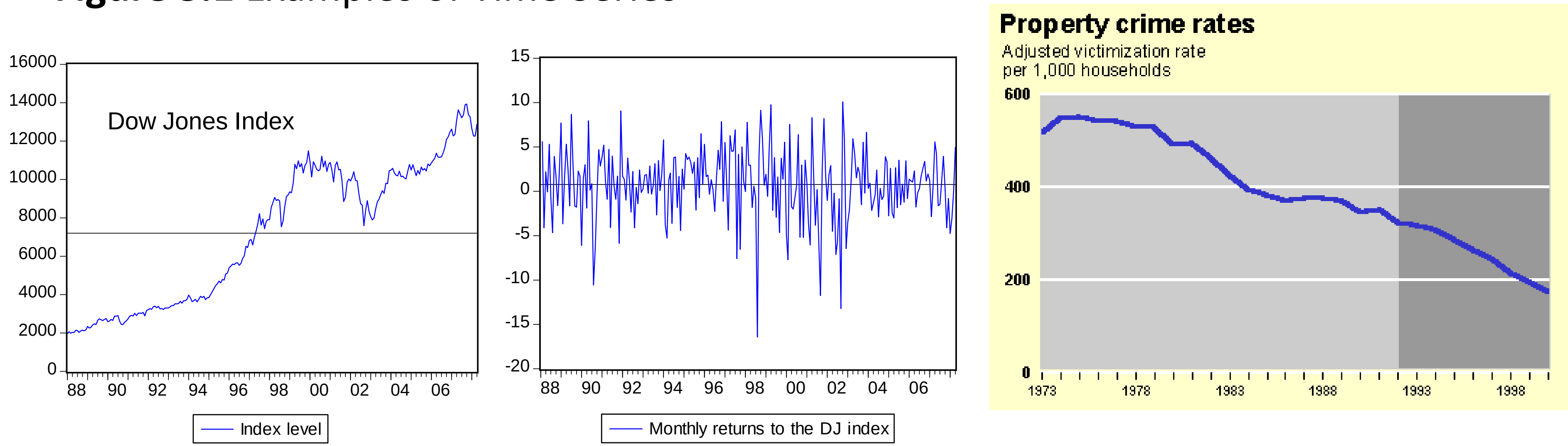
CHAPTER 3
STATISTICS AND TIME SERIES

A time series is a collection of records indexed by time.

$$\{y_t, t = 1, 2 \ldots . T\} = \{y_1, y_2, \ldots . . y_T\}, \text{ where } T \text{ is the number of periods}$$

Figure 3.1 Examples of Time Series



These slides are based on:
González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.
Slides adapted for this course. We thank Gloria González-Rivera and assume full responsibility for all errors due to our changes which are mainly in red

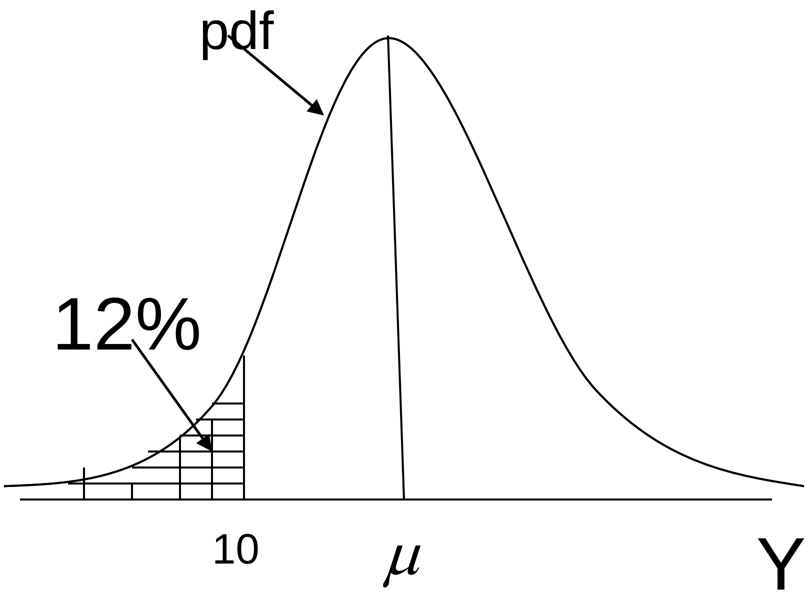
3.1 Stochastic Process and Time Series

Random variables can be characterized in two ways:

- probability density/mass function (full characterization)
- moments (partial characterization)

$$\bar{y} = \sum_{t=1}^{T} \frac{y_t}{T}$$

$$\overline{\sigma^2} = \sum_{t=1}^{T} \frac{(y_t - \bar{y})^2}{T}$$



$$P(Y \leq 10) = 12\%$$

Figure 3.2 Probability Density Function

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.1.1 Stochastic Process

A stochastic process is a collection of random variables indexed by time.

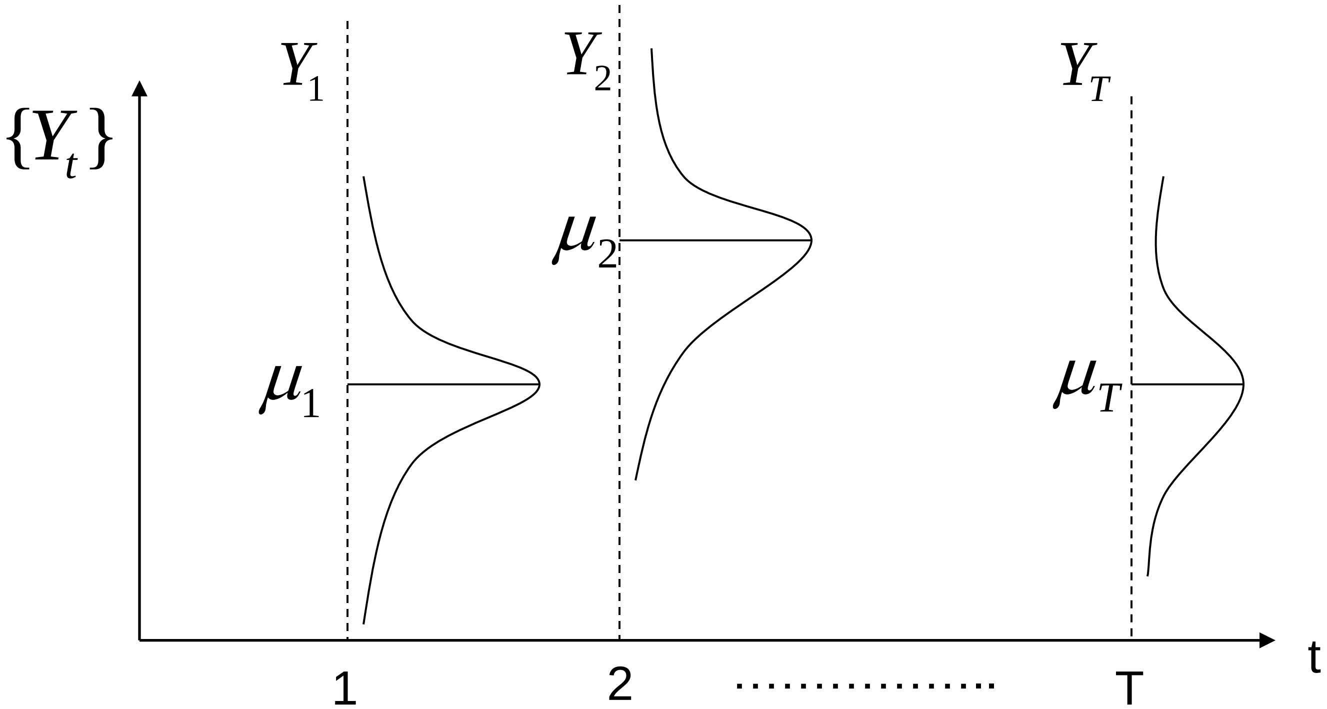


Figure 3.3 Graphical Representation of a Stochastic Process

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.1.1 Time Series

A time series is a sample realization of a stochastic process.

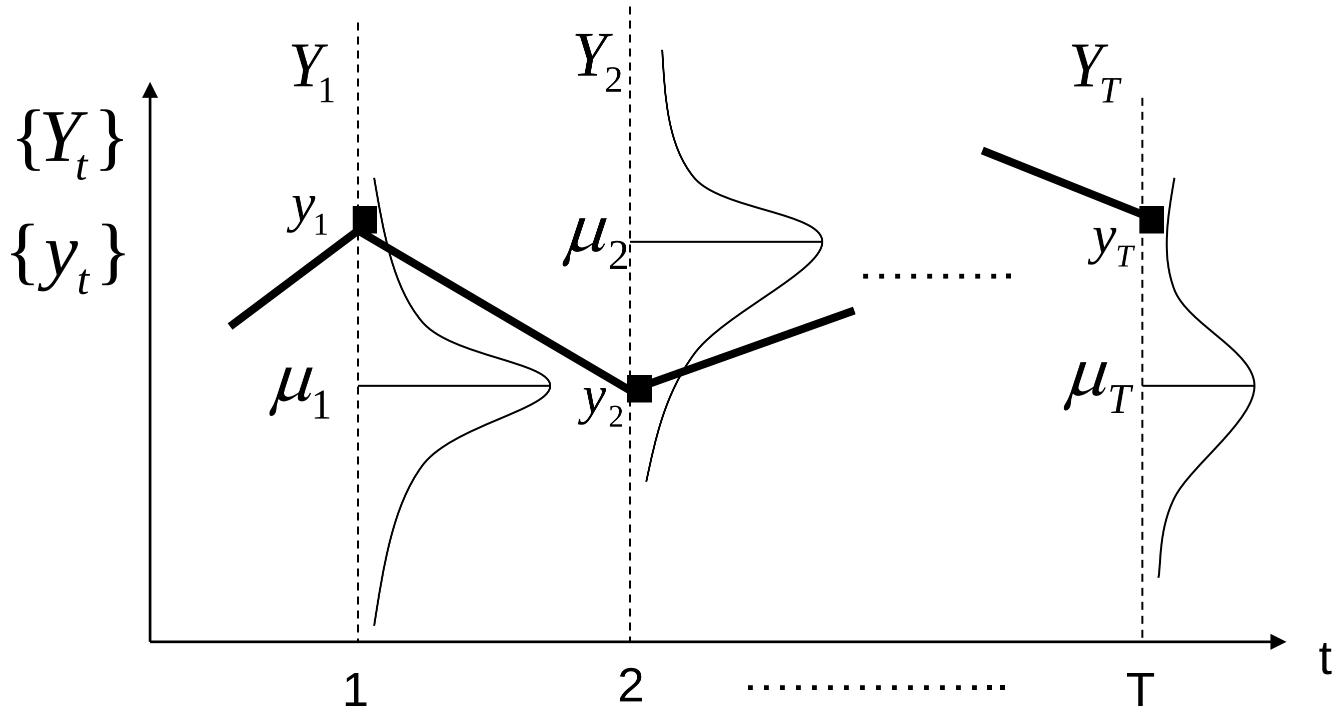
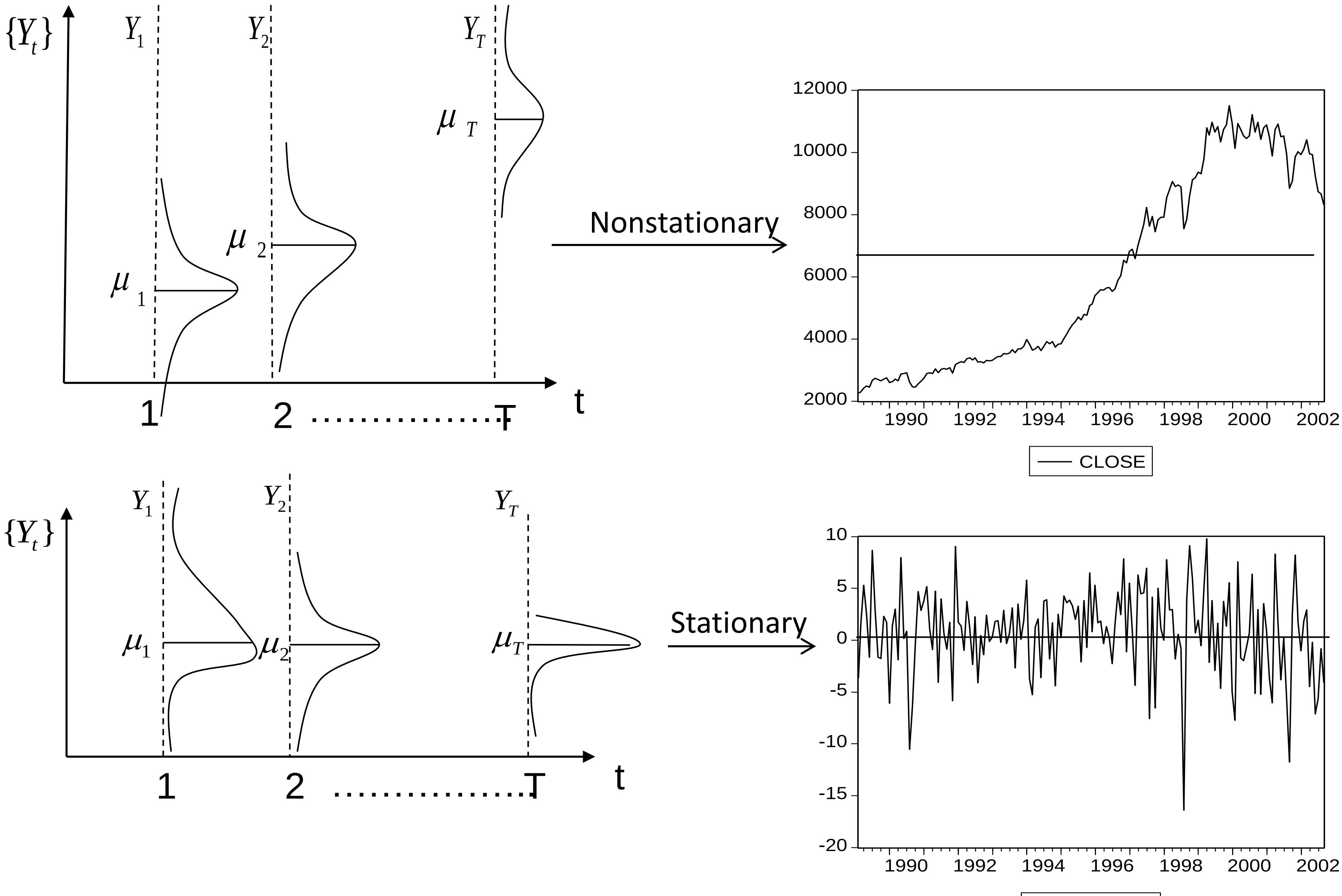


Figure 3.4 Graphical Representation of a Stochastic Process and a Time Series (Thick Line)

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.2.1 Stationarity

Figure 3.5 Nonstationary and Stationary Stochastic Process



González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

A stochastic process is said to be **first order strongly stationary** if all random variables have the **same probability mass/density function**.

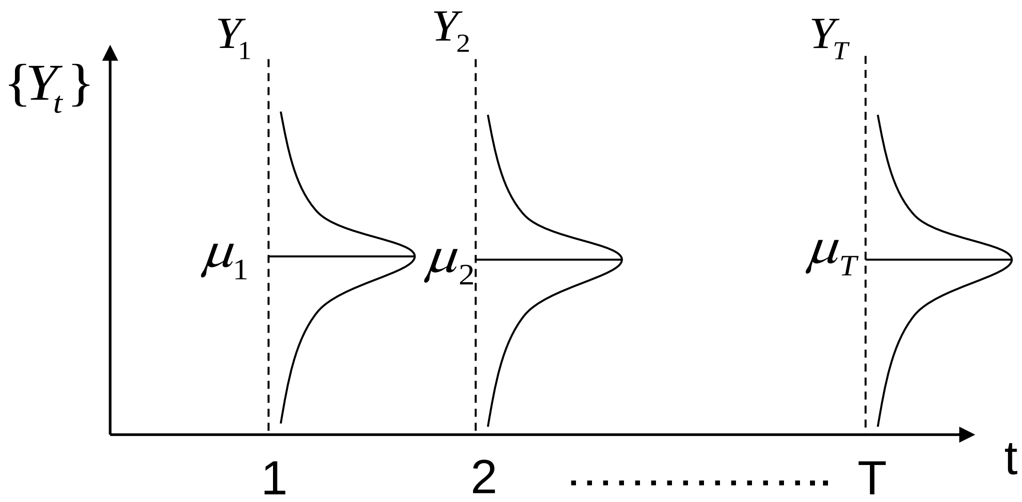


Figure 3.6 Strongly Stationary Stochastic Process

A stochastic process is said to be **first order weakly stationary** if all random variables have the same mean.

A stochastic process is said to be **second order weakly stationary** (or **covariance stationary**) if all random variables have the **same mean and the same variance and covariances do not depend on time, only on lag**.

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.2.2 Useful transformations of Nonstationary Processes

To transform a nonstationary series into a first-moment-stationary series, we can apply first differences :

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$$

Lag operator

To stabilize the variance we can use the Box-Cox transformation:
(before taking differences)

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0, \end{cases}$$

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

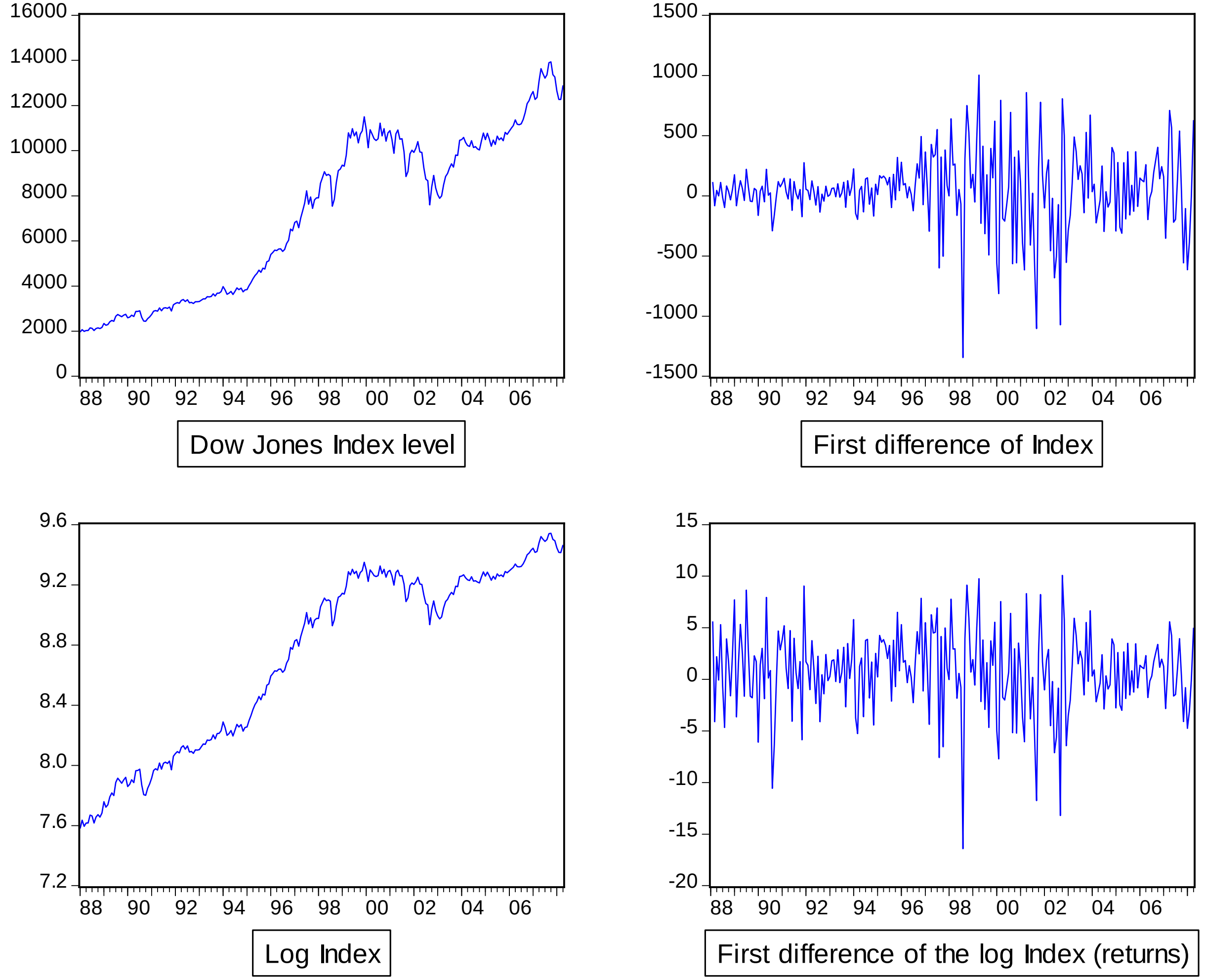
3.2.2 Useful transformations of Nonstationary Processes

Table 3.1 Dow Jones Index and Returns

Date	y_t	y_{t-1}	Δy_t	$R_t = \frac{\Delta y_t}{y_{t-1}} \times 100$	$R_t \approx \Delta\log(y_t) \times 100$
2001:01	10887.400	10788.000	99.40000	0.921394	0.917175
2001:02	10495.300	10887.400	−392.10000	−3.601411	−3.667862
2001:03	9878.800	10495.300	−616.50000	−5.874058	−6.053649
2001:04	10735.000	9878.800	856.20000	8.667045	8.311838
2001:05	10911.900	10735.000	176.90000	1.647881	1.634451
2001:06	10502.400	10911.900	−409.50000	−3.752784	−3.825013
2001:07	10522.800	10502.400	20.40000	0.194241	0.194053
2001:08	9949.800	10522.800	−573.00000	−5.445319	−5.599188
2001:09	8847.600	9949.800	−1102.20000	−11.077610	−11.740620
2001:10	9075.100	8847.600	227.50000	2.571319	2.538816
2001:11	9851.600	9075.100	776.50000	8.556380	8.209948
2001:12	10021.600	9851.600	170.00000	1.725608	1.710889
2002:01	9920.000	10021.600	−101.60000	−1.013810	−1.018984
2002:02	10106.100	9920.000	186.10000	1.876008	1.858628
2002:03	10403.900	10106.100	297.80000	2.946735	2.904153
2002:04	9946.200	10403.900	−457.70000	−4.399312	−4.499017
2002:05	9925.300	9946.200	−20.90000	−0.210131	−0.210352
2002:06	9243.300	9925.300	−682.00000	−6.871329	−7.118809
2002:07	8736.600	9243.300	−506.70000	−5.481808	−5.637787
2002:08	8663.500	8736.600	−73.10000	−0.836710	−0.840230
2002:09	8312.690	8663.500	−350.81000	−4.049287	−4.133554

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

Figure 3.7 Dow Jones Index and Its Transformation to Returns



González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.3 A New Tool of Analysis: The Autocorrelation Functions

Given two random variables *Y* and *X*, the correlation coefficient is a measure of the linear association.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \qquad -1 \leq \rho_{X,Y} \leq 1$$

Autocorrelation coefficient:

$$\rho_{Y_t,Y_{t-k}} = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-k})}}$$

For second order stationary processes :

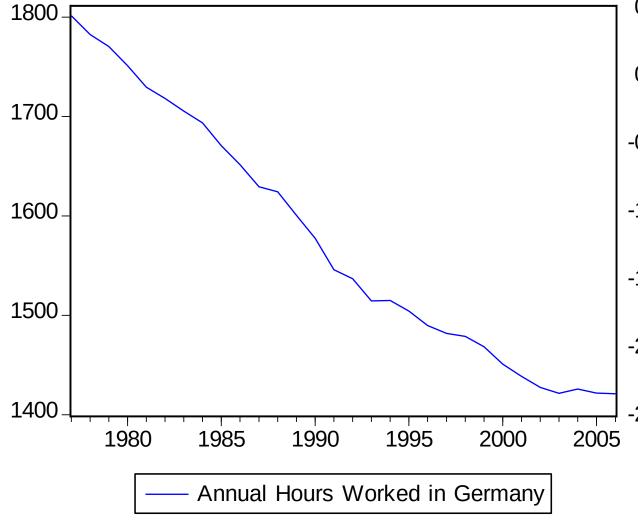
$$\rho_{Y_t,Y_{t-k}} = \rho_k = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} = \frac{\gamma_k}{\gamma_0} \qquad \rho_k = \rho_{-k} = \rho_{|k|}$$

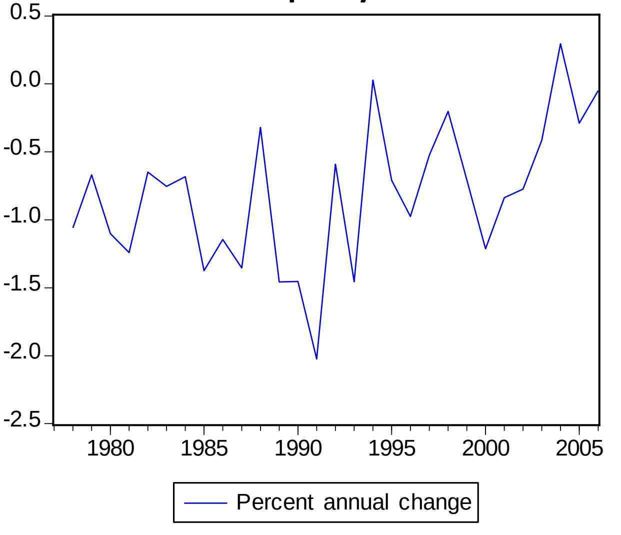
The **autocorrelation function** (ACF) is the function $\rho: k \rightarrow \rho_k$

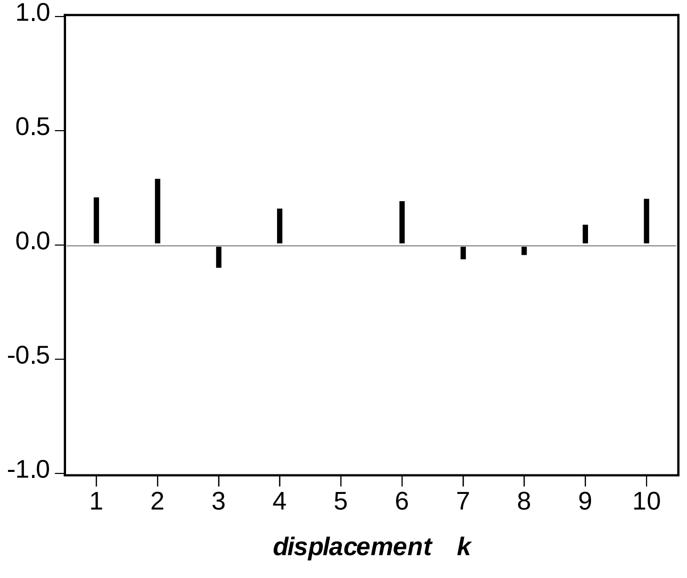
González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.3 A New Tool of Analysis: The Autocorrelation Functions

Figure 3.8 Annual Hours Worked per Person Employed in Germany







Autocorrelation function										
k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.22	.29	-.10	.16	-.01	.19	-.06	-.04	.09	.20

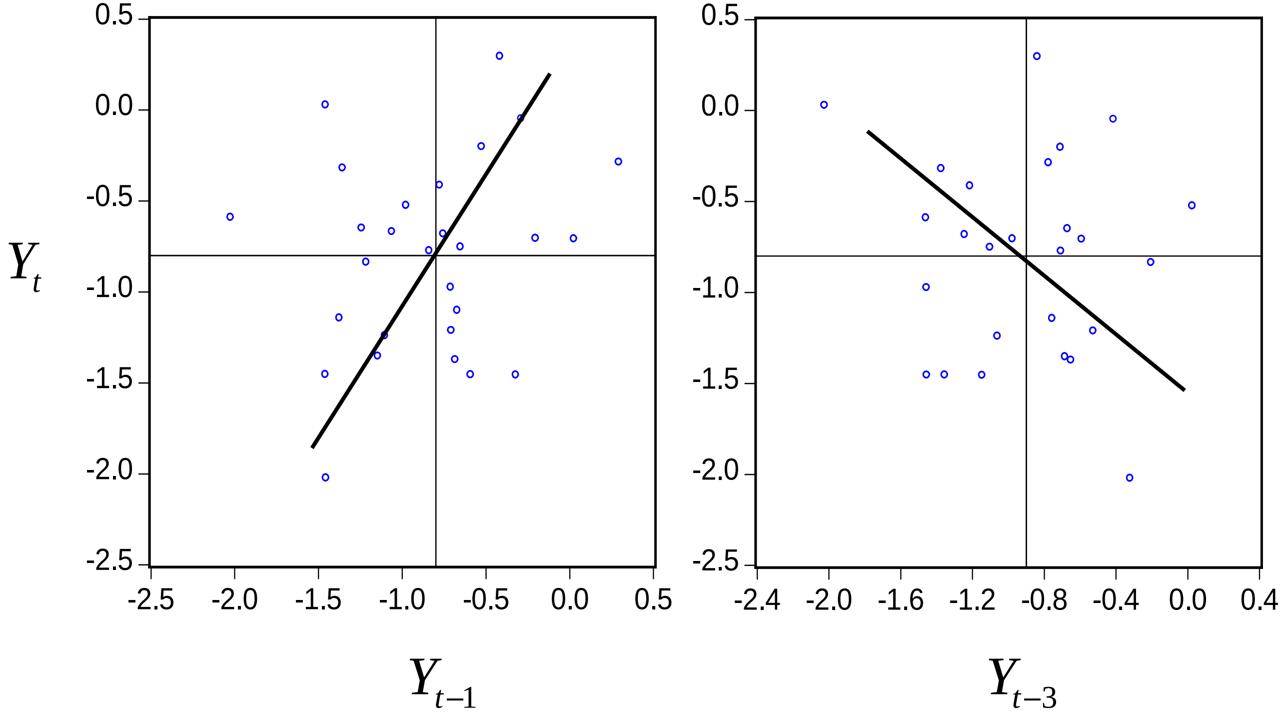
González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

Table 3.2 Percentage Change in Working Hours in Germany: Calculation of the Autocorrelation Coefficients

	Y_t	Y_{t-1}	Y_{t-3}
1978	-1.0604		
1979	-0.6699	-1.0604	
1980	-1.1018	-0.6699	
1981	-1.2413	-1.1018	-1.0604
1982	-0.6497	-1.2413	-0.6699
1983	-0.7536	-0.6497	-1.1018
1984	-0.6826	-0.7536	-1.2413
1985	-1.3733	-0.6826	-0.6497
1986	-1.1438	-1.3733	-0.7536
1987	-1.3533	-1.1438	-0.6826
1988	-0.3196	-1.3533	-1.3733
1989	-1.4574	-0.3196	-1.1438
1990	-1.4536	-1.4574	-1.3533
1991	-2.0234	-1.4536	-0.3196
1992	-0.5904	-2.0234	-1.4574
1993	-1.4550	-0.5904	-1.4536
1994	0.0264	-1.4550	-2.0234
1995	-0.7087	0.0264	-0.5904
1996	-0.9752	-0.7087	-1.4550
1997	-0.5249	-0.9752	0.0264
1998	-0.2026	-0.5249	-0.7087
1999	-0.7057	-0.2026	-0.9752
2000	-1.2126	-0.7057	-0.5249
2001	-0.8375	-1.2126	-0.2026
2002	-0.7745	-0.8375	-0.7057
2003	-0.4141	-0.7745	-1.2126
2004	0.2950	-0.4141	-0.8375
2005	-0.2879	0.2950	-0.7745
2006	-0.0492	-0.2879	-0.4141
Mean: $\hat{\mu}$	-0.8026		
Variance: $\hat{\gamma}_0$	0.2905		
$\hat{\gamma}_k$ $(k=1,3)$		0.0651	-0.0282
$\hat{\rho}_k$ $(k=1,3)$		0.2240	-0.0970

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

Figure 3.9 Percentage Change in Working Hours in Germany: Autocorrelations of Order 1 and 3



González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

Autocorrelation between Y_t and Y_{t+k} controlling for the effect of

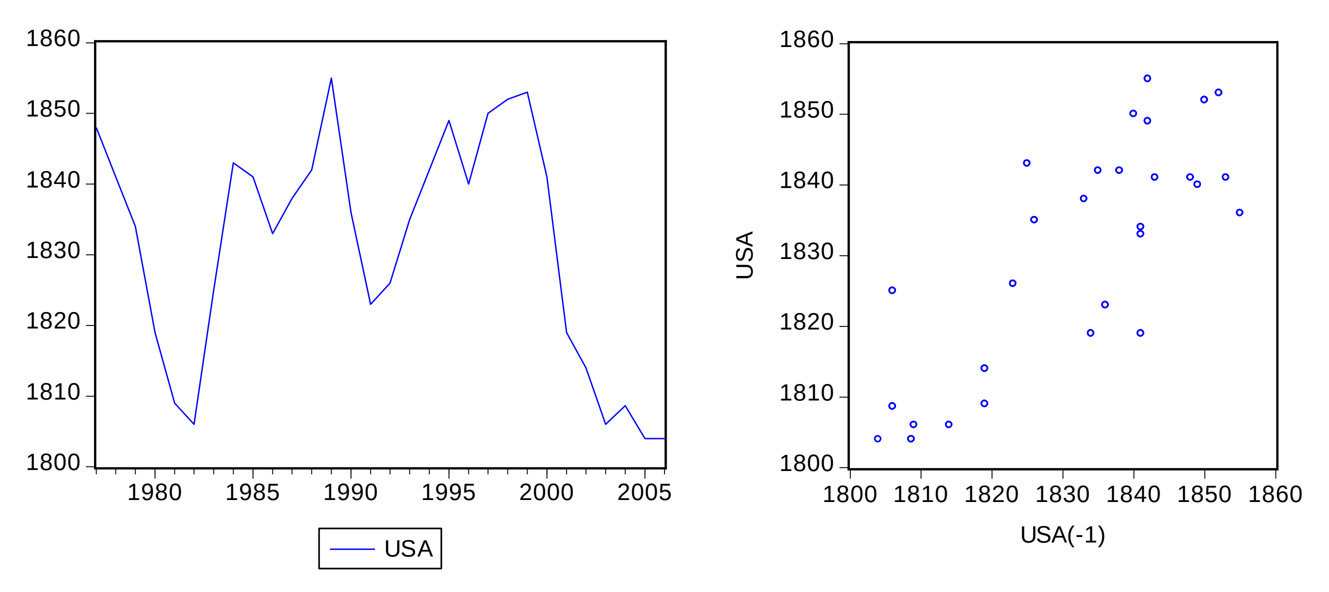
$Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$

Distance k	Regression	Partial autocorrelation coefficient r_k
1	$Y_{t+k} = \beta_0 + \beta_1 Y_{t+k-1} + \varepsilon_{t+k}$	β_1
2	$Y_{t+k} = \beta_0 + \beta_1 Y_{t+k-1} + \beta_2 Y_{t+k-2} + \varepsilon_{t+k}$	β_2
3	$Y_{t+k} = \beta_0 + \beta_1 Y_{t+k-1} + \beta_2 Y_{t+k-2} + \beta_3 Y_{t+k-3} + \varepsilon_{t+k}$	β_3
.....	....	......
k	$Y_{t+k} = \beta_0 + \beta_1 Y_{t+k-1} + \beta_2 Y_{t+k-2} + \cdots\cdots \beta_{k-1} Y_{t+1} + \beta_k Y_t + \varepsilon_{t+k}$	β_k

The partial **autocorrelation function** (PACF) is the function: $r: k \to r_k$

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

Figure 3.10 Annual Working Hours per Employee in the United States



Autocorrelation Function									
κ	1	2	3	4	5	6	7	8	9
$\hat{\rho}_k$	.74	.36	.06	−.09	−.16	−.29	−.35	−.25	−.06

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.

3.3.2 Statistical Tests for Autocorrelation Coefficients

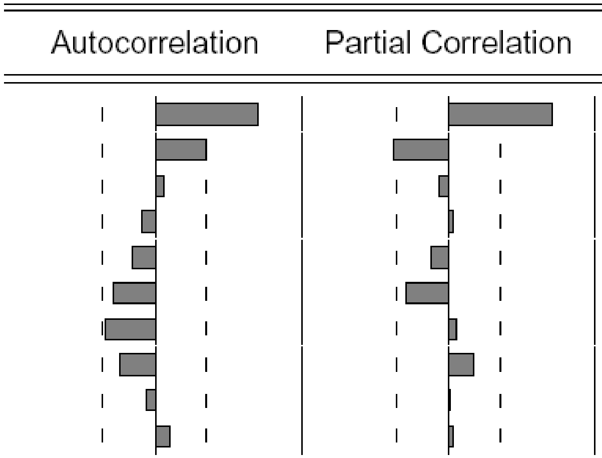
$$H_0: \rho_k = 0$$

$$\hat{\rho}_k \rightarrow N(0, 1/T)$$

$$H_0: \rho_1 = \rho_2 = \cdots \rho_k = 0$$

$$Q_k = T(T+2)\sum_{j=1}^{k} \frac{\hat{\rho}_k^{\,2}}{T-j} \rightarrow \chi^2(k)$$

Sample: 1977 2006
Included observations: 30



	AC	PAC	Q-Stat	Prob
1	0.737	0.737	17.987	0.000
2	0.364	-0.392	22.537	0.000
3	0.062	-0.058	22.676	0.000
4	-0.086	0.039	22.951	0.000
5	-0.162	-0.126	23.957	0.000
6	-0.288	-0.295	27.270	0.000
7	-0.352	0.052	32.432	0.000
8	-0.253	0.185	35.229	0.000
9	-0.064	0.001	35.416	0.000
10	0.114	0.034	36.035	0.000

Figure 3.11 Time Series: Annual Working Hours per Employee in the United States. Autocorrelation Function

González-Rivera: Forecasting for Economics and Business, Copyright © 2013 Pearson Education, Inc.