Session 3: Time Series Econometrics

Prepared by Ziyodullo Parpiev, PhD

for Regional Summer School

September 21, 2016

Outline

1. Stochastic processes

2. Stationary processes

3. Nonstationary processes

4. Integrated variables

5. Random walk models

6. Unit root tests

7. Cointegration and error correction models

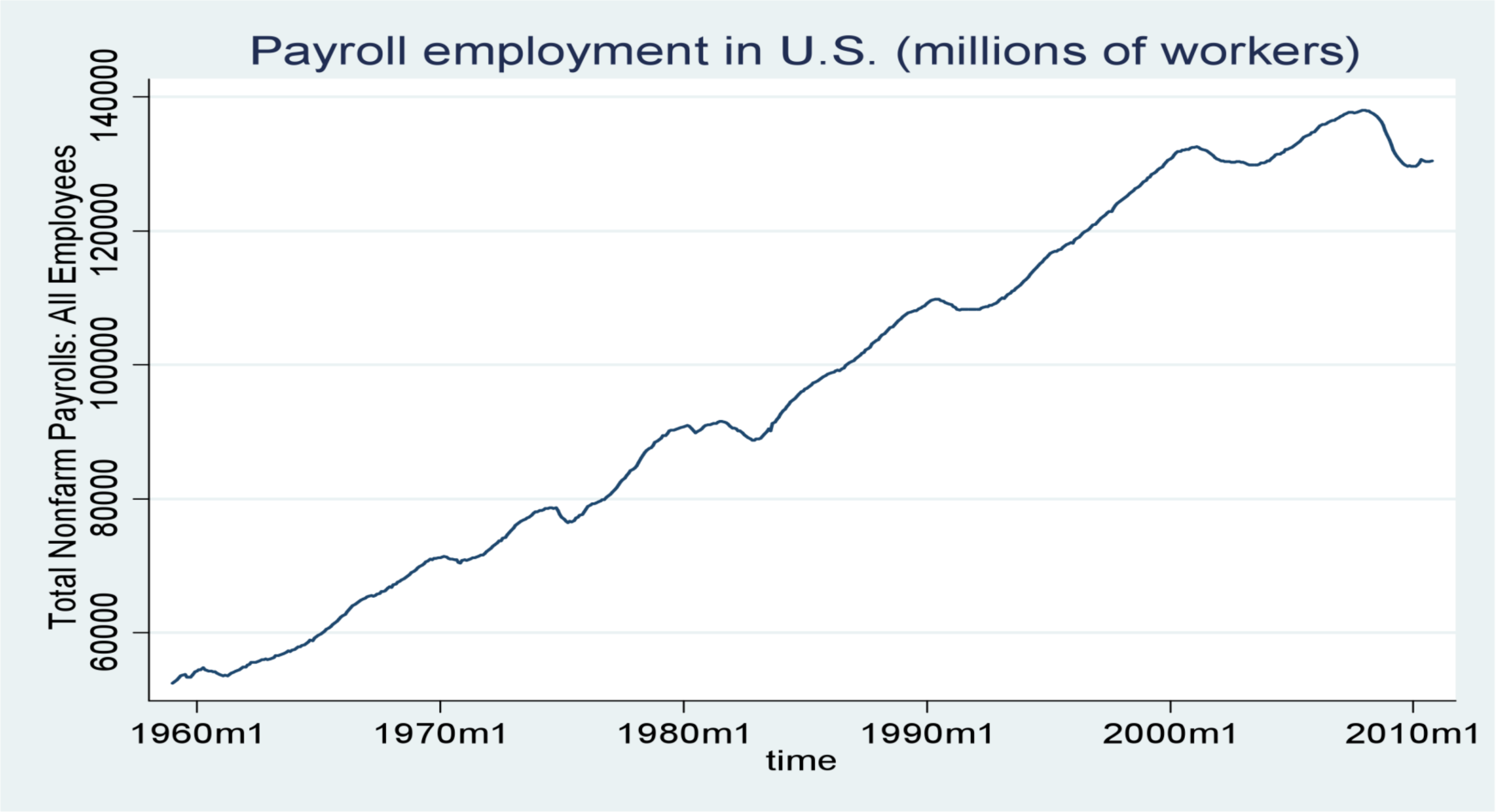
What is a time series?

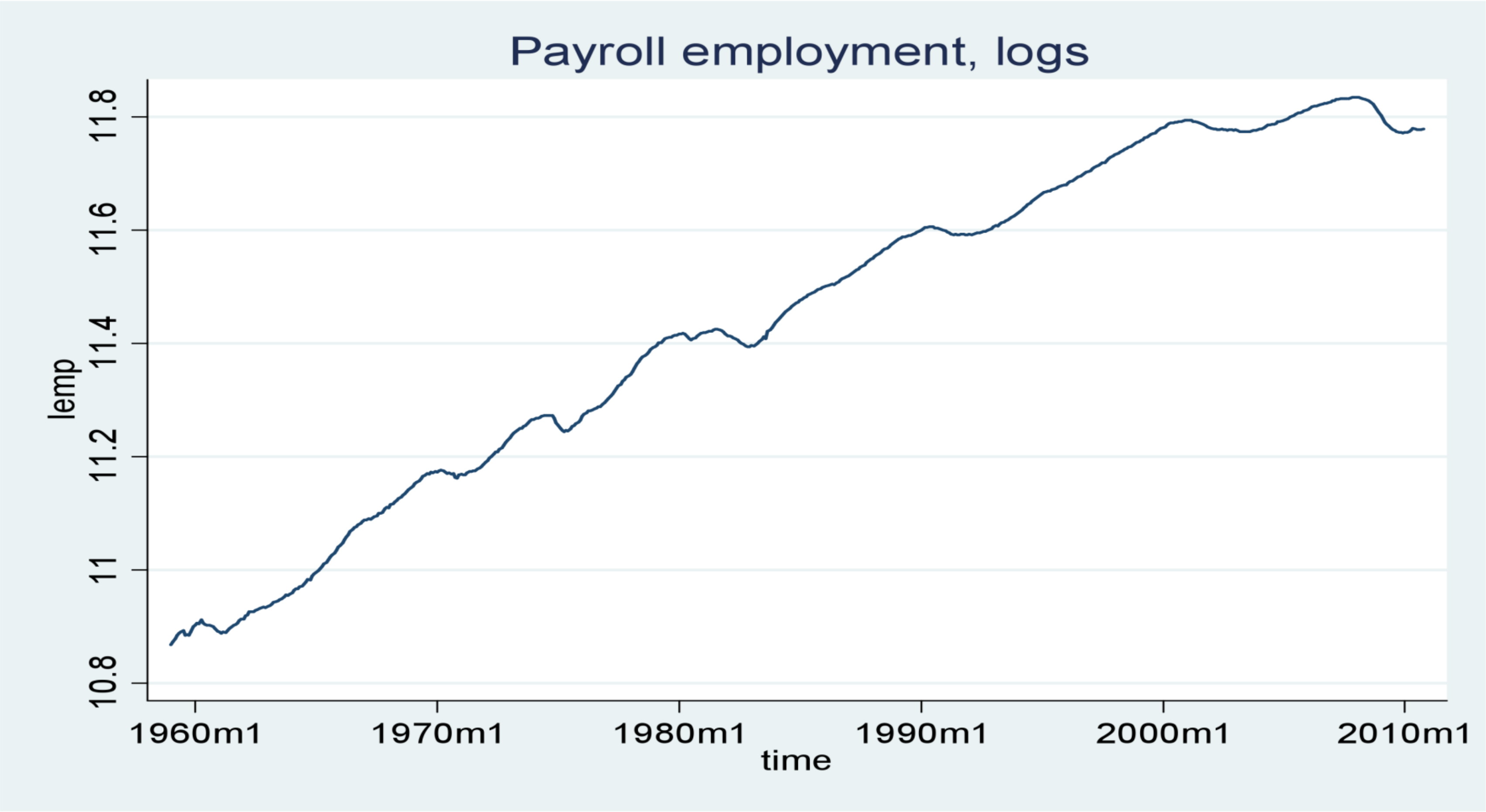
A time series is any series of data that varies over time. For example

- Payroll employment in the U.S.
- Unemployment rate
- 12-month inflation rate
- Daily price of stocks and shares
- Quarterly GDP series
- Annual precipitation (rain and snowfall)

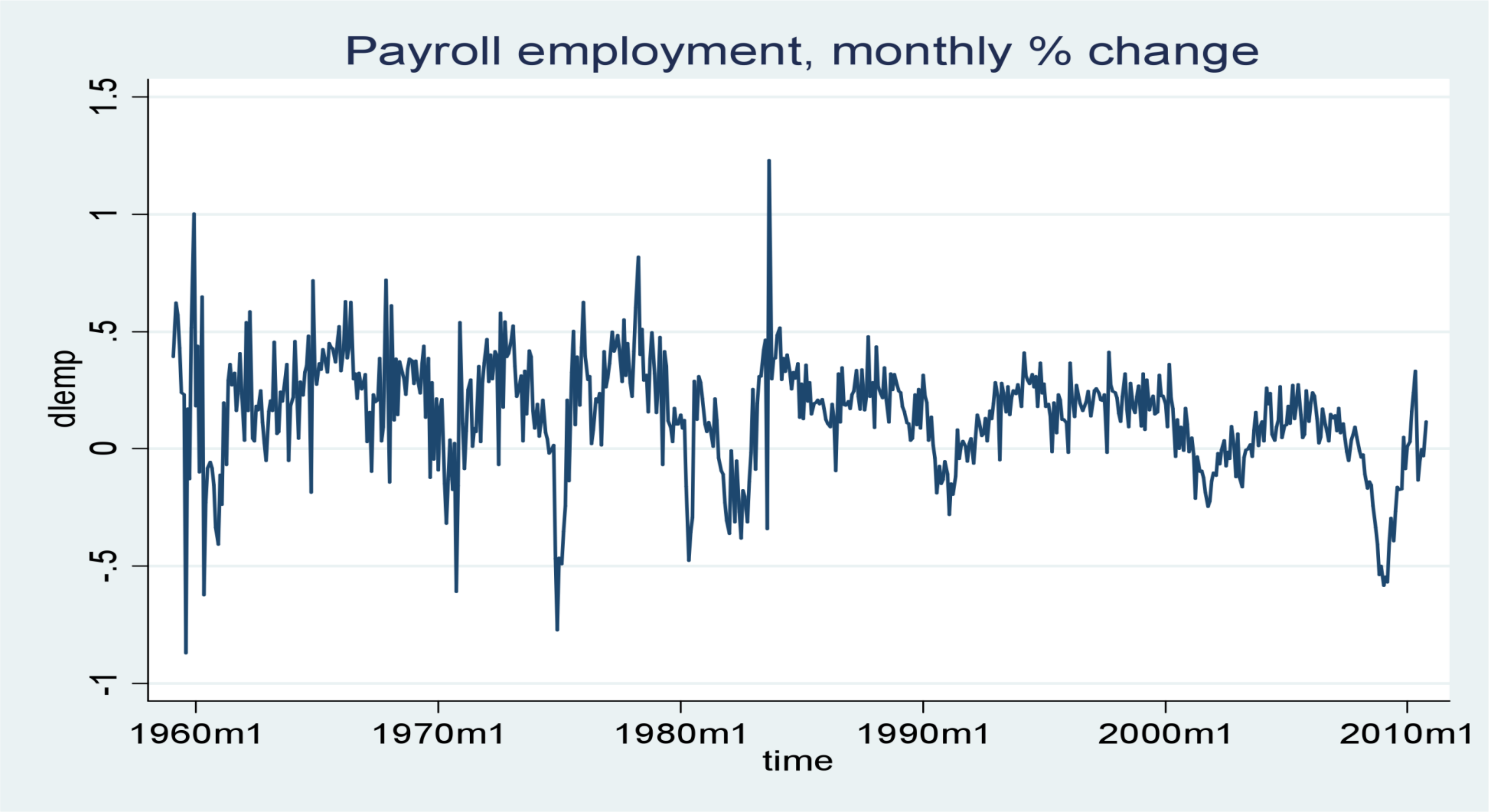
Because of widespread availability of time series databases most empirical studies use time series data.

Some monthly U.S. macro and financial time series

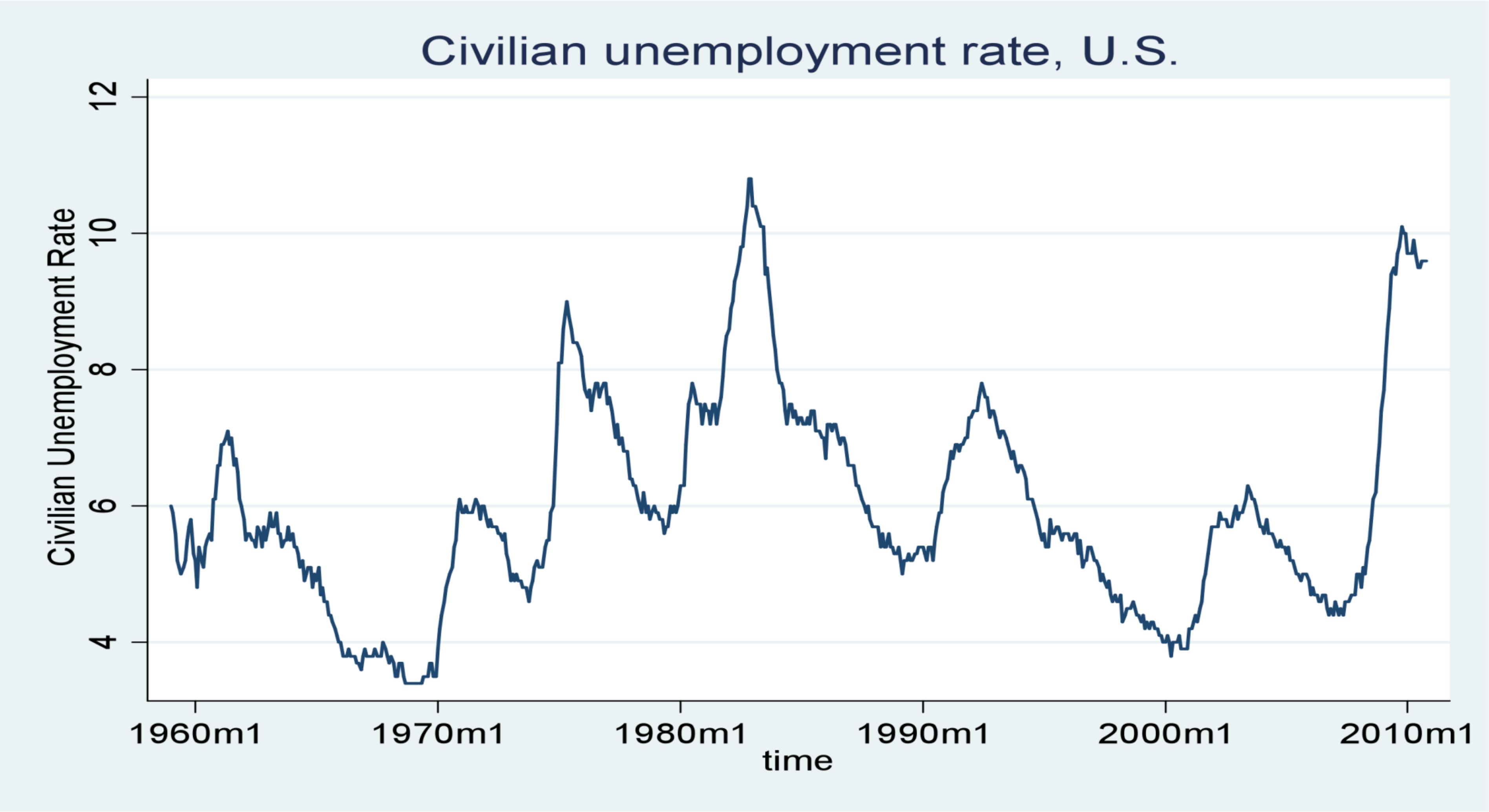


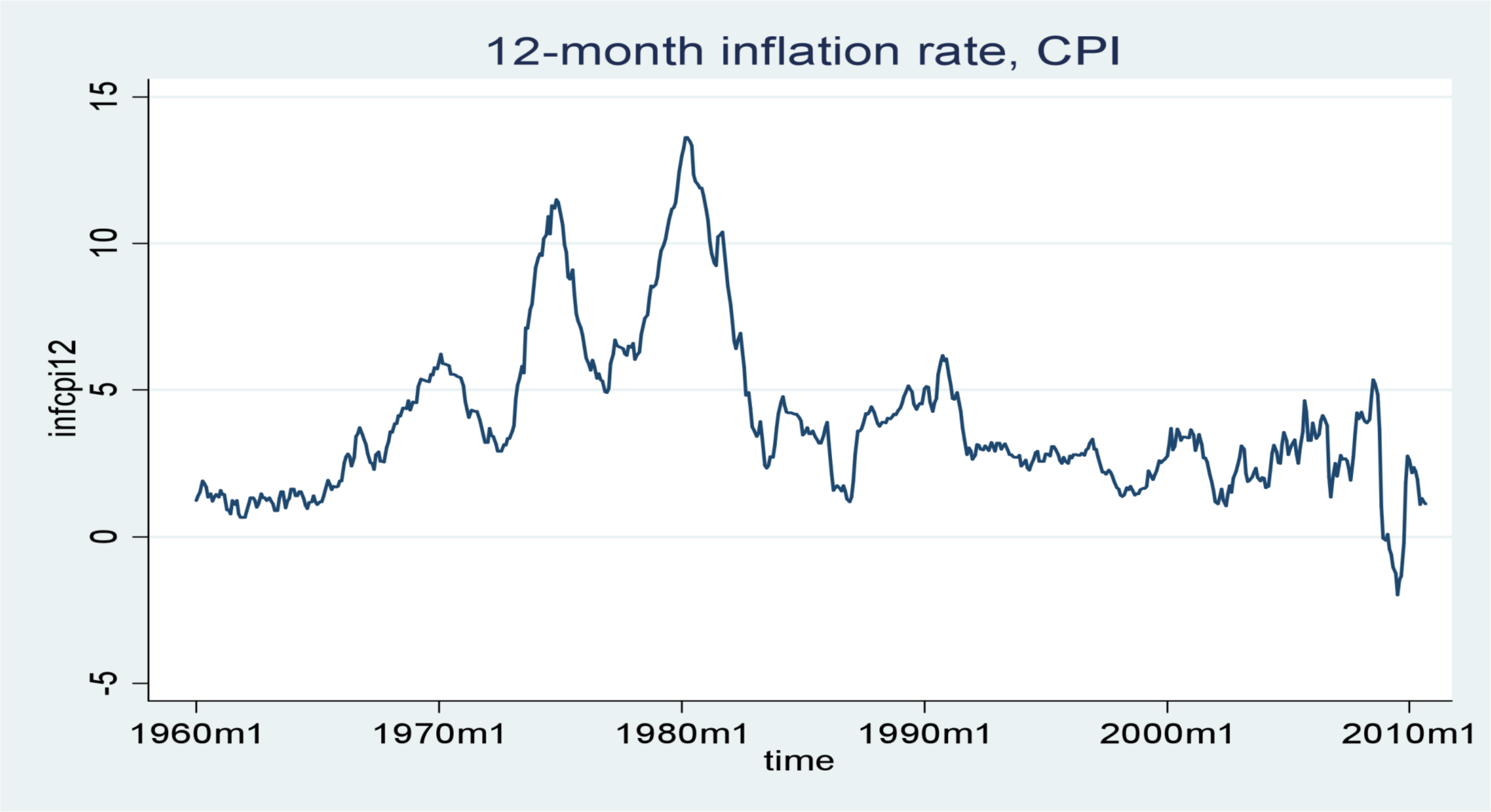


Payroll employment, logs
lemp
10.8
11
11.2
11.4
11.6
11.8
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time

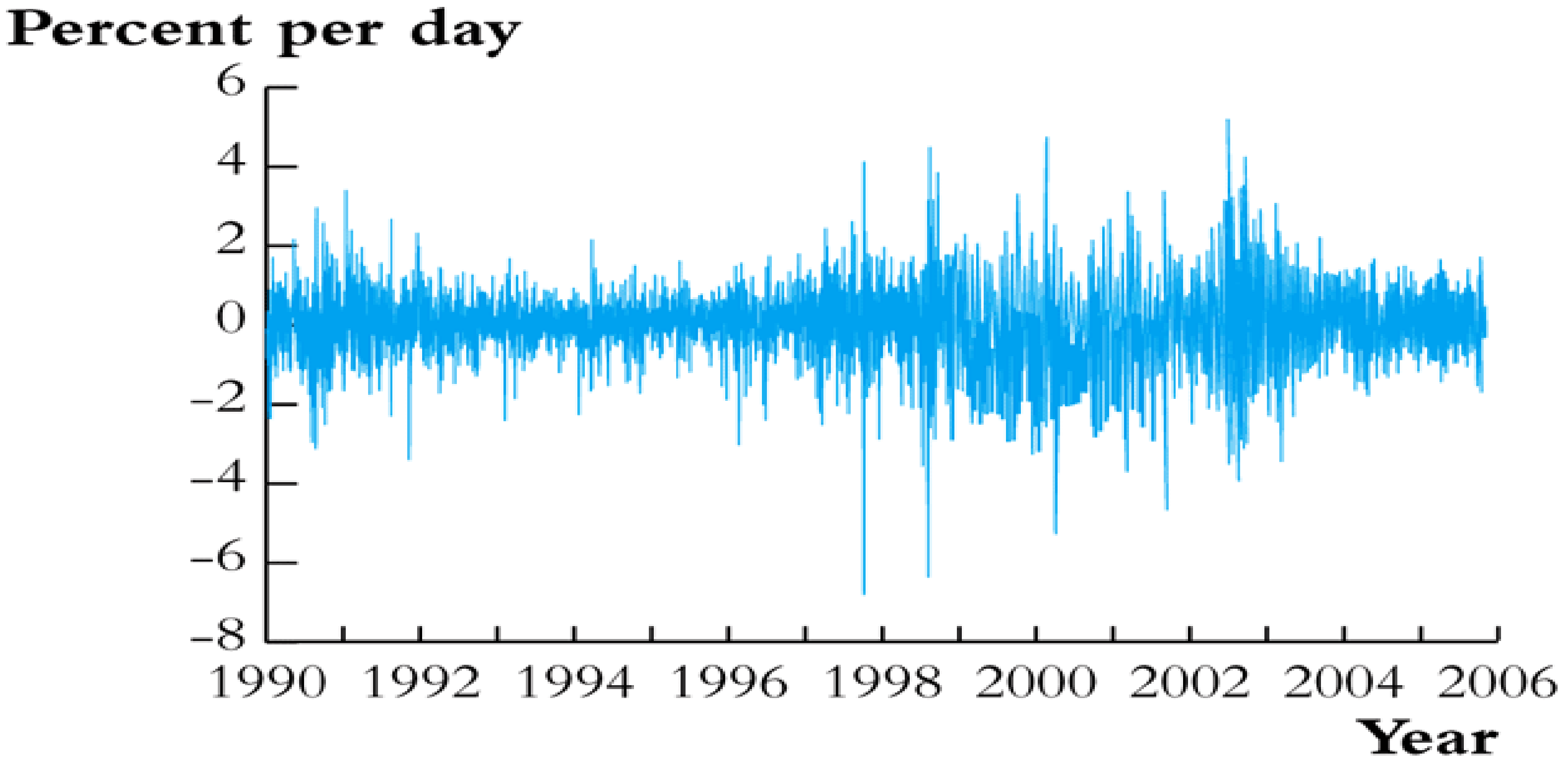


Payroll employment, monthly % change
dlemp
1.5
1
.5
0
-.5
-1
1960m1
1970m1
1980m1
1990m1
2000m1
2010m1
time


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


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

A daily financial time series:



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

STOCHASTIC PROCESSES

- *A random or stochastic process is a collection of random variables ordered in time.*
- If we let *Y* denote a random variable, and if it is continuous, we denote it as *Y*(*t*), but if it is discrete, we denoted it as *Yt*. An example of the former is an electrocardiogram, and an example of the latter is GDP, PDI, etc. Since most economic data are collected at discrete points in time, for our purpose we will use the notation *Yt* rather than *Y*(*t*).
- *Keep in mind that each of these Y's is a random variable.* In what sense can we regard GDP as a stochastic process? Consider for instance the GDP of $2872.8 billion for 1970–I. In theory, the GDP figure for the first quarter of 1970 could have been any number, depending on the economic and political climate then prevailing. The figure of 2872.8 is a particular **realization** of all such possibilities. The distinction between the stochastic process and its realization is akin to the distinction between population and sample in cross-sectional data.

Stationary Stochastic Processes

Forms of Stationarity: weak, and strong

(i) mean: $E(Y_t) = \mu$

(ii) variance: $var(Y_t) = E(Y_t – \mu)^2 = \sigma^2$

(iii) Covariance: $\gamma_k = E[(Y_t – \mu)(Y_{t-k} – \mu)^2$

where *γk*, the covariance (or autocovariance) at lag *k*, is the covariance between the values of *Yt* and *Yt*+*k*, that is, between two *Y* values *k* periods apart. If *k* = 0, we obtain *γ*0, which is simply the variance of *Y* (= *σ*2); if *k* = 1, *γ*1 is the covariance between two adjacent values of *Y*, the type of covariance we encountered in Chapter 12 (recall the Markov first-order autoregressive scheme).

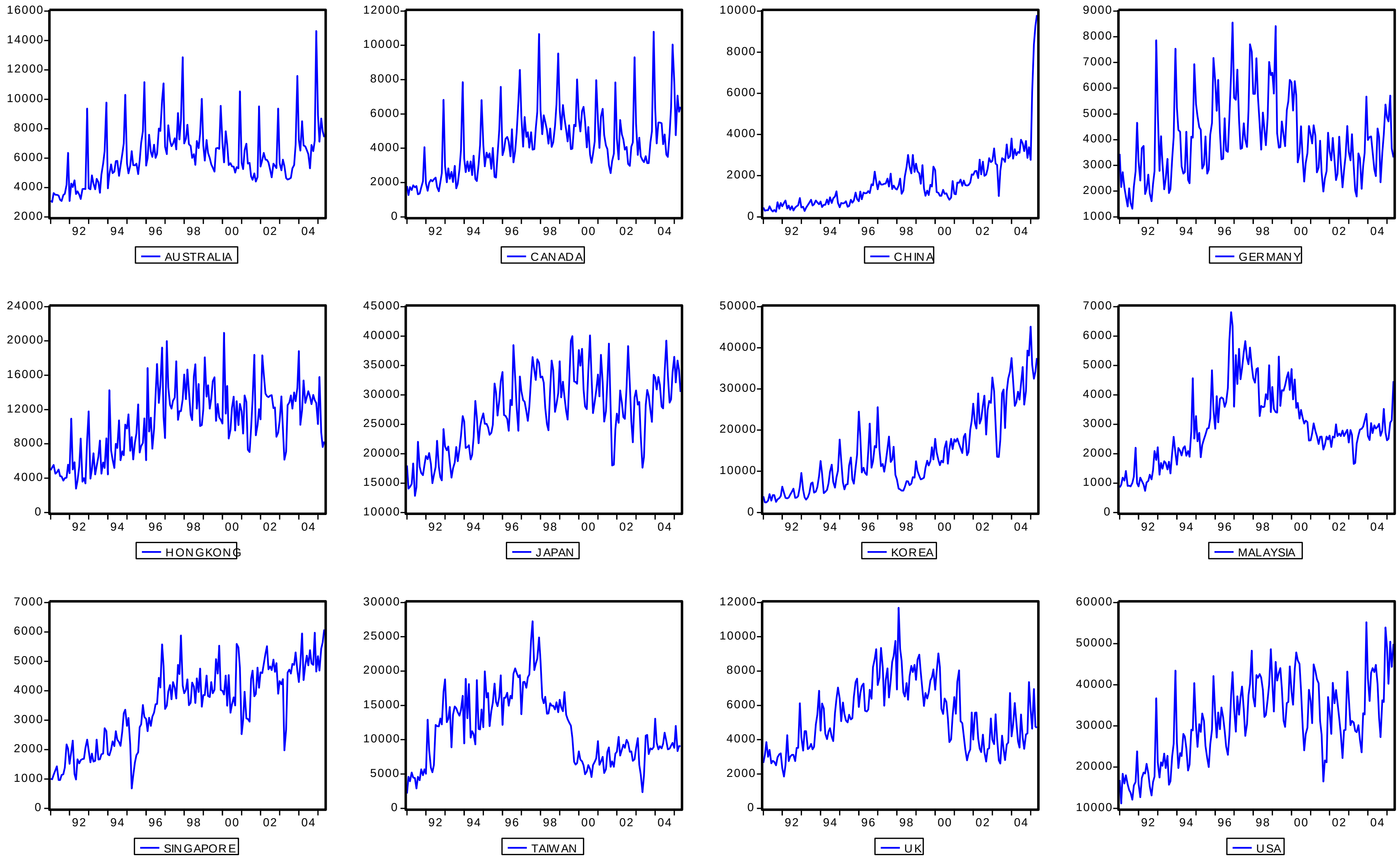
Types of Stationarity

- A time series is weakly stationary if its mean and variance are constant over time and the value of the covariance between two periods depends only on the distance (or lags) between the two periods.
- A time series if strongly stationary if for any values j_1, j_2,…j_n, the joint distribution of (Y_t, Y_{t+j1}, Y_{t+j2},…Y_{t+jn}) depends only on the intervals separating the dates (j_1, j_2,…,j_n) and not on the date itself (t).
- A weakly stationary series that is Gaussian (normal) is also strictly stationary.
- This is why we often test for the normality of a time series.

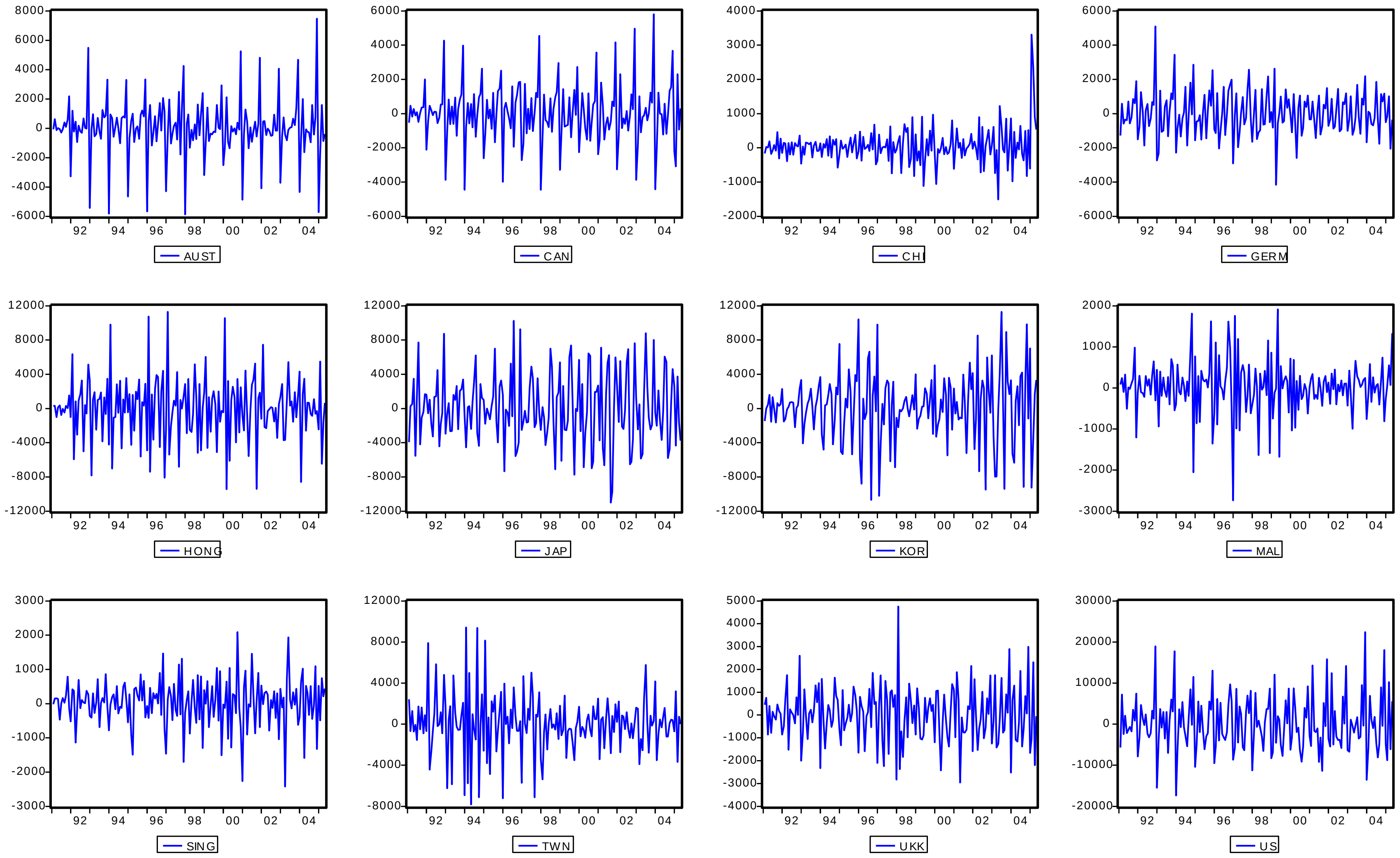
Stationarity vs. Nonstationarity

- *A time series is stationary, if its mean, variance, and autocovariance (at various lags) remain the same no matter at what point we measure them; that is, they are time invariant.*
- Such a time series will tend to return to its mean (called **mean reversion**) and fluctuations around this mean (measured by its variance) will have a broadly constant amplitude.
- If a time series is not stationary in the sense just defined, it is called a **nonstationary time series** (keep in mind we are talking only about weak stationarity). In other words, a nonstationary time series will have a *time-varying mean or a time-varying variance or both.*
- Why are stationary time series so important? Because if a time series is nonstationary, we can study its behavior only for the time period under consideration. Each set of time series data will therefore be for a particular episode. As a consequence, it is not possible to generalize it to other time periods. Therefore, for the purpose of forecasting, such time series may be of little practical value.

Examples of Non-Stationary Time Series



Examples of Stationary Time Series



“Unit Root” and order of integration

If a Non-Stationary Time Series Y_t has to be “differenced” d times to make it stationary, then Y_t is said to contain d “Unit Roots”. It is customary to denote Y_t ~ I(d) which reads “Y_t is integrated of order d”

If Y_t ~ I(0), then Y_t is Stationary

If Y_t ~ I(1), then $Z_t = Y_t - Y_{t-1}$ is Stationary

If Y_t ~ I(2), then $Z_t = Y_t - Y_{t-1} - (Y_t - Y_{t-2})$is Stationary

Unit Roots

- Consider an AR(1) process:

$y_t = a_1 y_{t-1} + \varepsilon_t$ (Eq. 1)

$\varepsilon_t \sim N(0, \sigma^2)$

- Case #1: Random walk ($a_1 = 1$)

$y_t = y_{t-1} + \varepsilon_t$

$\Delta y_t = \varepsilon_t$

Unit Roots

- In this model, the variance of the error term, ε_t, increases as t increases, in which case OLS will produce a downwardly biased estimate of a_1 (Hurwicz bias).
- Rewrite equation 1 by subtracting y_{t-1} from both sides:

$y_t – y_{t-1} = a_1 y_{t-1} – y_{t-1} + \varepsilon_t$ (Eq. 2)

$\Delta y_t = \delta\, y_{t-1} + \varepsilon_t$

$\delta = (a_1 – 1)$

Unit Roots

- H_0: $\delta = 0$ (there is a unit root)
- H_A: $\delta \neq 0$ (there is not a unit root)
- If $\delta = 0$, then we can rewrite Equation 2 as

$$\Delta y_t = \varepsilon_t$$

Thus first differences of a random walk time series are stationary, because by assumption, ε_t is purely random.

In general, a time series must be differenced *d* times to become stationary; it is integrated of order *d* or I(d). A stationary series is I(0). A random walk series is I(1).

Tests for Unit Roots

- Dickey-Fuller test
  - Estimates a regression using equation 2
  - The usual t-statistic is not valid, thus D-F developed appropriate critical values.
  - You can include a constant, trend, or both in the test.
  - If you accept the null hypothesis, you conclude that the time series has a unit root.
  - In that case, you should first difference the series before proceeding with analysis.

Tests for Unit Roots

- Augmented Dickey-Fuller test (dfuller in STATA)
  - We can use this version if we suspect there is autocorrelation in the residuals.
  - This model is the same as the DF test, but includes lags of the residuals too.
- Phillips-Perron test (pperron in STATA)
  - Makes milder assumptions concerning the error term, allowing for the ε_t to be weakly dependent and heterogenously distributed.
- Other tests include KPSS test, Variance Ratio test, and Modified Rescaled Range test.
- There are also unit root tests for panel data (Levin et al 2002, Pesaran et al).

Tests for Unit Roots

- These tests have been criticized for having low power (1-probability(Type II error)).
- They tend to (falsely) accept H_o too often, finding unit roots frequently, especially with seasonally adjusted data or series with structural breaks. Results are also sensitive to # of lags used in the test.
- Solution involves increasing the frequency of observations, or obtaining longer time series.

Trend Stationary vs. Difference Stationary

- Traditionally in regression-based time series models, a time trend variable, t, was included as one of the regressors to avoid spurious correlation.
- This practice is only valid if the trend variable is deterministic, not stochastic.
- A trend stationary series has a data generating process (DGP) of:

$$y_t = a_0 + a_1 t + \varepsilon_t$$

Trend Stationary vs. Difference Stationary

- A difference stationary time series has a DGP of:

$$y_t - y_{t-1} = a_0 + \varepsilon_t$$

$$\Delta y_t = a_0 + \varepsilon_t$$

- Run the ADF test with a trend. If the test still shows a unit root (accept H_o), then conclude it is difference stationary. If you reject H_o, you could simply include the time trend in the model.

What is a Spurious Regression?

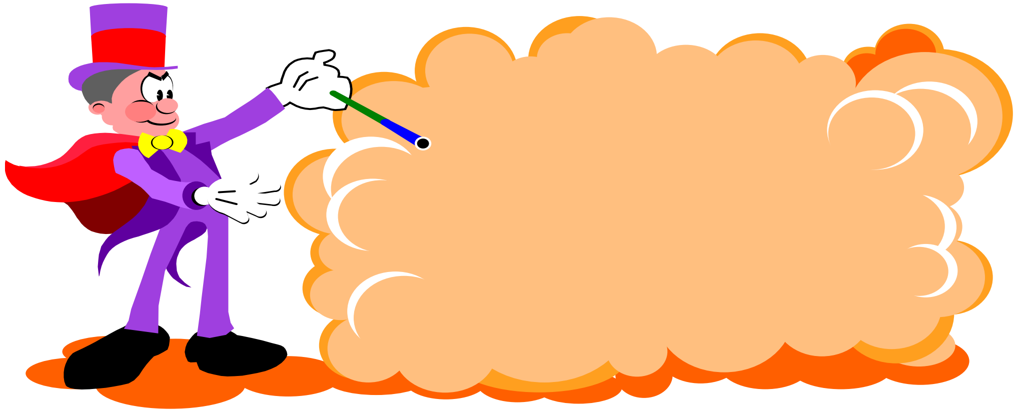
A Spurious or Nonsensical relationship may result when one Non-stationary time series is regressed against one or more Non-stationary time series

The best way to guard against Spurious Regressions is to check for “Cointegration” of the variables used in time series modeling

Symptoms of Likely Presence of Spurious Regression

- If the R^2 of the regression is greater than the Durbin-Watson Statistic

- If the residual series of the regression has a Unit Root



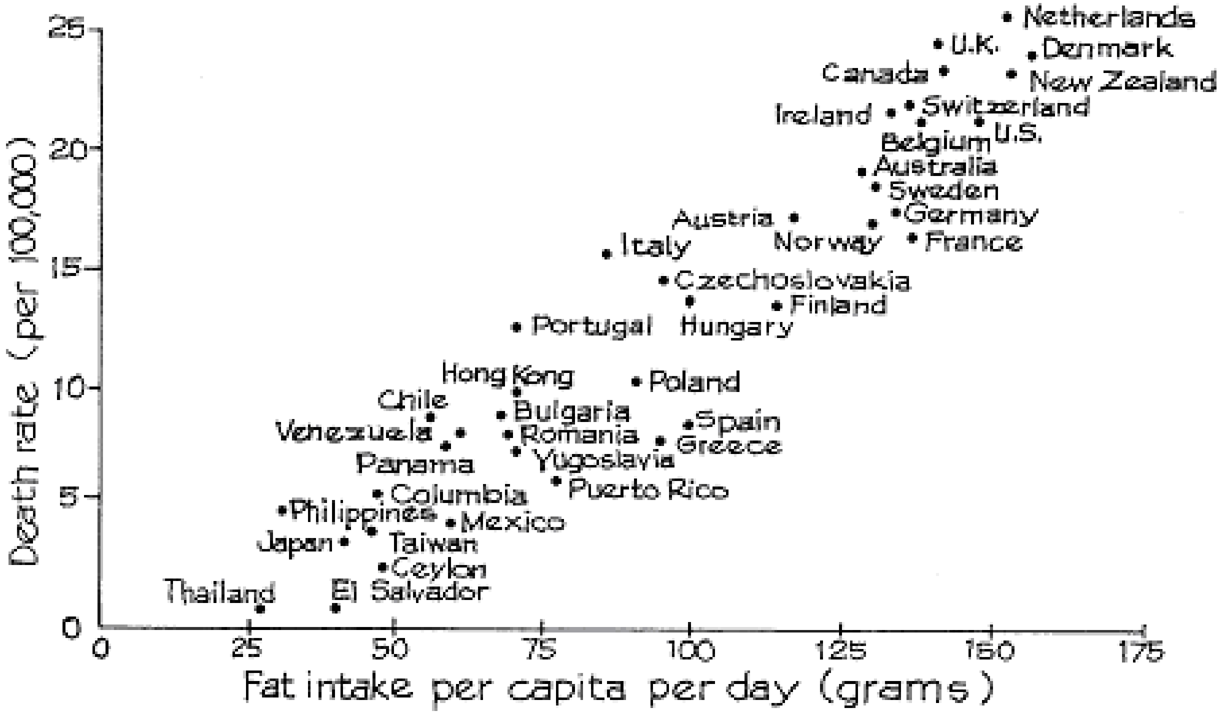
Examples of spurious relationships

- For school children, shoe size is strongly correlated with reading skills.
- Amount of ice cream sold and death by drowning in monthly data
- Number of doctors and number of people dying of disease in cities
- Number of libraries and number of people on drugs in annual data
- **Bottom line: Correlation measures association. But association is not the same as causation.**

More complicated case of spurious regression

- Fat in the diet seems to be correlated with cancer. Can we say the diagram is some evidence for the theory?
- But the evidence is quite weak, because other things aren't equal. For example, the countries with lots of fat in the diet also have lots of sugar. A plot of colon cancer rates against sugar consumption would look just like figure 8, and nobody thinks that sugar causes colon cancer. As it turns out, fat and sugar are relatively expensive. In rich countries, people can afford to eat fat and sugar rather than starchier grain products. Some aspects of the diet in these countries, or other factors in the life-style, probably do cause certain kinds of cancer and protect against other kinds. So far, epidemiologists can identify only a few of these factors with any real confidence. Fat is not among them

Figure 8. Cancer rates plotted against fat in the diet, for a sample of countries.



Source: K. Carroll, "Experimental evidence of dietary factors and hormone-dependent cancers," *Cancer Research* vol. 35 (1975) p. 3379. Copyright by *Cancer Research*. Reproduced by permission.

Cointegration

- Is the existence of a long run equilibrium relationship among time series variables
- Is a property of two or more variables moving together through time, and despite following their own individual trends will not drift too far apart since they are linked together in some sense

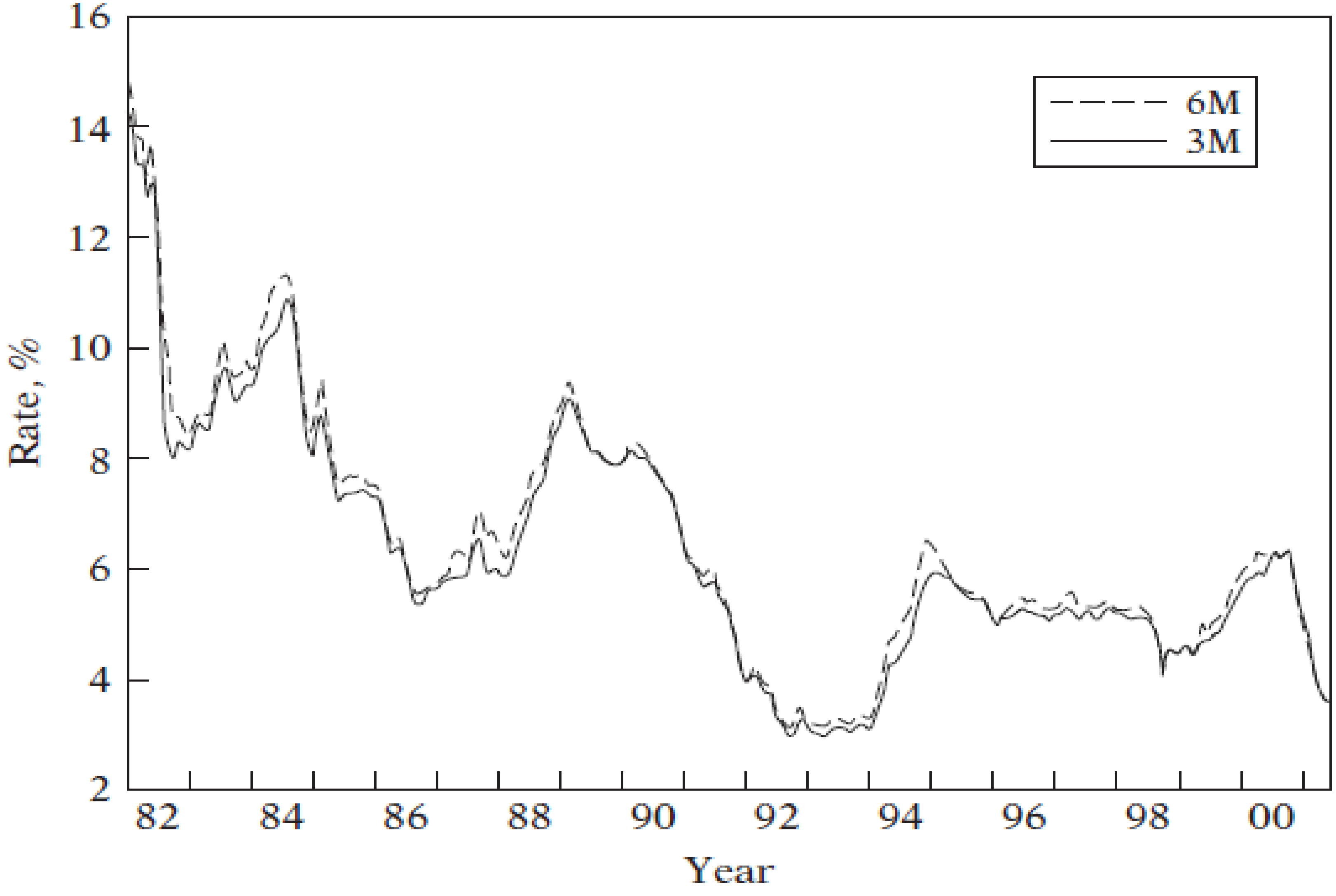


FIGURE 21.13
Three- and six-month Treasury bill rates (constant maturity).

Cointegration Analysis: Formal Tests

- Cointegrating Regression Durbin-Watson (CRDW) Test
- Augmented Engle-Granger (AEG) Test
- Johansen Multivariate Cointegration Tests or the Johansen Method

Error Correction Mechanism (ECM)

- Reconciles the Static LR Equilibrium relationship of Cointegrated Time Series with its Dynamic SR disequilibrium
- Based on the Granger Representation Theorem which states that “If variables are cointegrated, the relationship among them can be expressed as ECM”.