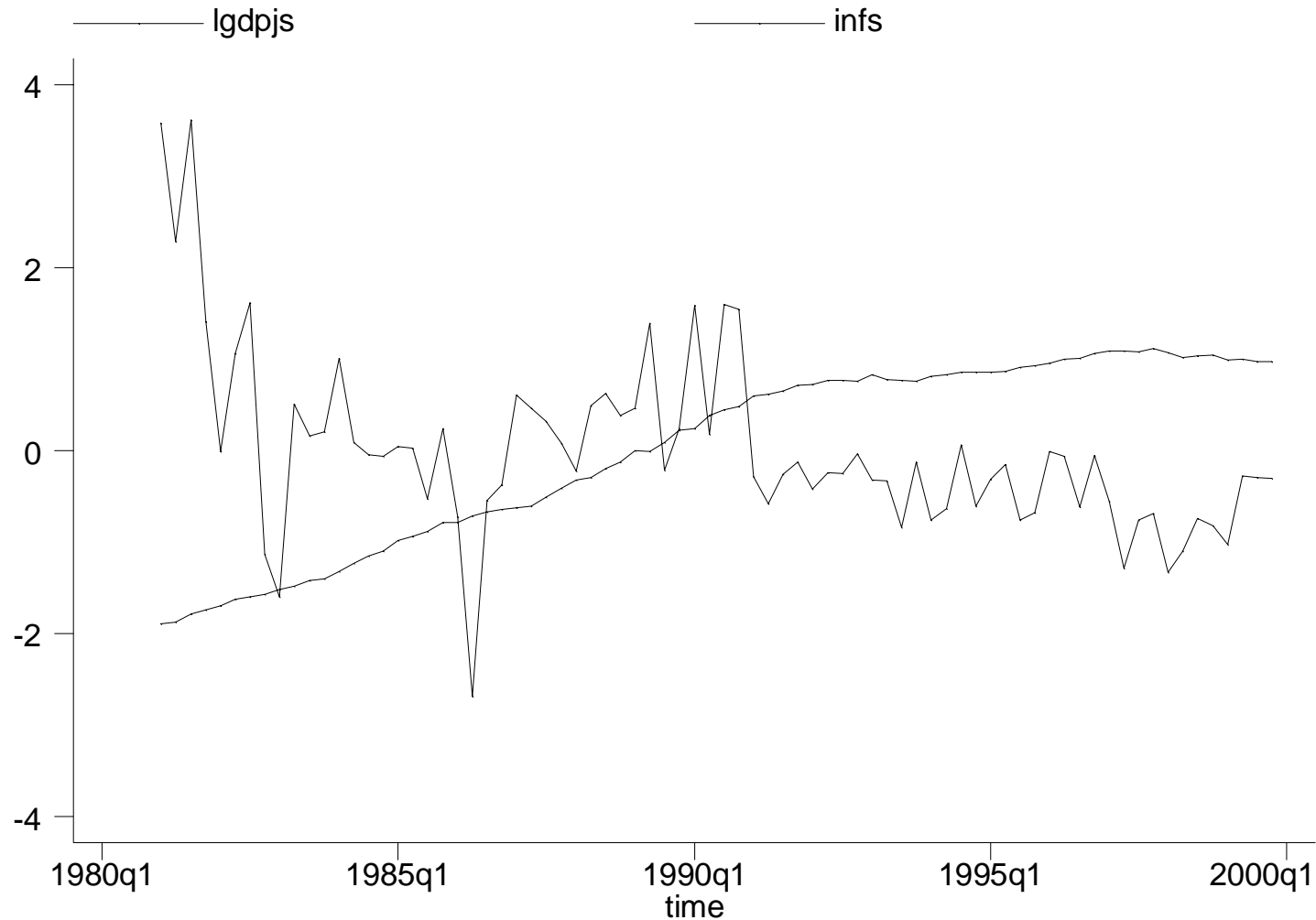
There are three main problems with stochastic trends:

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...

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



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



3. How do you detect trends?

1. Plot the data (think of the three examples we started with).
2. There is 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.

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

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

Test: compute the t -statistic testing $\delta = 0$

- 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:

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

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

- The two cases have different critical values!

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,

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

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

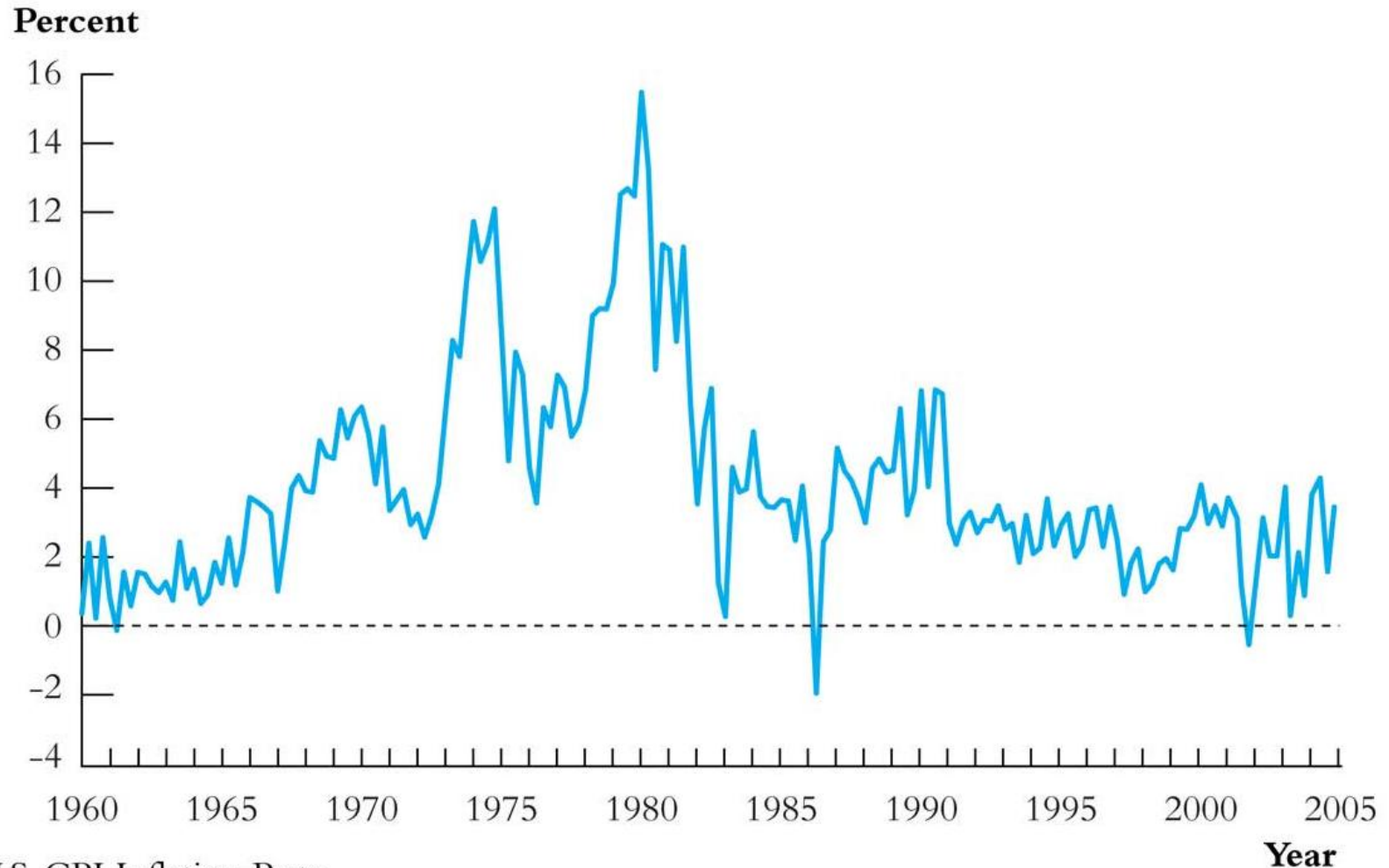
Modification for time trend: include t as a regressor in (*)

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
- 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	
Model	118.197526	5	23.6395052	F(5, 166)	= 10.31
Residual	380.599255	166	2.2927666	Prob > F	= 0.0000
				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

-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.

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

4. 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.

Nonstationarity II: Breaks and Model Stability (SW Section 14.7)

The second type of nonstationarity we consider is that the coefficients of the model might not be constant over the full sample. Clearly, it is a problem for forecasting if the model describing the historical data differs from the current model – you want the current model for your forecasts! (This is an issue of external validity.)

So we will:

- Go over 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 τ . 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 $\gamma_0 = \gamma_1 = \gamma_2 = 0$, then the coefficients are constant over the full sample.
- If at least one of γ_0 , γ_1 , or γ_2 are nonzero, the regression function changes at date τ .

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} \\ + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t$$

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

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

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

vs. H_1 : at least one of γ_0 , γ_1 , or γ_2 are nonzero

- Note that you can apply this to a subset of the coefficients, e.g. only the coefficient on X_{t-1} .
- Often, however, you don't have a candidate break date...

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 for stationarity (coefficient stability) against a general alternative that there has been a break sometime.
- Often, even if you think you know the break date, that “knowledge” is based on prior inspection of the series – so that in effect you “estimated” the break date. This invalidates the Chow test critical values (*why?*)

The Quandt Likelihood Ratio (QLR) Statistic (also called the “sup-Wald” statistic)

The QLR statistic = the maximal Chow statistics

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

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

- A conventional choice for τ_0 and τ_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.

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

- The large-sample null distribution of $F(\tau)$ for a given (fixed, not estimated) τ is $F_{q,\infty}$
- But if you get to compute two Chow tests and choose the biggest one, the critical value must be larger than the critical value for a single Chow test.
- If you compute very many Chow test statistics – for example, all dates in the central 70% of the sample – the critical value must be larger still!

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

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

where $\{B_i\}$, $i = 1, \dots, n$, are independent continuous-time “Brownian Bridges” on $0 \leq s \leq 1$ (a Brownian Bridge is a Brownian motion deviated from its mean), 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

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.

Has the postwar U.S. Phillips Curve been stable?

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

$$\overline{\Delta Inf}_t = 1.30 - .42\Delta Inf_{t-1} - .37\Delta Inf_{t-2} + .06\Delta Inf_{t-3} - .04\Delta Inf_{t-4}$$

(.44) (.08) (.09) (.08) (.08)

$$- 2.64Unem_{t-1} + 3.04Unem_{t-2} - 0.38Unem_{t-3} + .25Unemp_{t-4}$$

(.46) (.86) (.89) (.45)

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

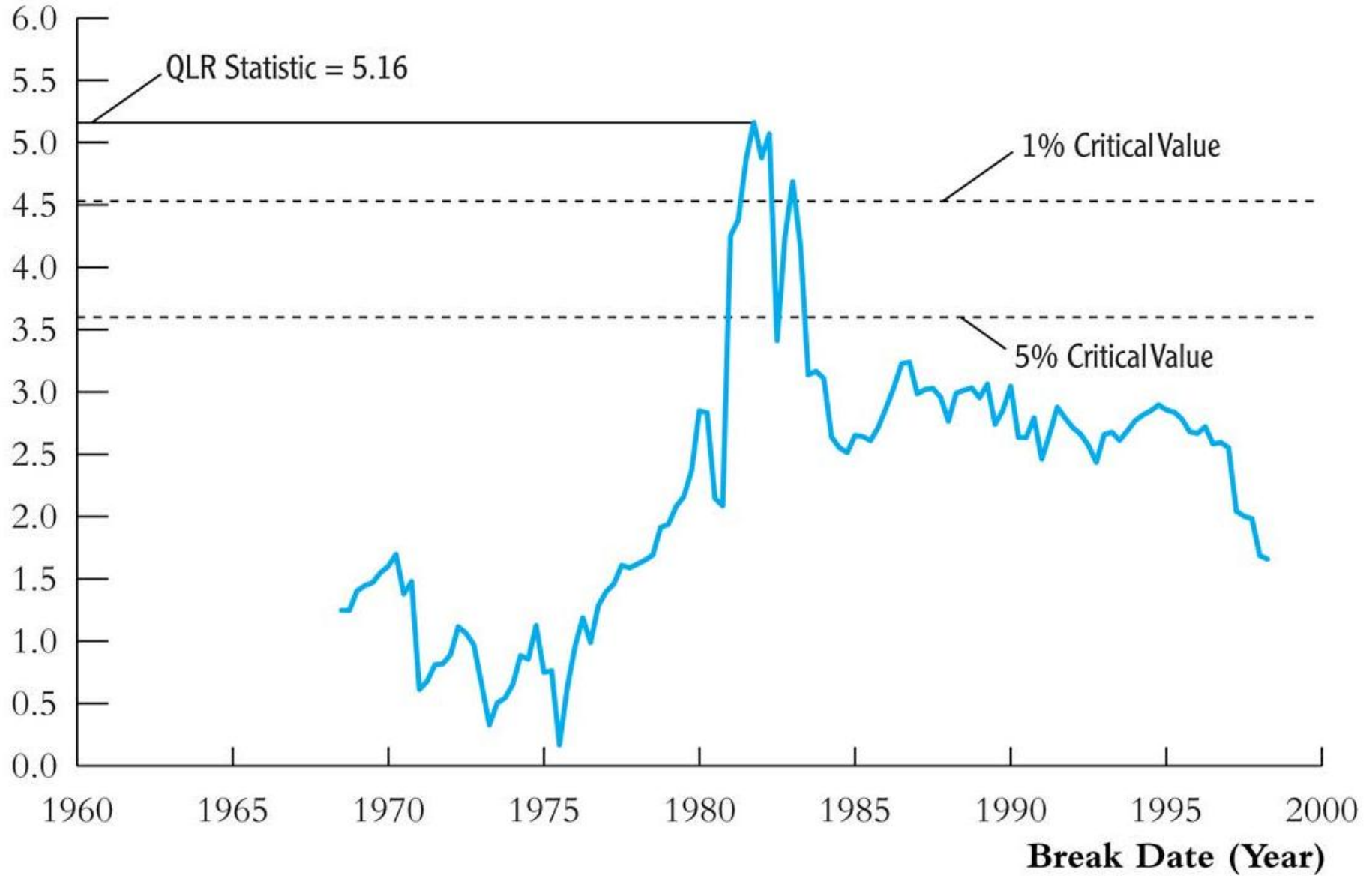
QLR tests of the stability of the U.S. Phillips curve.

dependent variable: ΔInf_t

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

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

F-Statistic



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

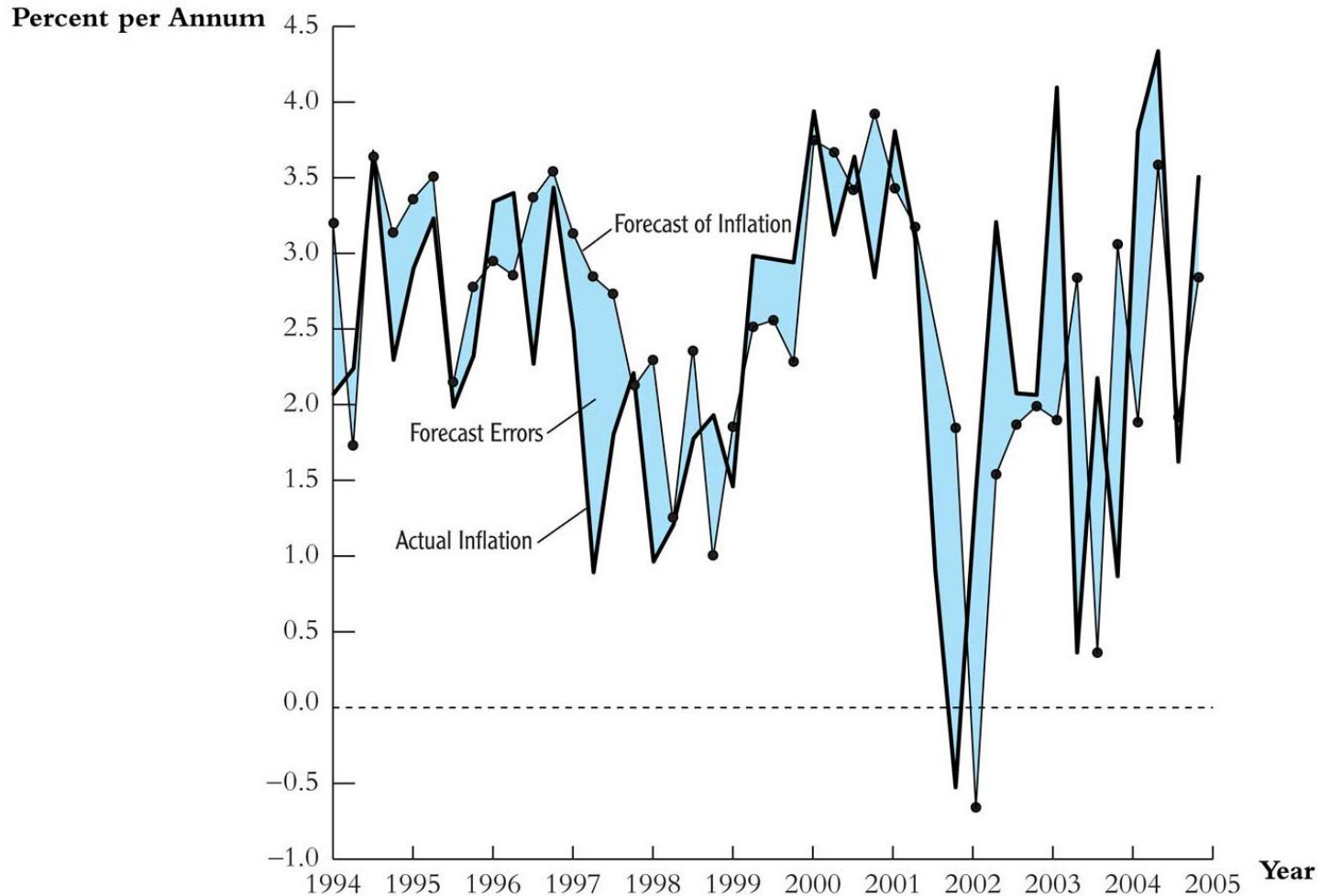
- The QLR test does not work well towards the very end of the sample – but this is usually the most interesting part – it is the most recent history and you want to know if the forecasting model still works in the very recent past.
- One way to check whether the model is working is to see whether the pseudo out-of-sample (*poos*) forecasts are “on track” in the most recent observations. Because the focus is on only the most recent observations, this is an informal diagnostic (not a formal test) – this complements formal testing using the QLR.

Application to the U.S. Phillips Curve:

Was the U.S. Phillips Curve stable towards 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”).
- Regression model:
dependent variable: ΔInf_t
regressors: $1, \Delta Inf_{t-1}, \dots, \Delta Inf_{t-4}, Unemp_{t-1}, \dots, Unemp_{t-4}$
- Pseudo out-of-sample forecasts:
 - Compute regression over $t = 1982:I, \dots, P$
 - Compute *poos* forecast, $\overline{\Delta Inf}_{P+1|P}$, and forecast error
 - Repeat for $P = 1994:I, \dots, 2005:I$

POOS forecasts of ΔInf 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 too high or too low
- *poos RMSFE*, 1999:I – 2004:IV: 1.32
- *SER*, model fit 1982:I – 1998:IV: 1.30
 - The *poos RMSFE* \approx 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

Summary: 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 .

Summary, 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