Time Series Regression

Outline

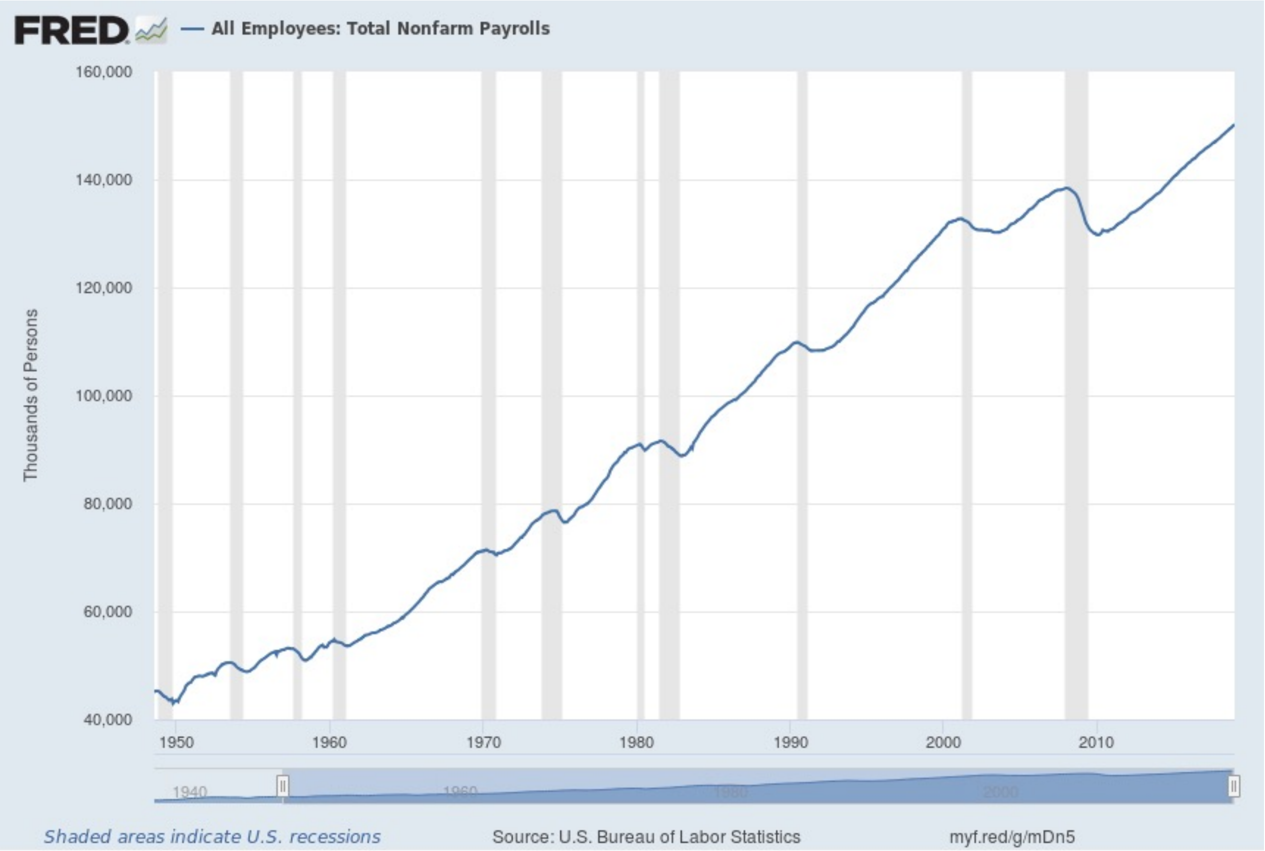
1. Time Series Data: What' s Different?
2. Lags, Differences, Autocorrelation
3. Forecasting, Stationarity, and the Mean Squared Forecast Error
4. Autoregressions
5. The Autoregressive – Distributed Lag (ADL) Model
6. Estimation of the MSFE and Forecast Intervals
7. Lag Length Selection: Information 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

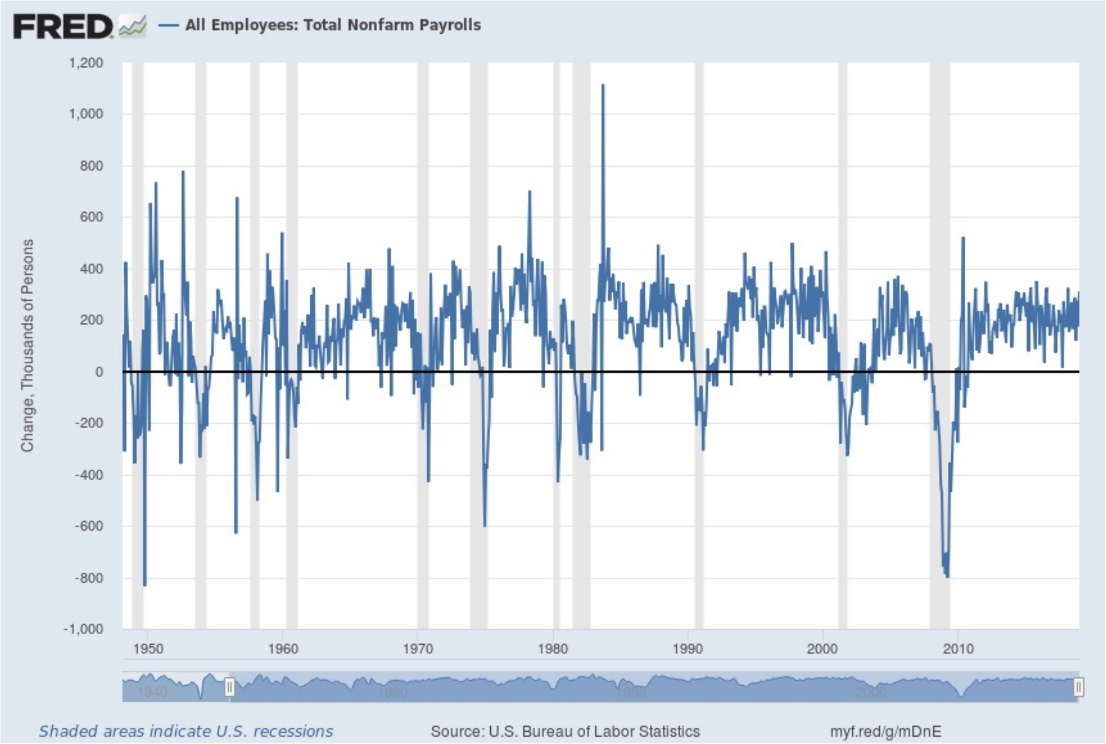
- Aggregate consumption and GDP for a country (for example, 20 years of quarterly observations = 80 observations)
- Yen/$, pound/$ and Euro/$ exchange rates (daily data for 1 year = 365 observations)
- Cigarette consumption per capita in California, by year (annual data)

Some monthly U.S. macro and financial time series (1 of 7)



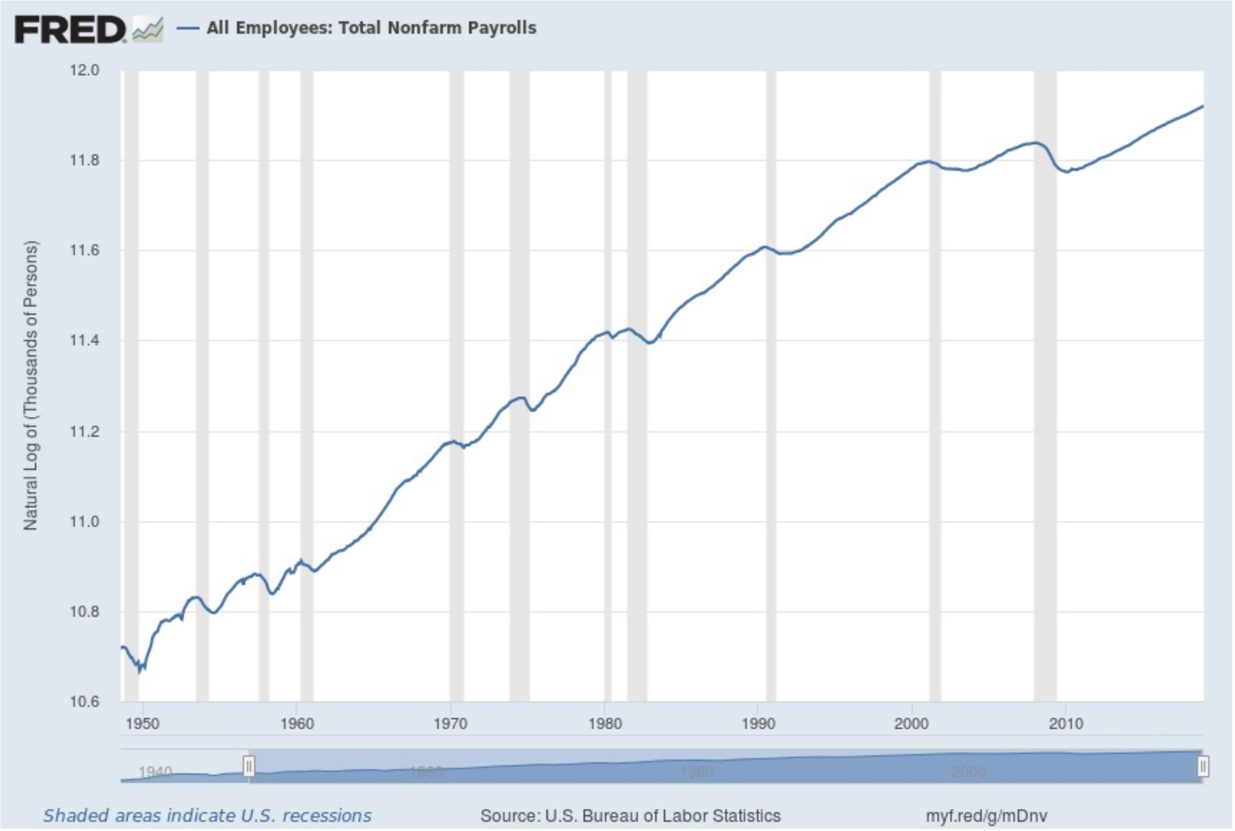
Some monthly U.S. macro and financial time series (2 of 7)

Monthly
Change



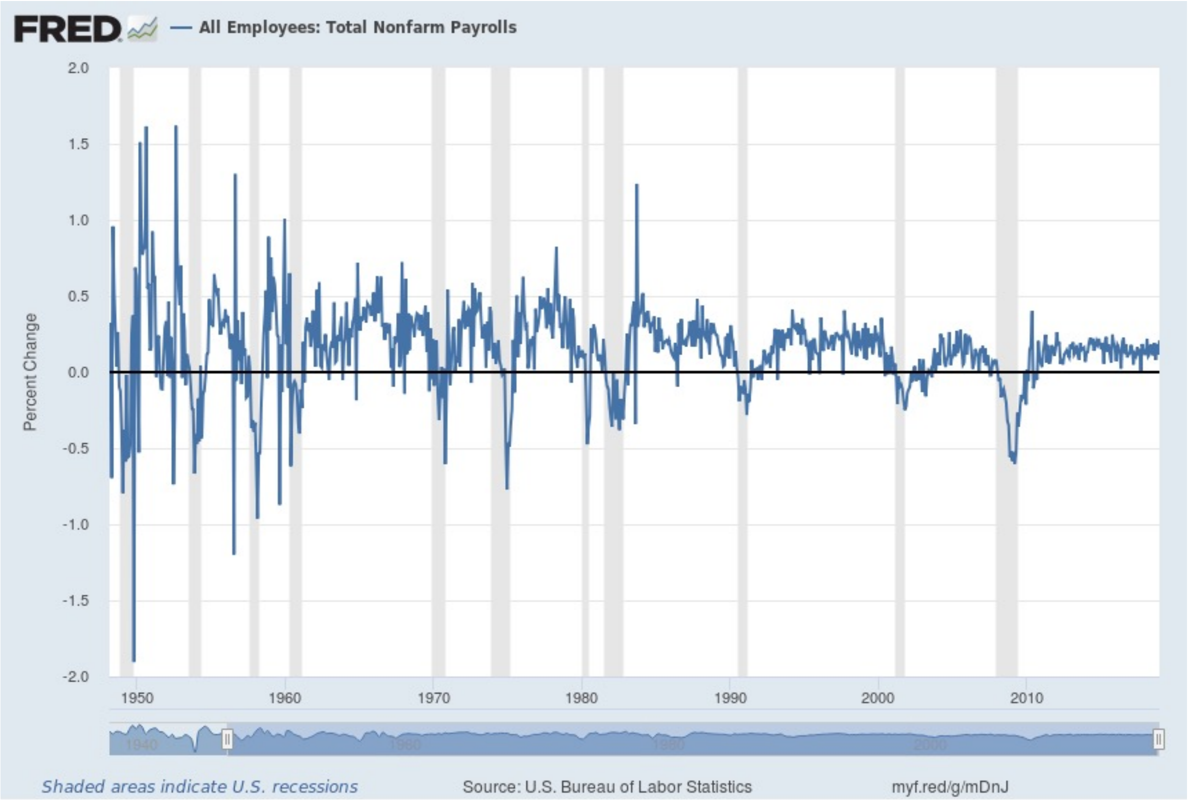
Some monthly U.S. macro and financial time series (3 of 7)

Logarithm

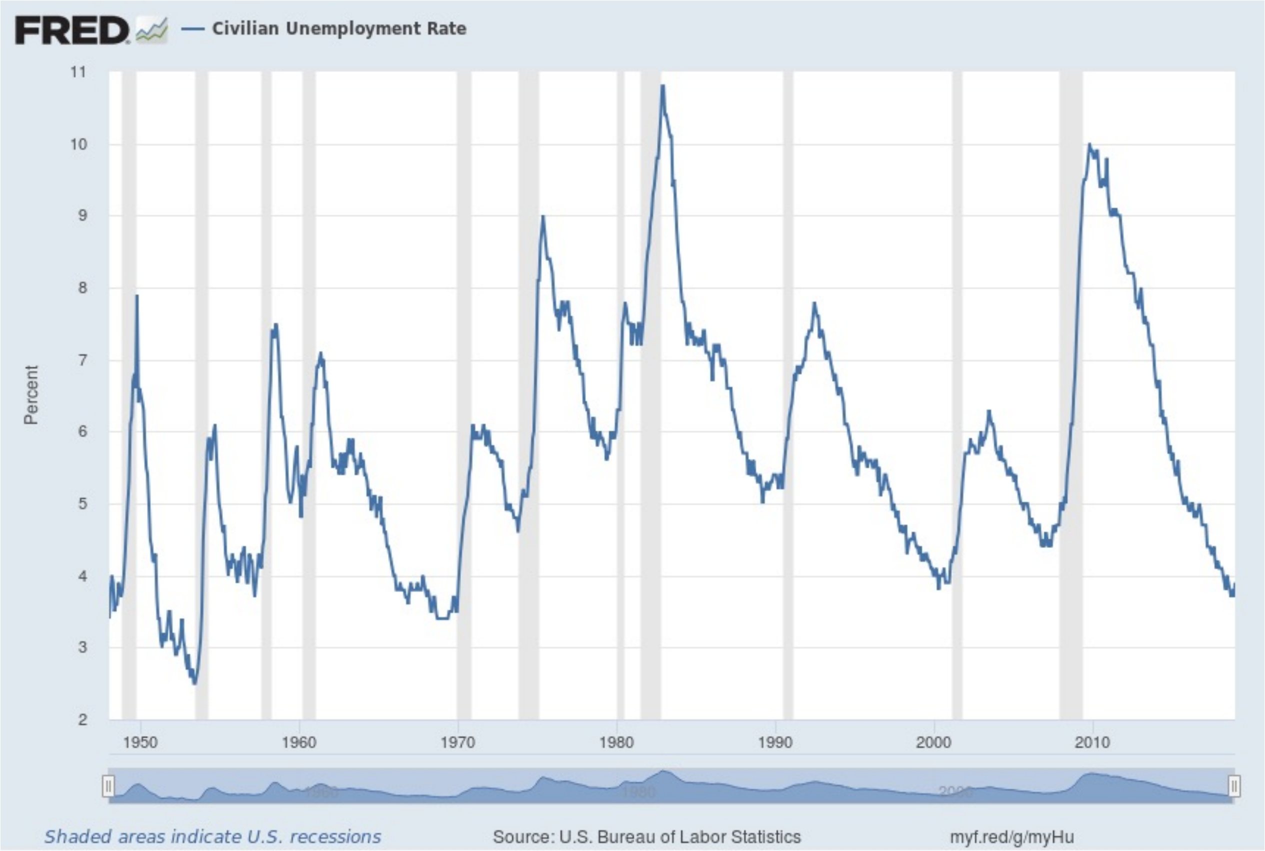


Some monthly U.S. macro and financial time series (4 of 7)

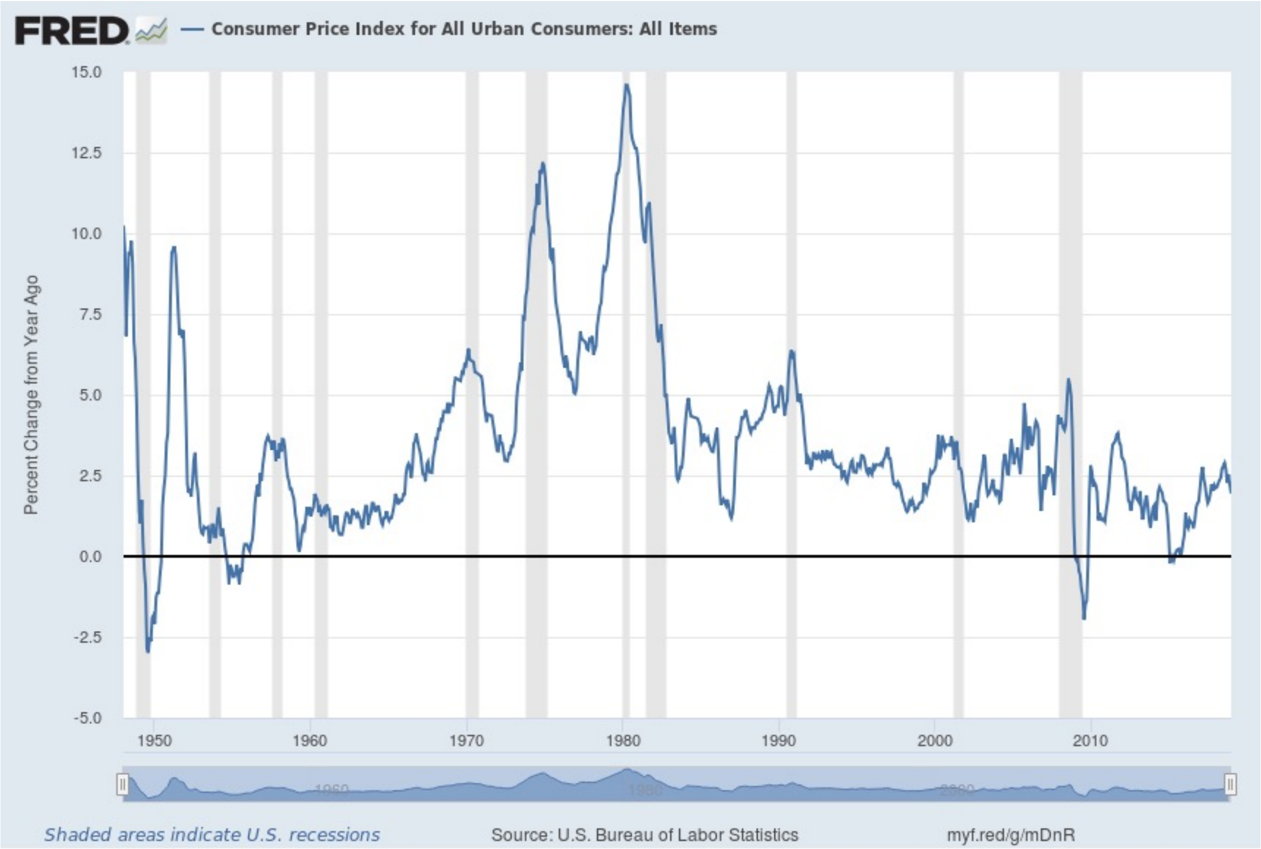
Monthly
Percent Change



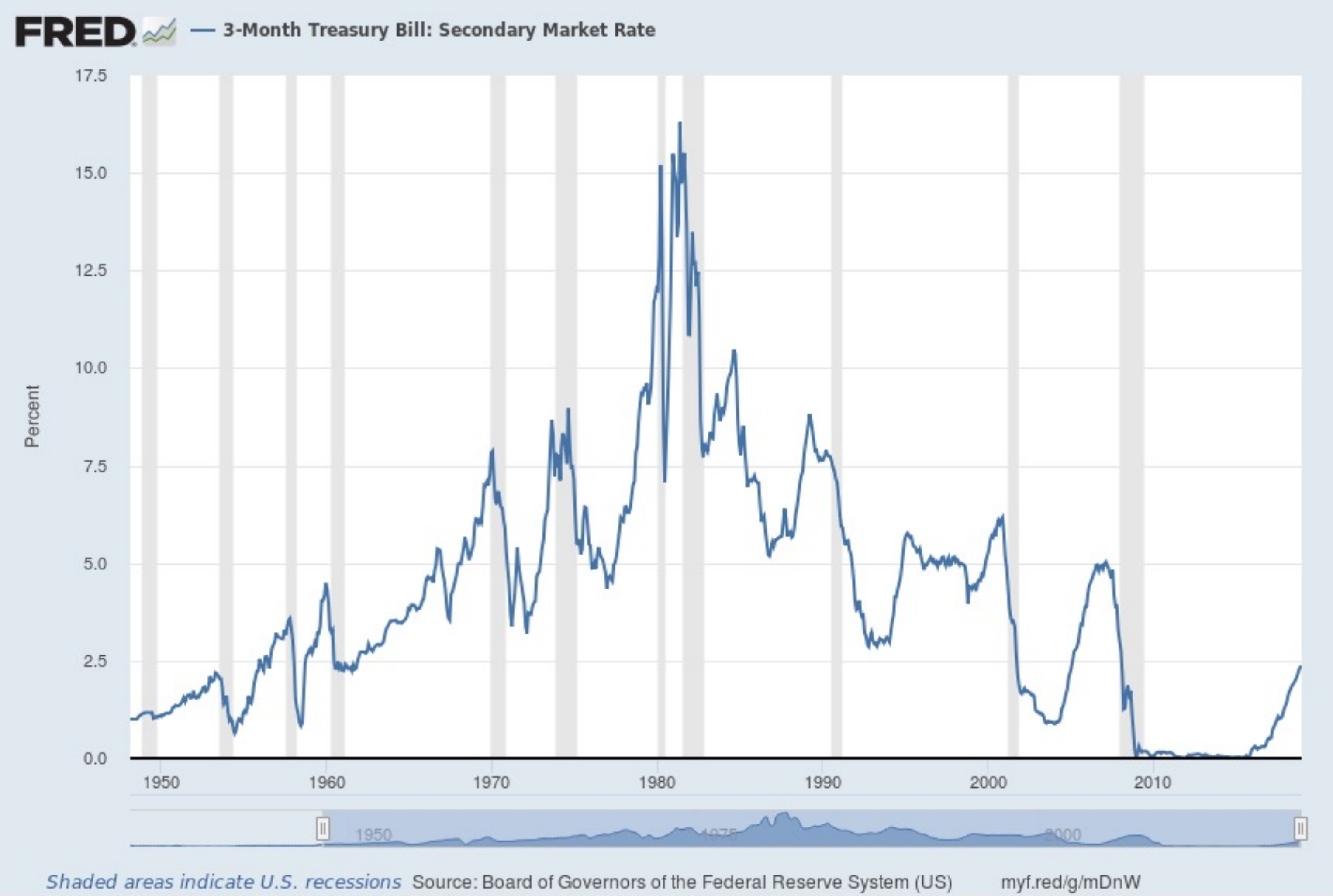
Some monthly U.S. macro and financial time series (5 of 7)



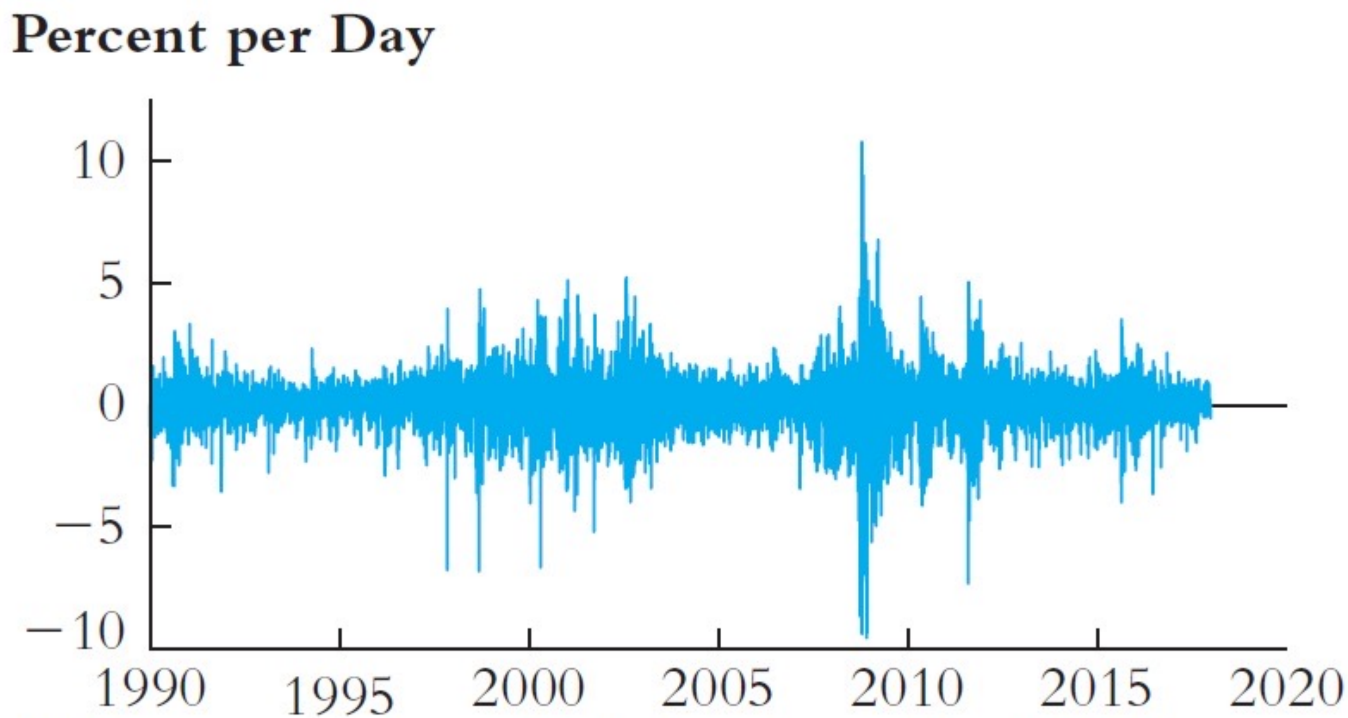
Some monthly U.S. macro and financial time series (6 of 7)



Some monthly U.S. macro and financial time series (7 of 7)



A daily U.S. financial time series:



(d) Percentage change in daily value of the Wilshire 5000 Total Market Index

Some uses of time series data

- Forecasting (SW Ch. 15)
- Estimation of dynamic causal effects (SW Ch. 16)
  - 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?
- Modeling risks, which is used in financial markets (one aspect of this, modeling changing variances and "volatility clustering," is discussed in SW Ch. 17)
- Applications outside of economics include environmental and climate modeling, engineering (system dynamics), computer science (network dynamics),…

Time series data raises new technical issues

- Time lags
- Correlation over time (*serial correlation*, a.k.a. *autocorrelation* – which we encountered in panel data)
- Calculation of standard errors when the errors are serially correlated

A good way to learn about time series data is to investigate it yourself! 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.

2. Introduction to Time Series Data and Serial Correlation (SW Section 15.1)

Time series basics:

A. Notation
B. Lags, first differences, and growth rates, and first log diference approximation to growth rates
C. Autocorrelation (serial correlation)

A. Notation

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

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^{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 growth of U.S. GDP at an annual rate

GDP = Real GDP in the US (Billions of $2009)

- GDP in the fourth quarter of 2016 (2016:Q4) = 16851
- GDP in the first quarter of 2017 (2017:Q1) = 16903
- Percentage change in GDP, 2016:Q4 to 2017:Q1

$$= 100 \times \left(\frac{16903 - 16851}{16851} \right) = 0.31\%$$

- Percentage change in GDP, 2012:Q1 to 2012:Q2, *at an annual rate* = 4 × 0.31% = **1.23**% ≈ **1.2**% (percent per year)
- Using the logarithmic approximation to percent changes yields 4 × 100 × [log(16903) – log(16851)] = **1.232**%

Example: GDP and its rate of change

TABLE 15.1 GDP in the United States in the Last Quarter of 2016 and in 2017

Quarter	U.S. GDP (billions of $2009), GDP_t	Logarithm of GDP, $\ln(GDP_t)$	Growth Rate of GDP at an Annual Rate, $GDPGR_t = 400 \times \Delta\ln(GDP_t)$	First Lag, $GDPGR_{t-1}$
2016:Q4	16,851	9.732	1.74	2.74
2017:Q1	16,903	9.735	1.23	1.74
2017:Q2	17,031	9.743	3.01	1.23
2017:Q3	17,164	9.751	3.11	3.01
2017:Q4	17,272	9.757	2.50	3.11

Note: The quarterly rate of GDP growth is the first difference of the logarithm. This is converted into percentages at an annual rate by multiplying by 400. The first lag is its value in the previous quarter. 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 cov(Y_t, Y_{t-1})
- The first ***autocorrelation*** of Y_t is corr(Y_t, Y_{t-1})
- Thus

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

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

Pearson

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,

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

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

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

Sample autocorrelations (1 of 2)

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

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

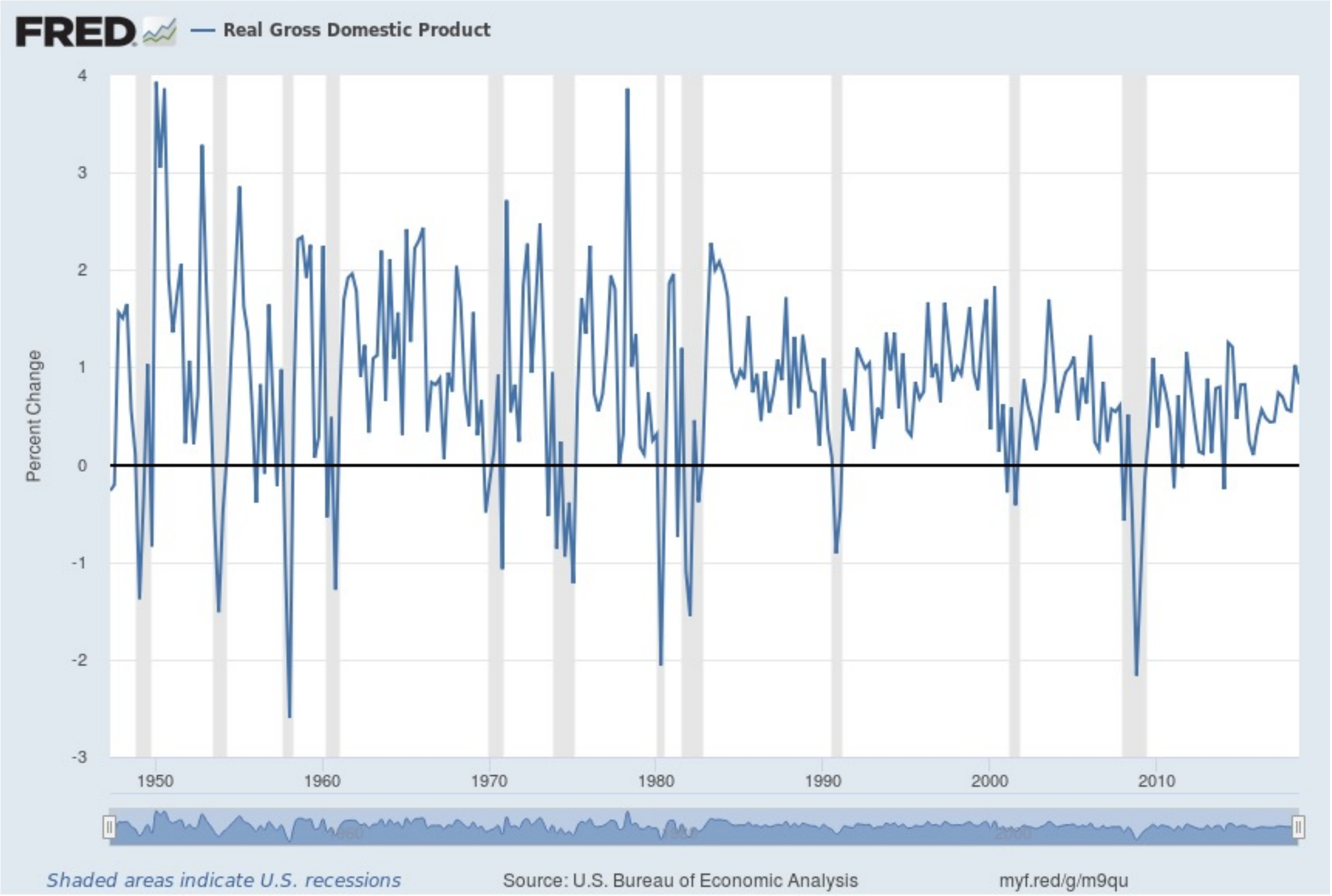
where $\text{cov}(Y_t, Y_{t-j}) = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \bar{Y}_{j+1,T})(Y_{t-j} - \bar{Y}_{1,T-j})$

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

Sample autocorrelations (2 of 2)

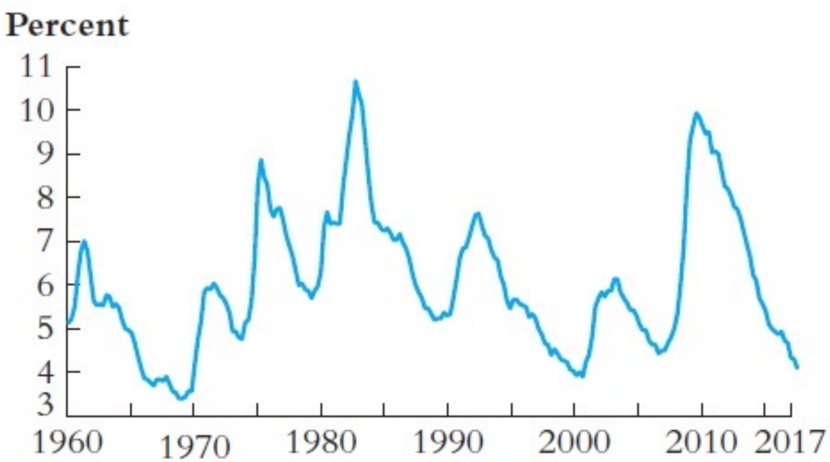
Quarterly percent change



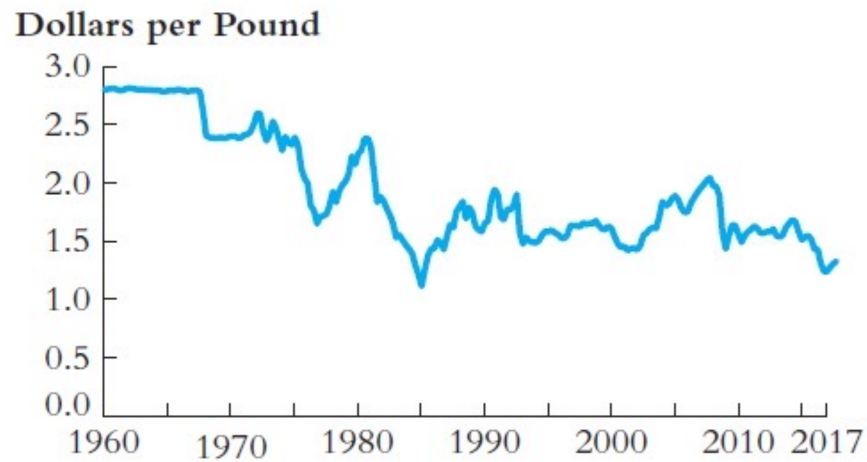
The first four autocorrelations are:

$$\hat{\rho}_1 = 0.33,\ \hat{\rho}_2 = 0.26,\ \hat{\rho}_3 = 0.10,\ \text{and}\ \hat{\rho}_4 = 0.11$$

Other economic time series: do these series look serially correlated (is Y_t strongly correlated with Y_{t+1}?) (1 of 2)



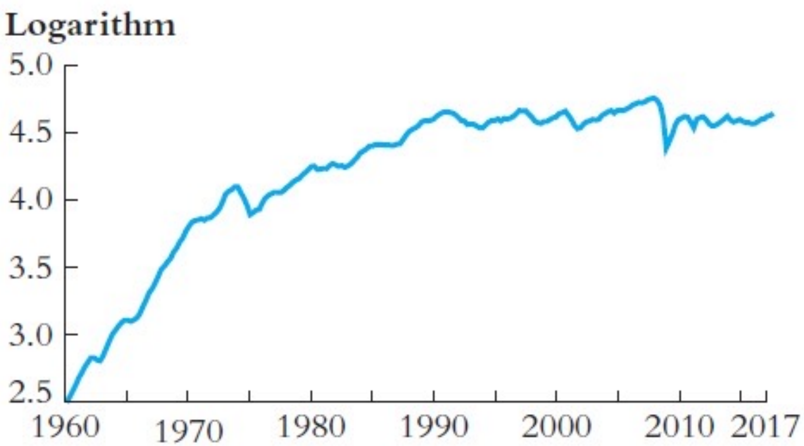
(a) U.S. unemployment rate



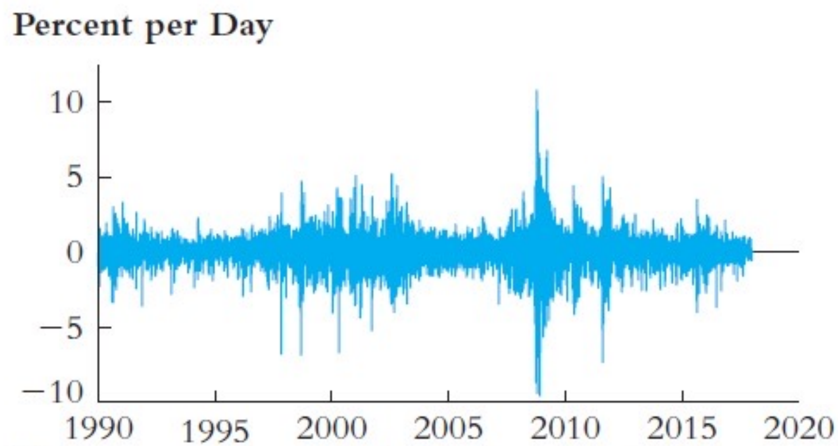
(b) U.S. dollar/British pound exchange rate

Pearson

Other economic time series: do these series look serially correlated (is Y_t strongly correlated with Y_{t+1}?) (2 of 2)



(c) Logarithm of Japan Index of Industrial Production



(d) Percentage change in daily value of the Wilshire 5000 Total Market Index

3. Forecasting, Stationarity, and the Mean Squared Forecast Error (SW Section 15.2)

- Forecasting and estimation of causal effects are quite different objectives.
- For estimation of causal effects, we were very concerned about omitted variable bias, control variables, and exogeneity.
- For forecasting,
  - Omitted variable bias isn't a problem!
  - We won't worry about interpreting coefficients in forecasting models – no need to estimate causal effects if all you want to do is forecast!
  - What is paramount is, instead, that the model provide an out-of-sample prediction that is as accurate as possible.
    - For the in-sample model to be useful out-of-sample, the out-of-sample period (near future) must be like the in-sample data (the historical data) – a condition called *stationarity*…

Stationarity

A time series Y_t is *stationary* if its probability distribution does not change over time, that is, if the joint distribution of $(Y_{s+1}, Y_{s+2}, \cdots, Y_{s+T})$ 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 $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \cdots, X_{s+T}, Y_{s+T})$ does not depend on s. Stationarity requires the future to be like the past, at least in a probabilistic sense.

- Stationarity says that history is relevant. Stationarity is a key requirement for external validity of a forecast made using time series data.

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

Forecasts and Forecast Errors

- A forecast is a prediction of the future based on time series data. Because the forecasting model is estimated using historical data available at the time the model is estimated, a forecast is an out-of-sample prediction of the future.
- Forecasts can either be one-step ahead (next month, if you are using monthly data), or multiple steps ahead (4 months from now)
  - Let the estimation sample be $t = 1,\cdots, T$. The one-step ahead forecast is:

"Hat" means based on an estimated model

$$\hat{Y}_{T+1|T}$$

Forecast of Y at date $T+1$

Forecast is based on data through date T

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

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},$\cdots$, 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, which are estimated using data through period T.
- A multi-step ahead forecast is for a date more than one period in the future. Specifically, an h-period ahead forecast, made using data through date T, is denoted,

$$h\text{-period ahead forecast} = \hat{Y}_{T+h|T}$$

Mean Squared Forecast Error

- The Mean Squared Forecast Error (MSFE) is a measure of the quality of a forecast.
- Specifically, the MSFE is the expected value of the square of the forecast error, when the forecast is made for an observation not used to estimate the forecasting model (i.e., for an observation in the future).
  - Using the square of the error penalizes large forecast error much more heavily than small ones.
  - A small error might not matter, but a very large error could call into question the entire forecasting exercise.

$$MSFE = E\left[\left(Y_{T+1} - \hat{Y}_{T+1|T}\right)^2\right]$$

4. Autoregressions (SW Section 15.3)

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

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 an OLS regression of Y_t against Y_{t-1} *(mechanically, how would you run this regression??)*
- 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 for the growth rate of GDP

AR(1) estimated using data from 1962:Q1 – 2017:Q3:

$$\underset{}{GDPGR_t} = \underset{(0.322)}{1.950} + \underset{(0.073)}{0.341}GDPGR_{t-1}, \qquad \bar{R}^2 = 0.11$$

where *GDPGR* is the percentage growth of GDP at an annual rate using the log approximation = $400\ln(GDP_t/GDP_{t-1})$

- Is the lagged growth rate of GDP a useful predictor of the current growth rate of GDP?
  - t = 0.341/.073 = 4.67 > 1.96 (in absolute value)

- Yes, the lagged growth rate of GDP is a useful of the current growth rate-but the $\bar{R}^2$ is pretty low.

Example: forecasting GDP growth using an AR(1)

AR(1) estimated using data from 1962:Q1 – 2017:Q3:

$$GDPGR_t = 1.950 + 0.341 GDPGR_{t-1}$$

$GDPGR_{2017:Q3}$ = 3.11 (units are percent, at an annual rate)

The forecast of $GDPGR_{2017:Q4}$ is:

$$GDPGR_{2017:Q4|2017:Q3} = 1.950 + 0.341 \times 3.11 = 3.0\%$$

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

The pth order autoregressive model (AR(p)) is

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

- The AR(p) model uses p lags of Y as regressors
- The AR(1) model is a special case with $p = 1$
- The coefficients do not have a causal interpretation
- To test the hypothesis that $Y_{t-2},\cdots,Y_{t-p}$ 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(2) model for the growth rate of GDP

AR(2) estimated using data from 1962:Q1 – 2017:Q3:

$$\underset{}{GDPGR_t} = \underset{(0.37)}{1.60} + \underset{(0.08)}{0.28}GDPGR_{t-1} + \underset{(0.08)}{0.18}GDPGR_{t-2}, \qquad \bar{R}^2 = 0.14$$

- To test the hypothesis that Y_{t-2} doe not further help forecast Y_t, beyond Y_{t-1}, use the t-test: 0.18/0.08 = 2.30 > 1.96
- The adjusted R^2 increases slightly relative to the AR(2) and the second lag is statistically significant. Still, most of the variation in next-quarter' s GDP growth remains unforecasted. We would like to do better!!

Pearson

5. Time Series Regression with Additional Predictors and the Autoregressive Distributed Lag (ADL) Model (SW Section 15.4)

- 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:

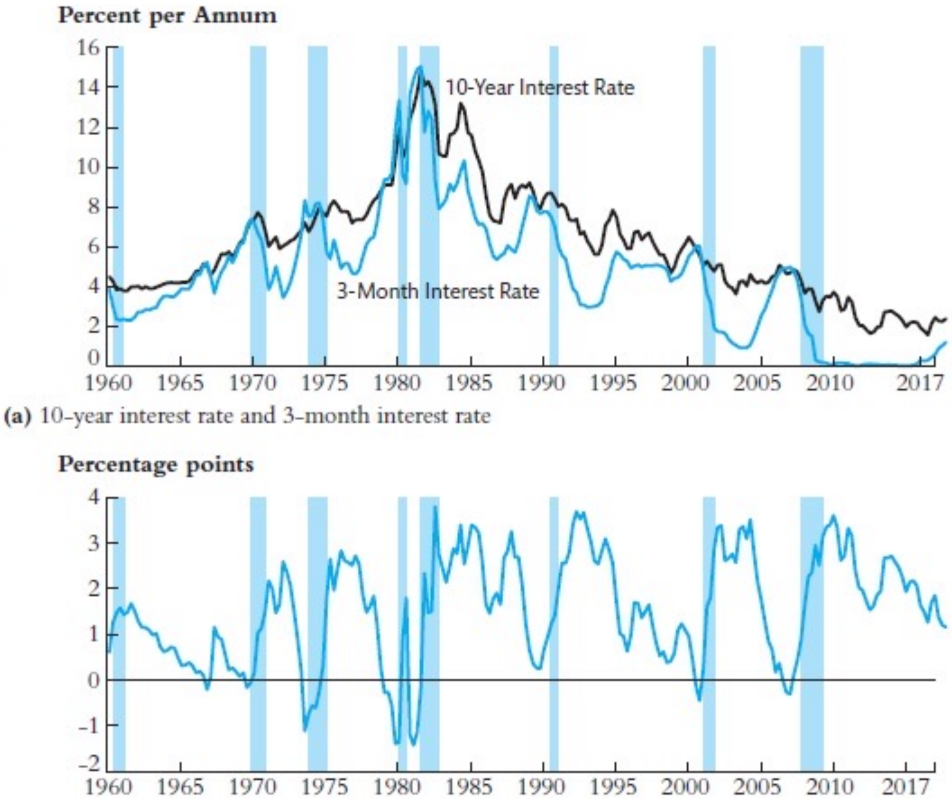
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \cdots + \delta_r X_{t-r} + u_t$$

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

Example: interest rates and the term spread

FIGURE 15.3 Interest Rates and the Term Spread, 1960–2017

Long-term and short-term interest rates on bonds move together but not one-for-one. The difference between long-term rates and short-term rates is called the term spread. The term spread has fallen sharply before U.S. recessions, which are shown as shaded regions in the figures.



(b) Term spread

Example: interest rates and the term spread

ADL(2,2) model (1962:Q1 – 2017:Q3):

$$\begin{aligned} GDPGR_t = {} & \underset{(0.46)}{0.94} + \underset{(0.08)}{0.25}GDPGR_{t-1} + \underset{(0.08)}{0.18}GDPGR_{t-2} \\ & \underset{(0.42)}{-0.13}\,TSpread_{t-1} + \underset{(0.43)}{0.62}\,TSpread_{t-2} \qquad \bar{R}^2 = 0.16 \end{aligned}$$

- The ADL(2,1) model has the first lag of *Tspread* significant, however the second lag in the ADL(2,2) model is not significant (see SW equation (15.14))

6. Estimation of the MSFE and Forecast Intervals (SW Section 15.5)

Why do you need a measure of forecast uncertainty?

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

Consider the forecast

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

The forecast error is:

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

The *mean squared forecast error (MSFE)* is

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

$$= E(u_{T+1})^2 + E[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_1 - \beta_2)X_T]^2$$

- The first term, $E[(u_{T+1})^2]$, is the MSFE of the oracle forecast – that is, the forecast you would make if you knew the true value of the parameters.
  - If the sample size is large and the number of predictors is small, then most of the MSFE arises from this term.
  - Under stationarity, the mean of the forecast error will be zero, so $E[(u_{T+1})^2] = \text{var}(u_{T+1})$.
  - In general, however, the forecast error might have a nonzero mean.
- The second term arises because you need to estimate the β's
  - If the number of predictors is moderate or large relative to the sample size, then this term can be large, in fact it can be larger than the first term.

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 of 4)

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

$$= E(u_{T+1})^2 + E[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_1 - \beta_2)X_T]^2$$

1. Approximate the MSFE by the SER

- If the data are stationary then the forecast error will have mean zero, so $E[(u_{T+1})^2] = \sigma_u^2$
- If the number of predictors is small compared to the sample size, the contribution of estimation error can be ignored.
- Use the approximation MSFE $\approx E[(u_{T+1})^2] = \sigma_u^2$, so estimate the RMSFE by the square of the standard error of the regression (SER). Let SSR denote the (in-sample) sum of squared residuals:

$$MSFE_{SER} = s_{\hat{u}}^2 = \frac{SSR}{T-p-1}$$

Three ways to estimate the RMSFE (2 of 4)

2. Estimate the MSFE using the final prediction error (FPE)

- The Final Prediction Error is an estimate of the MSFE that incorporates both terms in the MSFE formula (squared future *u* and contribution from estimating the coefficients), under the additional assumption that the error are homoskedastic.
- It is shown in SW Appendix 19.7 that this leads to,

$$MSFE_{FPE} = \left(\frac{T+p+1}{T}\right)s_{\hat{u}}^{2}$$

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

Three ways to estimate the RMSFE (3 of 4)

3. Estimate the MSFE using by pseudo out-of-sample (POOS) forecasting

- The SER method of estimating the MSFE assumes stationarity and ignores estimation error.
- The FPE method of estimating the MSFE incorporates estimation error, but requires stationarity and homoskedasticity
- The third method for estimating the MSFE incorporates the estimation error and requires neither stationarity nor assumptions like homoskedasticity.
- This method, pseudo out-of-sample (POOS) forecasting, is based on simulating how the forecasting model would have done, had you been using it in real time.

Three ways to estimate the RMSFE (4 of 4)

3. Estimate the MSFE using by pseudo out-of-sample (POOS) forecasting

- Choose a date P for the start of your POOS sample, for example $P = 0.9T$.
- Re-estimate your model every period, $s = P–1,\ldots,T–1$
- Compute your pseudo out-of-sample forecast for date $s+1$, using the model estimated through s. This is $\hat{Y}_{s+1|s}$.
- Compute the POOS forecast error, $\tilde{u}_{s+1} = Y_{s+1} - \hat{Y}_{s+1|s}$
- Plug this forecast error into the MSFE formula,

$$MSFE_{POOS} = \frac{1}{P}\sum_{s=T-P+1}^{T} \tilde{u}_s^2$$

Why are these "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

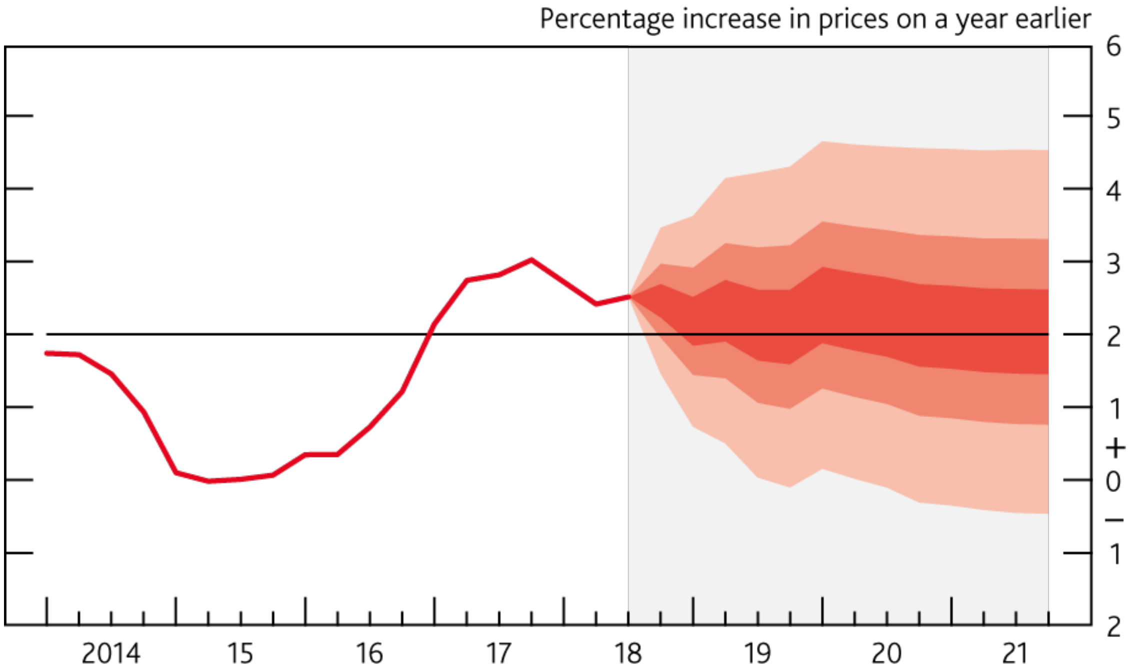
$$\hat{Y}_{T+1|T} \pm 1.96 \times RMSFE$$

Note:

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

Example #1: the Bank of England "Fan Chart", November 2018

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



https://www.bankofengland.co.uk/inflation-report/2018/november-2018/prospects-for-inflation

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

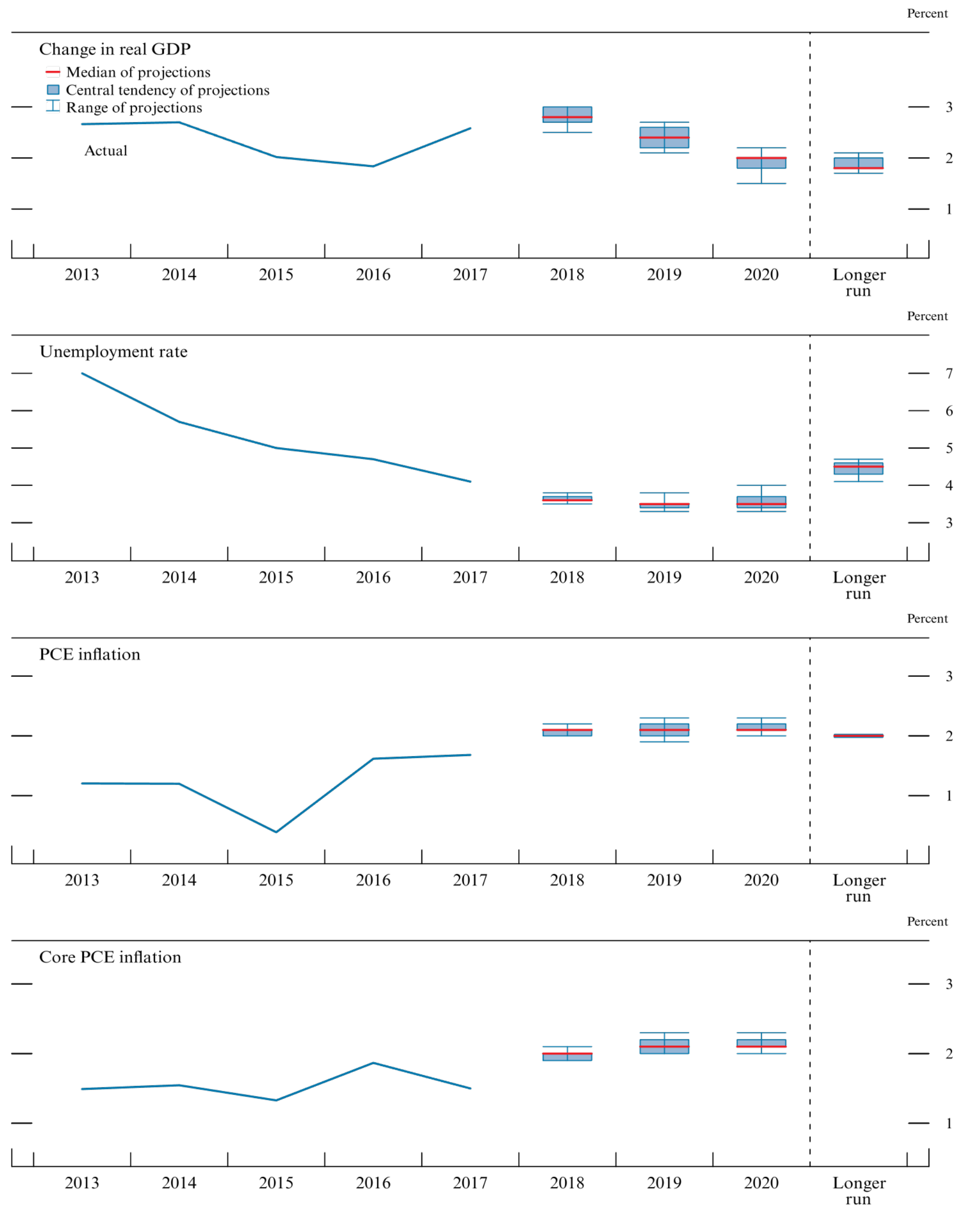
Table A Macroeconomic projections for the euro area
(average annual percentage changes; GDP only)

	December 2018			
	2018	2019	2020	2021
Real GDP	1.9	1.7	1.7	1.5
	[1.8 - 2.0][2)]	[1.1 - 2.3][2)]	[0.8 - 2.6][2)]	[0.5 - 2.5][2)]

Source: https://www.ecb.europa.eu/pub/projections/html/index.en.html

Example #3: Fed, Semiannual Report to Congress, July 2018

Figure 1. Medians, central tendencies, and ranges of economic projections, 2018–20 and over the longer run



Note: Definitions of variables and other explanations are in the notes to table 1. The data for the actual values of the variables are annual.

Source:

https://www.federalreserve.gov/monetarypolicy/2018-07-mpr-part3.htm

7. Lag Length Selection Using Information Criteria (SW Section 15.6)

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

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

The Bayes Information Criterion (BIC)

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

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

  - The result is that $\hat{p}^{BIC} \xrightarrow{p} p$! (*SW*, *App. 15.5*)

Another information criterion: *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 GDP Growth, lags 0 – 6:

TABLE 15.3 The Bayes Information Criterion (BIC) and the R^2 for Autoregressive Models of U.S. GDP Growth Rates, 1962:Q1–2017:Q3

p	$SSR(p)/T$	$\ln[SSR(p)/T]$	$(p+1)\ln(T)/T$	$BIC(p)$	R^2
0	10.477	2.349	0.024	2.373	0.000
1	9.247	2.224	0.048	2.273	0.117
2	8.954	2.192	0.073	2.265	0.145
3	8.954	2.192	0.097	2.289	0.145
4	8.920	2.188	0.121	2.310	0.149
5	8.788	2.173	0.145	2.319	0.161
6	8.779	2.172	0.170	2.342	0.162

- BIC chooses 2 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,

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

- 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)

8. Nonstationarity I: Trends (SW Section 15.7)

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

If stationarity doesn't hold, the series are said to be ***nonstationary***.

Two important types of nonstationarity are:

- Trends (SW Section 15.7)
- Structural breaks (model instability) (SW Section 15.8)

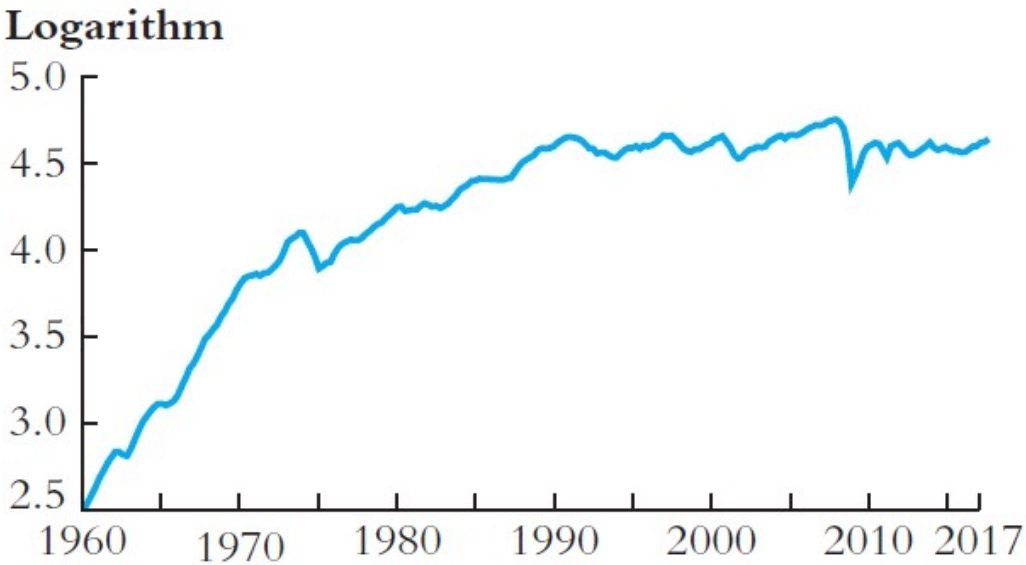
Outline of discussion of trends in time series data:

A. What is a trend?
B. Deterministic and stochastic (random) 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? (1 of 4)

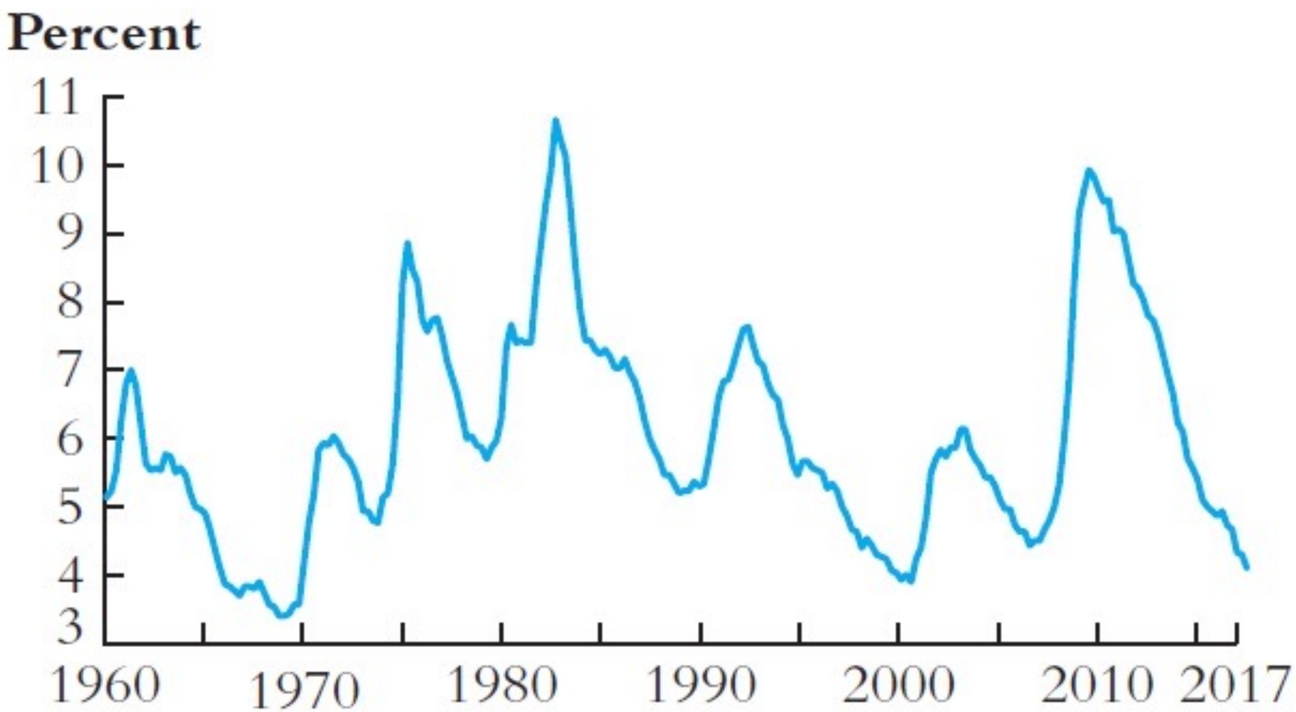
A trend is a persistent, long-term movement or tendency in the data. Trends need not be just a straight line!

Which of these series has a trend?



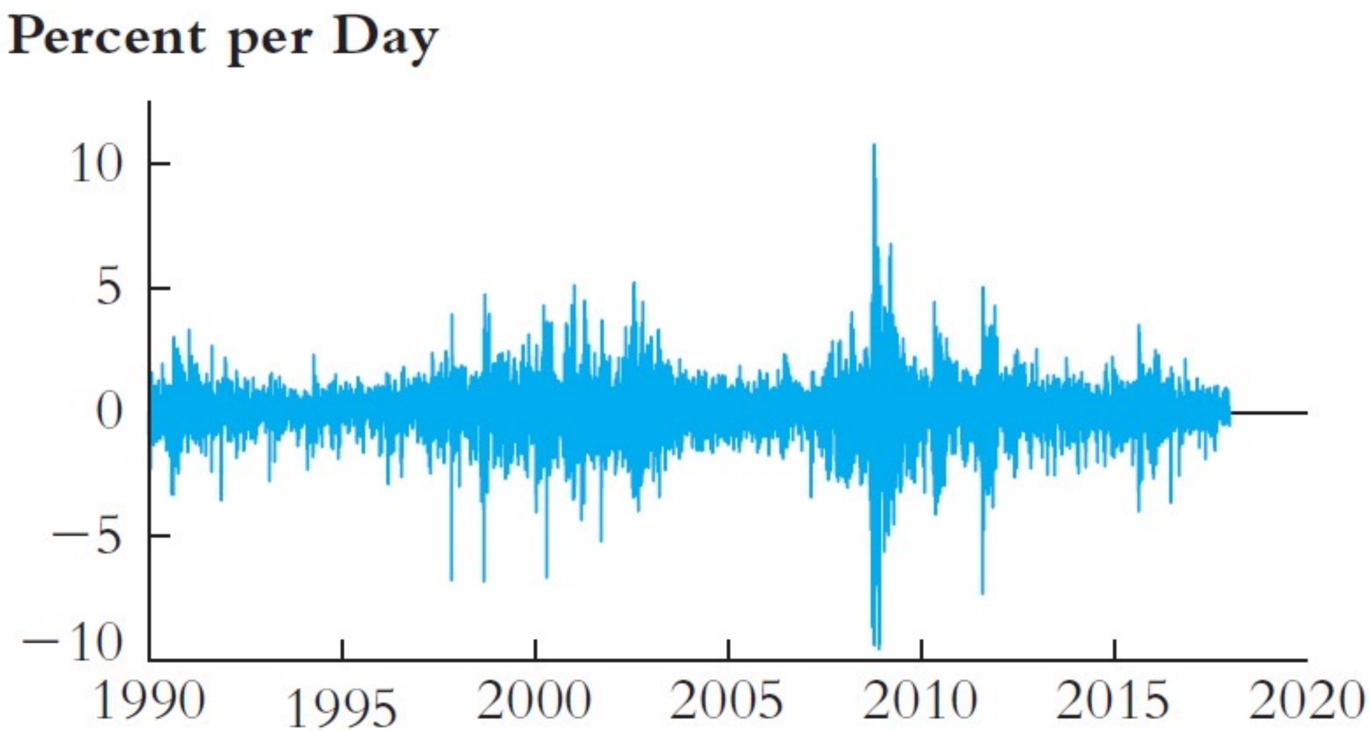
(c) Logarithm of Japan Index of Industrial Production

A. What is a trend? (2 of 4)



(a) U.S. unemployment rate

A. What is a trend? (3 of 4)



(d) Percentage change in daily value of the Wilshire 5000 Total Market Index

A. What is a trend? (4 of 4)

The three series:

- log Japan industrial production 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.
- The unemployment rate has long-term swings – a general increase from 1970 to 1990, and then a decrease from 1990 to 2007. But these long-term fluctuations are interrupted by large increases in recessions and decreases in expansions. Maybe it has a small trend – hard to tell.
- Daily stock price returns has no apparent trend. There are periods of persistently high volatility – but this isn’t a trend.

B. Deterministic and stochastic trends (1 of 5)

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

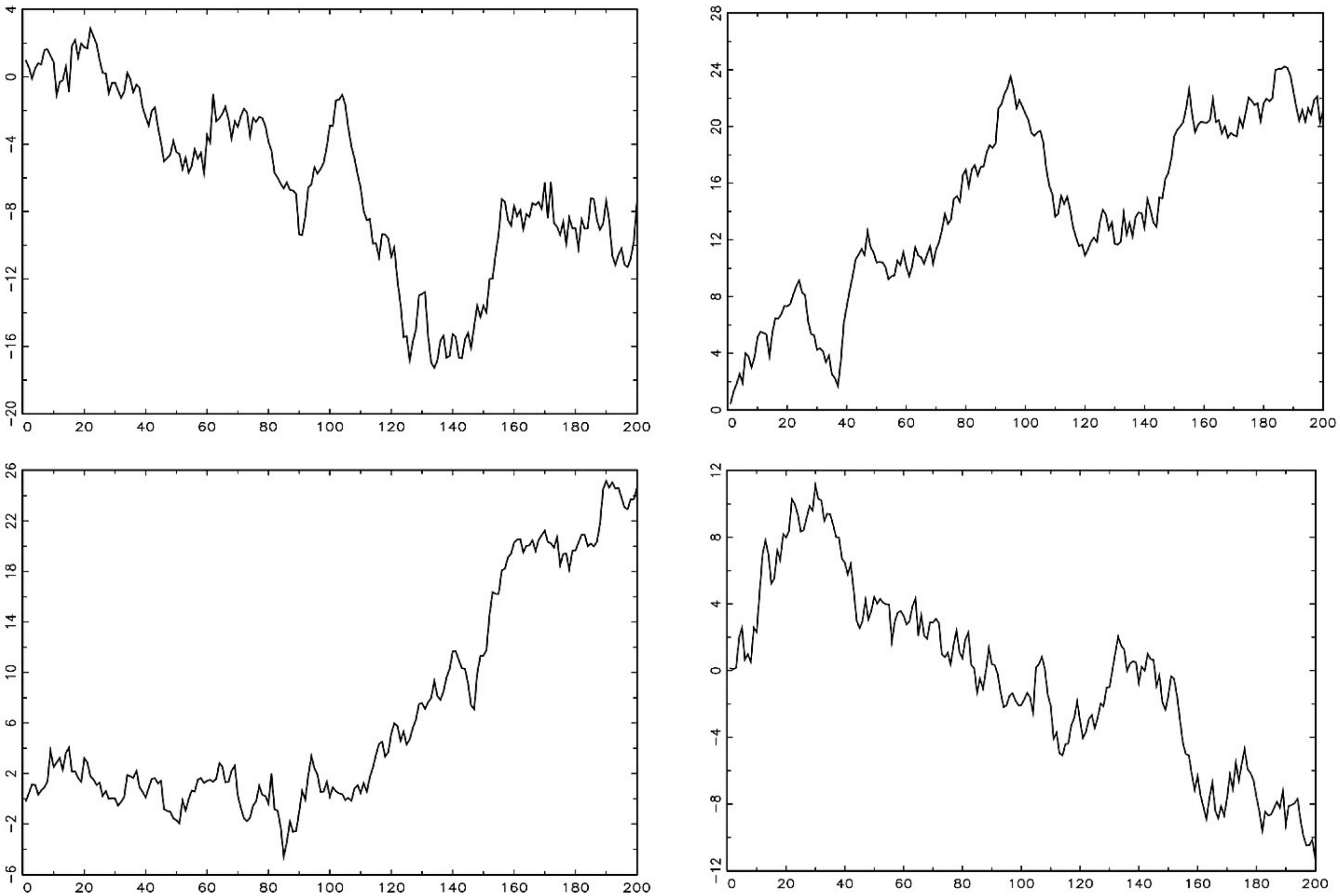
- A **deterministic trend** is a nonrandom function of time (e.g. $Y_t = t$, or $Y_t = t^2$).
- A **stochastic trend** is random and varies over time
- An important example of a stochastic trend is a **random walk**:

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

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

B. Deterministic and stochastic trends (2 of 5)

Four artificially generated random walks, *T* = 200:



How would you produce a random walk on the computer?

B. Deterministic and stochastic trends (3 of 5)

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 $\text{var}(Y_t) = t\sigma_u^2$.

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

B. Deterministic and stochastic trends (4 of 5)

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.

B. Deterministic and stochastic trends (5 of 5)

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")
- The Dickey-Fuller test for whether a Y_t has a random walk trend

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) (1 of 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 $\gamma_1 = -\beta_2 \cdots$

Unit roots in an AR(2) (2 of 2)

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 $\beta_1 + \beta_2 = 1$ (*show this yourself – find the roots!*)

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 (1 of 2)

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

This regression can be rearranged as,

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

where

$$\delta = \beta_1 + \beta_2 + \cdots + \beta_{p-1}$$

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

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

$$\cdots$$

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

Unit roots in the AR(p) model (2 of 2)

The AR(p) model can be written as,

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

where $\delta = \beta_1 + \beta_2 + \cdots + \beta_{p-1}$.

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:**

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

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…

U.S. Unemployment Rate and Log Japan IP

1965–1981:

$$\widehat{US\ URate} = -2.37 + \mathbf{2.22} \times \ln(Japanese\ IP),\ \bar{R}^2 = 0.34$$
$$(1.19)\ \mathbf{(0.32)}$$

1986–2012:

$$\widehat{US\ URate} = 41.78 - \mathbf{7.78} \times \ln(Japanese\ IP),\ \bar{R}^2 = 0.15$$
$$(1.19)\ \mathbf{(1.75)}$$

D. How do you detect stochastic trends? (1 of 2)

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.

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

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

or

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

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

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

D. How do you detect stochastic trends? (2 of 2)

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

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

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) $\Delta Y_t = \beta_0 + \mu t + \delta Y_{t-1} + u_t$ (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 15.4 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,

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

where $\delta = \beta_1 + \beta_2 + \cdots + \beta_{p-1}$. 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 15.4

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

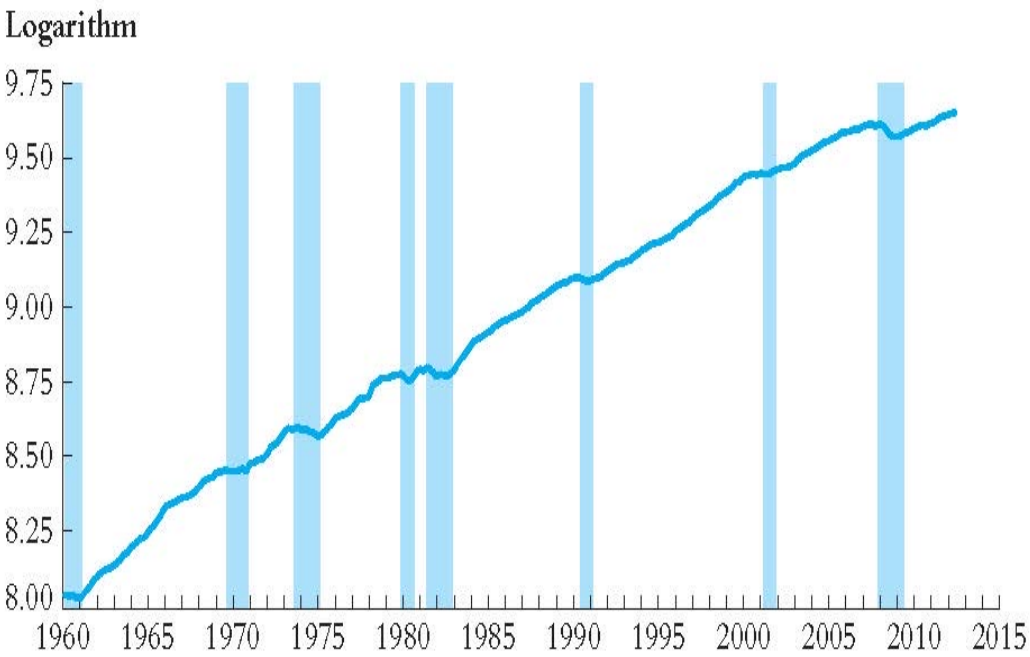
When should you include a time trend in the DF test?

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. GDP have a stochastic trend? (1 of 2)

The alternative is that logarithm of GDP is stationary around a linear time trend



(a) US GDP ($1996, Billions)

Example: Does U.S. GDP have a stochastic trend? (2 of 2)

$$\Delta \ln(GDP_t) = 0.162 + 0.0001t - \mathbf{0.019} \ln(GDP_{t-1})$$
$$(0.080) \quad (0.0001) \quad \mathbf{(0.010)}$$

$$+\ 0.261\Delta\ln(GDP_{t-1}) + 0.165\Delta\ln(GDP_{t-2})$$
$$(0.066) \qquad\qquad (0.066)$$

DF *t*-statstic = **–1.95**

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

DF t-statistic = –1.95 (intercept and time trend):

TABLE 15.4 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 = –1.95 does no reject a unit root at 10% level.

Note: you can choose the lag length in the DF regression by BIC or AIC. (For GDP two lags are chosen as in the ADF regression above.)

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 (ΔY_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 ΔY_t for regression analysis and forecasting. If no unit root, use Y_t.

9. Nonstationarity II: Breaks and Model Stability (SW Section 15.8)

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: predicting GDP with the term spread

A. Tests for a break (change) in regression coefficients (1 of 2)

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.

A. Tests for a break (change) in regression coefficients (2 of 2)

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

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}.
- Unfortunately, you often don’t have a candidate break date, that is, you don’t know τ…

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 (*why?*)

The Quandt Likelihood Ratio (QLR) Statistic (also called the "sup-Wald" statistic) (1 of 5)

The QLR statistic = the maximum Chow statistic

- 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, $\tau_0 \leqslant \tau \leqslant \tau_1$:

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

- 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 Quandt Likelihood Ratio (QLR) Statistic (also called the "sup-Wald" statistic) (2 of 5)

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

- 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!

The Quandt Likelihood Ratio (QLR) Statistic (also called the "sup-Wald" statistic) (3 of 5)

- ***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 $\{B_i\}$, $i = 1, \cdots, n$, are independent continuous-time "Brownian Bridges" on $0 \leqslant s \leqslant 1$ (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 15.5⋯

The Quandt Likelihood Ratio (QLR) Statistic (also called the "sup-Wald" statistic) (4 of 5)

TABLE 15.5 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.

The Quandt Likelihood Ratio (QLR) Statistic (also called the "sup-Wald" statistic) (5 of 5)

TABLE 15.5 (*Continued*)

Number of Restrictions (q)	10%	5%	1%
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

Note: 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).

Example: Has the GDP/Term Spread relationship been stable?

Recall the ADL(2,2) model of $GDPGR_t$ and $TSpread_t$, estimated over (1962:Q1 – 2017:Q3):

$$\underset{}{GDPGR_t} = \underset{(0.46)}{0.94} + \underset{(0.08)}{0.25}GDPGR_{t-1} + \underset{(0.08)}{0.18}GDPGR_{t-2}$$

$$\underset{(0.42)}{-0.13}\,TSpread_{t-1} + \underset{(0.43)}{0.62}\,TSpread_{t-2} \qquad \bar{R}^2 = 0.16$$

Has this model been stable over the full period 1962–2017?

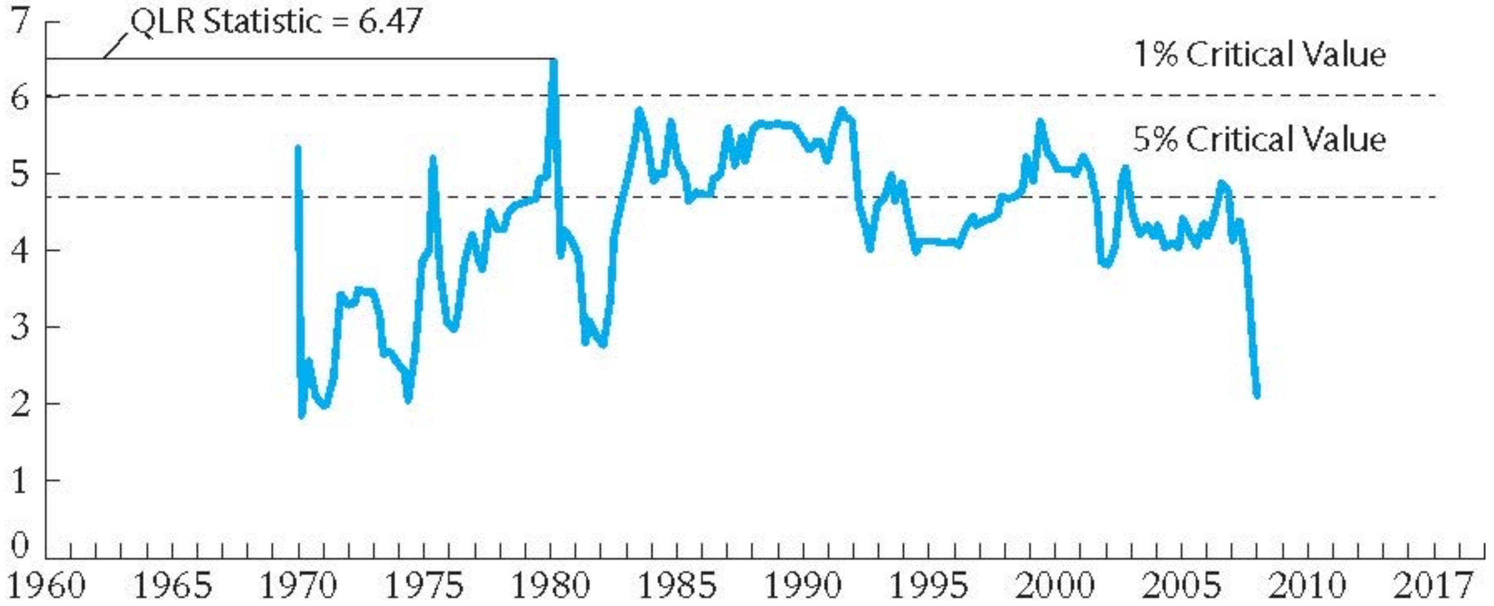
QLR tests of the ADL(2,2) model (1 of 2)

dependent variable: $GDPGR_t$

regressors: intercept, $GDPGR_{t-1}$, $GDPGR_{t-2}$, $TSpread_{t-1}$, and $TSpread_{t-2}$

- test for constancy of intercept and coefficients on $TSpread_{t-1}$, and $TSpread_{t-2}$ (coefficients on $\Delta GDPGR_{t-1},\cdots, GDPGR_{t-2}$ are constant): $QLR = 6.47$ ($q = 3$)
  - 1% critical value = 6.02 $\Rightarrow$ reject at 1% level
  - Estimate break date: maximal F occurs in 1980:Q4
- Conclude that there is a break in the TSpread/GDP relation, with estimated date of 1980:Q4

QLR tests of the ADL(2,2) model (2 of 2)



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 GDP-Term Spread model:

- We found a break in 1980:Q4 – so for this analysis, we only consider regressions that start in 1981:Q1 – ignore the earlier data from the "old" model ("regime").
- <u>Regression model:</u>

  dependent variable: $GDPGR_t$

  regressors: intercept, $GDPGR_{t-1}$, $GDPGR_{t-2}$, $TSpread_{t-1}$, and $TSpread_{t-2}$
- Pseudo out-of-sample forecasts:
  - Compute regression over t = 1980:Q1,⋯, P

  – Compute *poos* forecast, $\widehat{GDPGR}_{P+1|P}$, and forecast error

  - Repeat for P = 2003:Q1,⋯, 2017:Q4

Pearson

POOS forecasts of *GDPGR* using ADL(2,2) model with *TSpread*

The forecasts generally track the growth rate of GDP from 2003–2017 but fail to forecast the sharp decline during the financial crisis recession (2008-2009)

10. Summary: Time Series Forecasting Models (1 of 2)

- 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.

10. Summary: Time Series Forecasting Models (2 of 2)

- 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