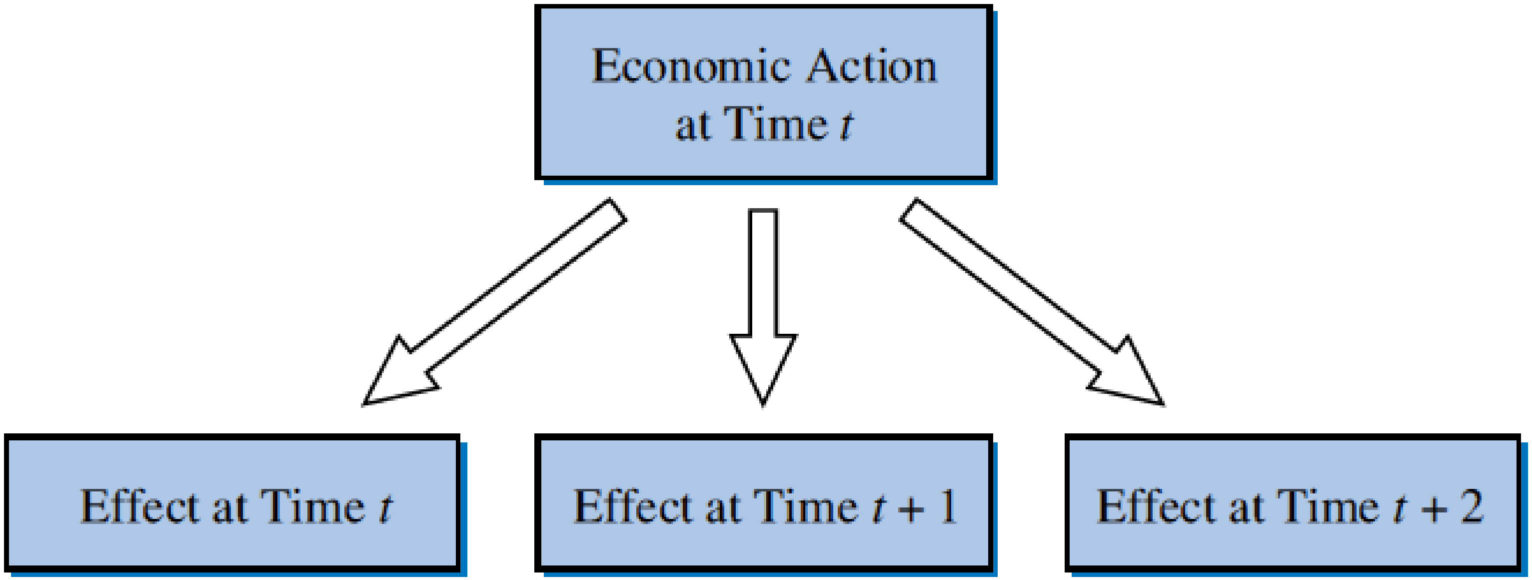
Lecture 1
Dynamic Nature of Relationships, Serial Correlation

9.1 Introduction

- When modeling relationships between variables, the nature of the data that have been collected has an important bearing on the appropriate choice of an econometric model
  - Two features of time-series data to consider:
    1. Time-series observations on a given economic unit, observed over a number of time periods, are likely to be correlated
    2. Time-series data have a natural ordering according to time

- There is also the possible existence of dynamic relationships between variables
  - A dynamic relationship is one in which the change in a variable now has an impact on that same variable, or other variables, in one or more future time periods
  - These effects do not occur instantaneously but are spread, or **distributed**, over future time periods

FIGURE 9.1 The distributed lag effect



- Ways to model the dynamic relationship:
  1. Specify that a dependent variable *y* is a function of current and past values of an explanatory variable *x*

Eq. 9.1

$$y_t = f(x_t, x_{t-1}, x_{t-2}, \ldots)$$

  - Because of the existence of these lagged effects, Eq. 9.1 is called a distributed lag model

- Ways to model the dynamic relationship (Continued):

2. Capturing the dynamic characteristics of time-series by specifying a model with a lagged dependent variable as one of the explanatory variables

Eq. 9.2

$$y_t = f(y_{t-1}, x_t)$$

- Or have:

Eq. 9.3

$$y_t = f(y_{t-1}, x_t, x_{t-1}, x_{t-2})$$

– Such models are called **autoregressive distributed lag** (**ARDL**) models, with "autoregressive" meaning a regression of y_t on its own lag or lags

- Ways to model the dynamic relationship (Continued):

3. Model the continuing impact of change over several periods via the error term

Eq. 9.4

$$y_t = f(x_t) + e_t \qquad e_t = f(e_{t-1})$$

- In this case e_t is correlated with e_{t-1}
- We say the errors are **serially correlated** or **autocorrelated**

- The primary assumption is Assumption MR4:

$$\operatorname{cov}(y_i, y_j) = \operatorname{cov}(e_i, e_j) = 0 \quad \text{for } i \neq j$$

  - For time series, this is written as:

$$\operatorname{cov}(y_t, y_s) = \operatorname{cov}(e_t, e_s) = 0 \quad \text{for } t \neq s$$

    – The dynamic models in Eqs. 9.2, 9.3 and 9.4 imply correlation between y_t and y_{t-1} or e_t and e_{t-1} or both, so they clearly violate assumption MR4

- A stationary variable is one that is not explosive, nor trending, and nor wandering aimlessly without returning to its mean

FIGURE 9.2 (a) Time series of a stationary variable

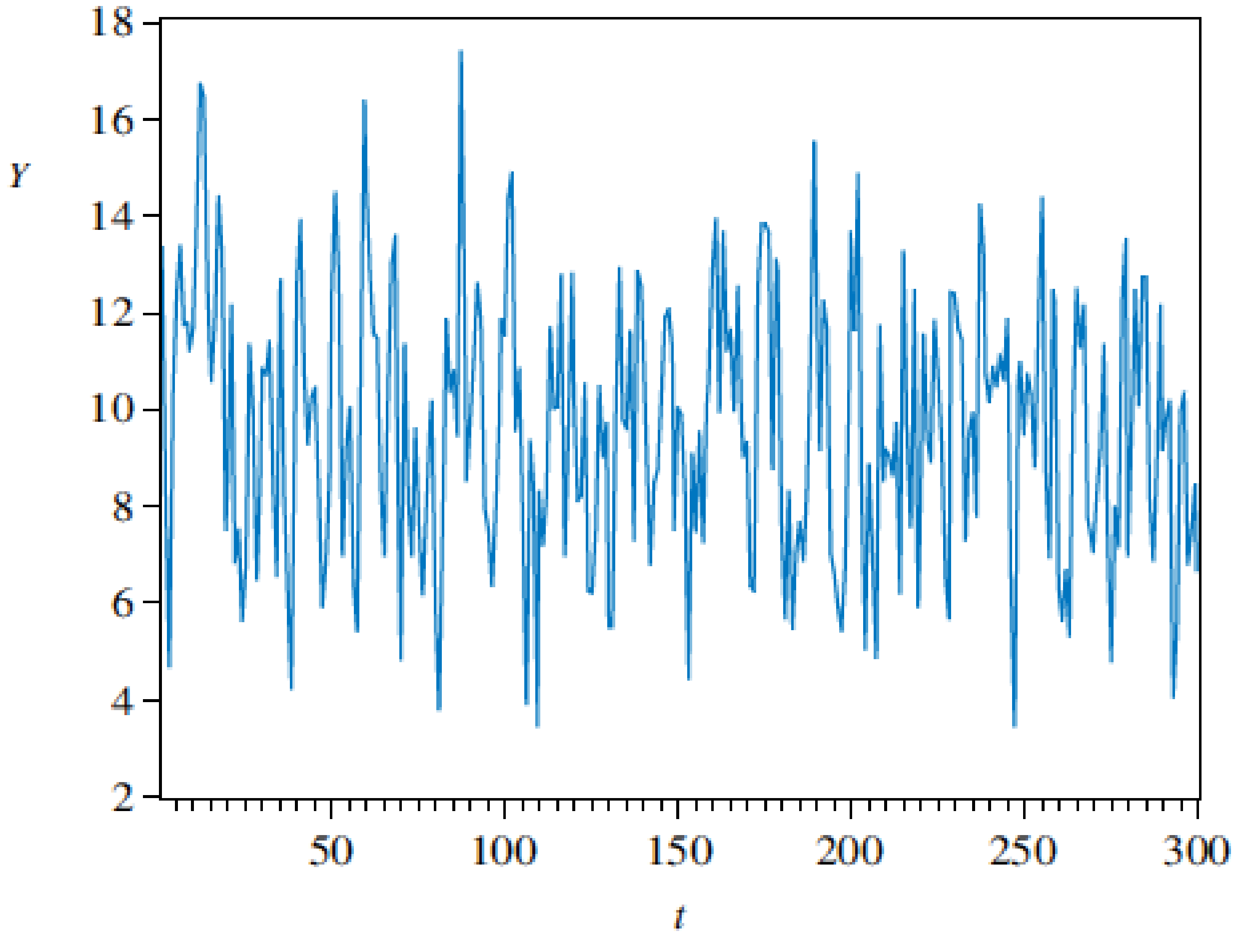


FIGURE 9.2 (b) time series of a nonstationary variable that is "slow-turning" or "wandering"

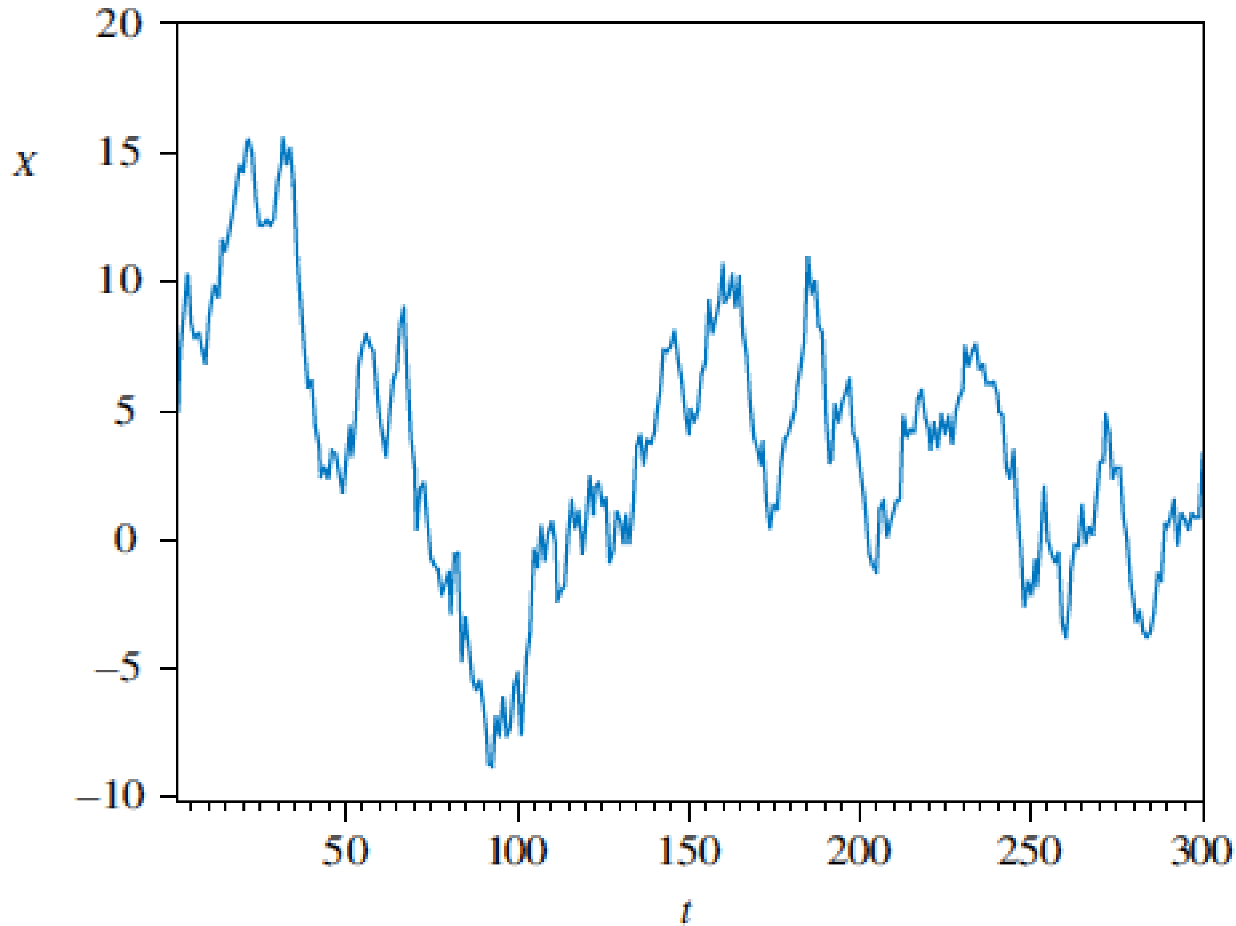
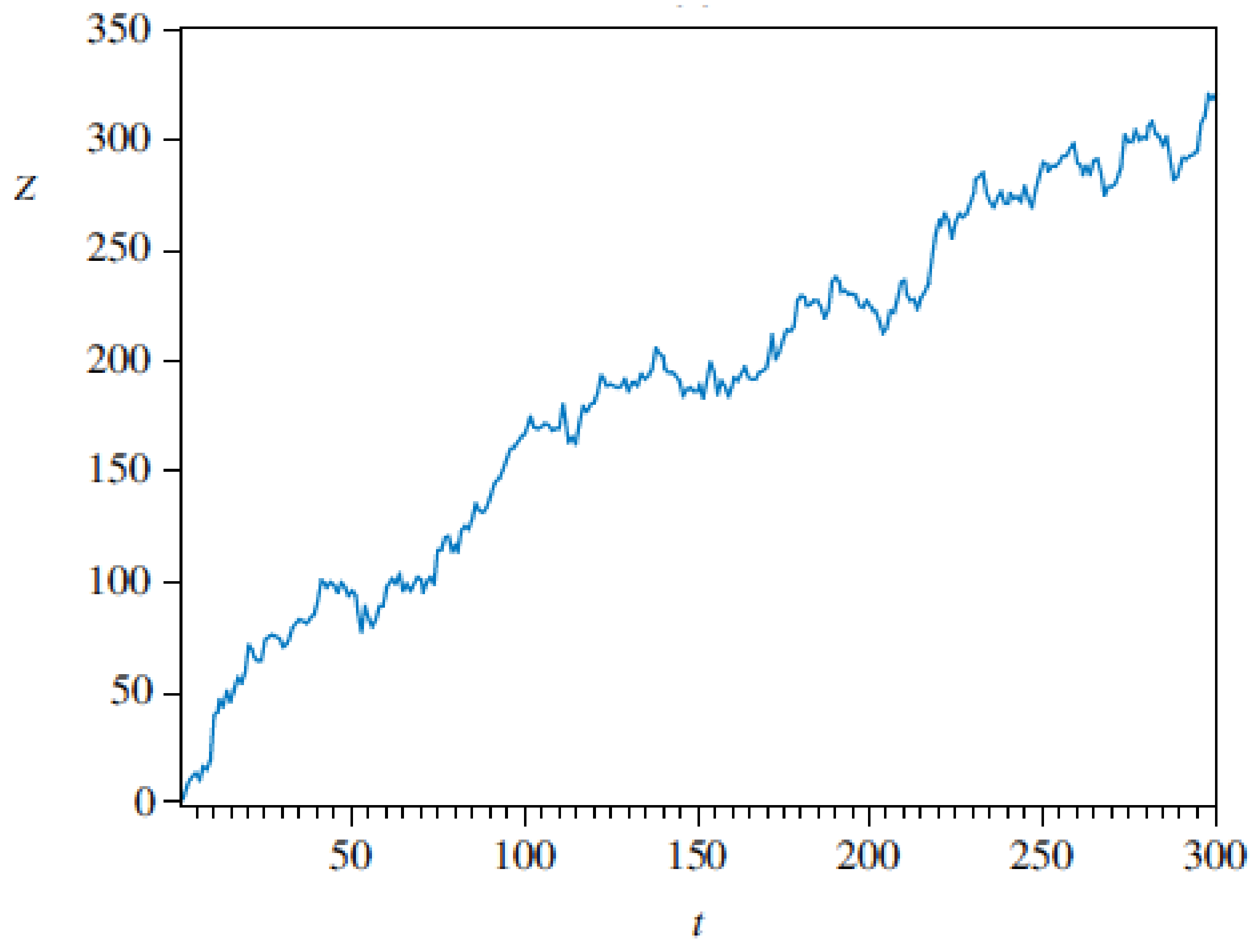


FIGURE 9.2 (c) time series of a nonstationary variable that "trends"



9.2
Finite Distributed Lags

- Consider a linear model in which, after q time periods, changes in x no longer have an impact on y

Eq. 9.5

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t$$

  – Note the notation change: β_s is used to denote the coefficient of x_{t-s} and α is introduced to denote the intercept

- Model 9.5 has two uses:
  - Forecasting

Eq. 9.6 $$y_{T+1} = \alpha + \beta_0 x_{T+1} + \beta_1 x_T + \beta_2 x_{T-1} + \cdots + \beta_q x_{T-q+1} + e_{T+1}$$

  - Policy analysis
    - What is the effect of a change in *x* on *y*?

Eq. 9.7 $$\frac{\partial E(y_t)}{\partial x_{t-s}} = \frac{\partial E(y_{t+s})}{\partial x_t} = \beta_s$$

- Assume x_t is increased by one unit and then maintained at its new level in subsequent periods
  - The immediate impact will be β_0
  - the total effect in period $t + 1$ will be $\beta_0 + \beta_1$, in period $t + 2$ it will be $\beta_0 + \beta_1 + \beta_2$, and so on
    - These quantities are called **interim multipliers**
  - The **total multiplier** is the final effect on y of the sustained increase after q or more periods have elapsed $\sum_{s=0}^{q} \beta_s$

- The effect of a one-unit change in x_t is **distributed** over the current and next q periods, from which we get the term “distributed lag model”
  - It is called a **finite distributed lag model of order *q***
    - It is assumed that after a finite number of periods q, changes in x no longer have an impact on y
  - The coefficient β_s is called a **distributed-lag weight** or an ***s*-period delay multiplier**
  - The coefficient β_0 ($s = 0$) is called the **impact multiplier**

ASSUMPTIONS OF THE DISTRIBUTED LAG MODEL

TSMR1. $y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t, \qquad t = q+1, \ldots, T$

TSMR2. y and x are stationary random variables, and e_t is independent of current, past and future values of x.

TSMR3. $E(e_t) = 0$

TSMR4. $\text{var}(e_t) = \sigma^2$

TSMR5. $\text{cov}(e_t, e_s) = 0 \quad t \neq s$

TSMR6. $e_t \sim N(0, \sigma^2)$

9.2.2
An Example:
Okun's Law

- Consider Okun's Law
  - In this model the change in the unemployment rate from one period to the next depends on the rate of growth of output in the economy:

Eq. 9.8

$$U_t - U_{t-1} = -\gamma(G_t - G_N)$$

  - We can rewrite this as:

Eq. 9.9

$$DU_t = \alpha + \beta_0 G_t + e_t$$

where $DU = \Delta U = U_t - U_{t-1}$, $\beta_0 = -\gamma$, and $\alpha = \gamma G_N$

- We can expand this to include lags:

Eq. 9.10

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \cdots + \beta_q G_{t-q} + e_t$$

- We can calculate the growth in output, *G*, as:

Eq. 9.11

$$G_t = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}} \times 100$$

FIGURE 9.4 (a) Time series for the change in the U.S. unemployment rate: 1985Q3 to 2009Q3

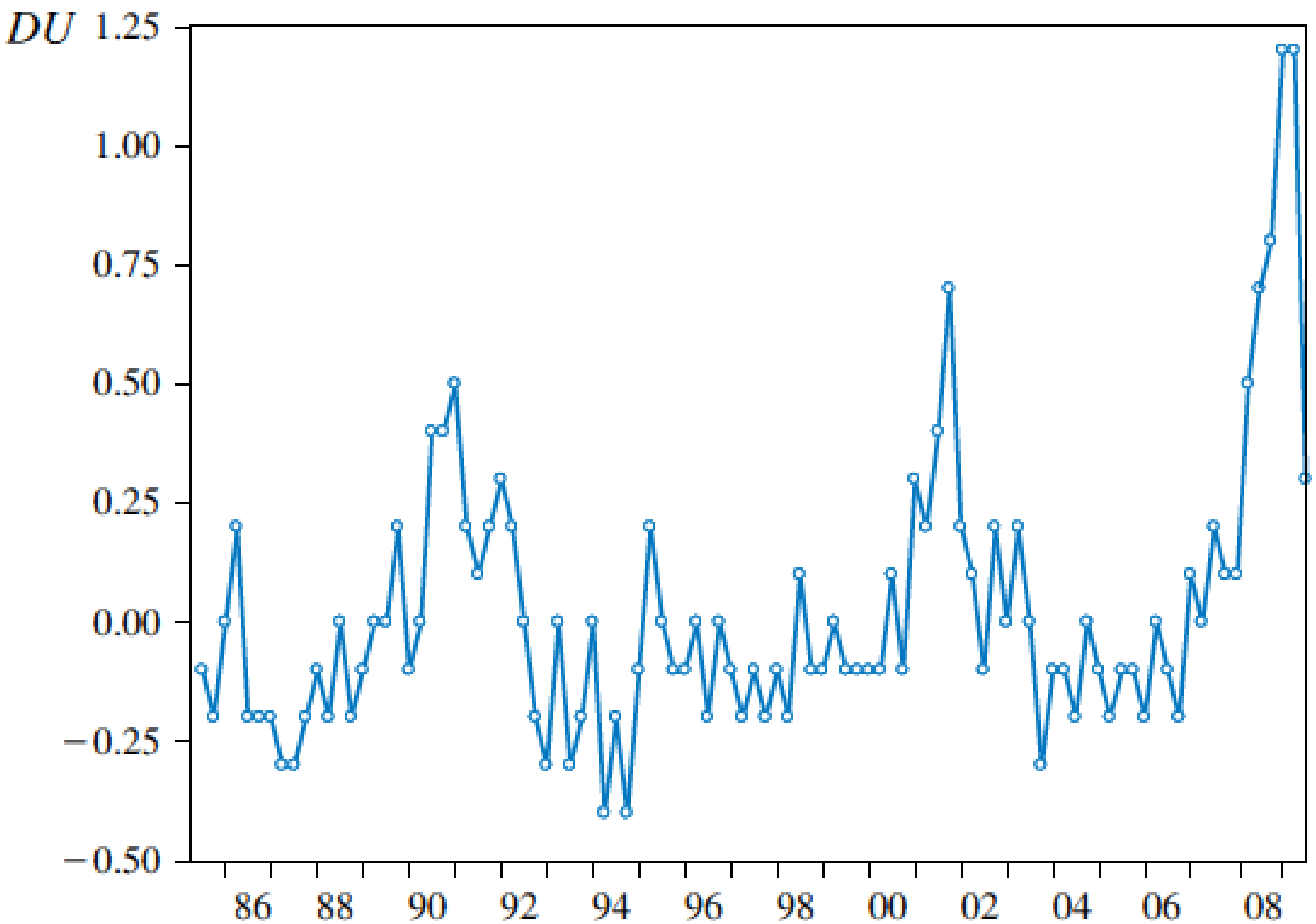


FIGURE 9.4 (b) Time series for U.S. GDP growth: 1985Q2 to 2009Q3

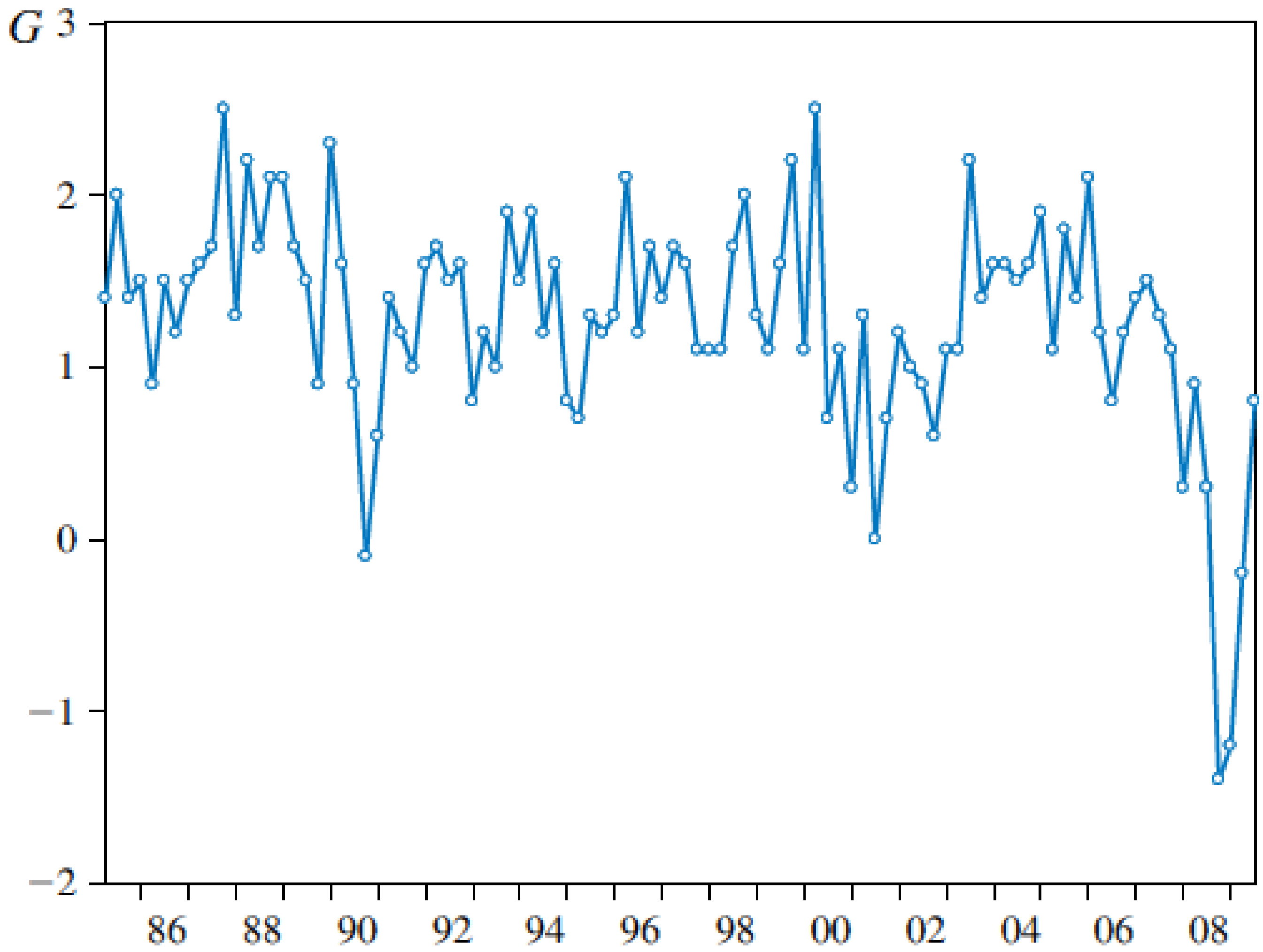


Table 9.1 Spreadsheet of Observations for Distributed Lag Model

t	Quarter	U_t	U_{t-1}	DU_t	G_t	G_{t-1}	G_{t-2}	G_{t-3}
1	1985Q2	7.3	•	•	1.4	•	•	•
2	1985Q3	7.2	7.3	−0.1	2.0	1.4	•	•
3	1985Q4	7.0	7.2	−0.2	1.4	2.0	1.4	•
4	1986Q1	7.0	7.0	0.0	1.5	1.4	2.0	1.4
5	1986Q2	7.2	7.0	0.2	0.9	1.5	1.4	2.0
96	2009Q1	8.1	6.9	1.2	−1.2	−1.4	0.3	0.9
97	2009Q2	9.3	8.1	1.2	−0.2	−1.2	−1.4	0.3
98	2009Q3	9.6	9.3	0.3	0.8	−0.2	−1.2	−1.4

Table 9.2 Estimates for Okun's Law Finite Distributed Lag Model

Lag Length $q = 3$

Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5810	0.0539	10.781	0.0000
G_t	−0.2021	0.0330	6.120	0.0000
G_{t-1}	−0.1645	0.0358	−4.549	0.0000
G_{t-2}	−0.0716	0.0353	−2.027	0.0456
G_{t-3}	0.0033	0.0363	0.091	0.9276

Observations = 95 $R^2 = 0.652$ $\hat{\sigma} = 0.1743$

Lag Length $q = 2$

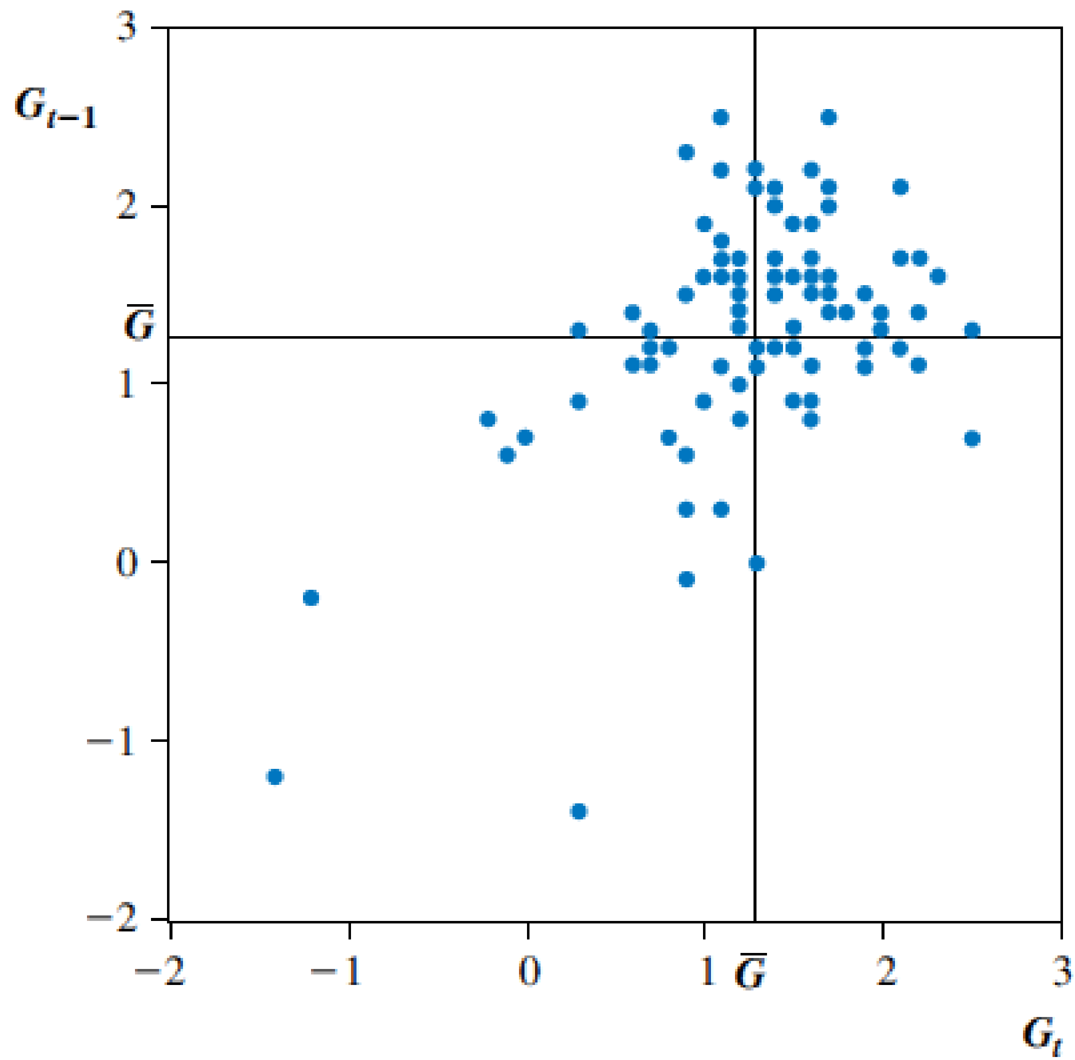
Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5836	0.0472	12.360	0.0000
G_t	−0.2020	0.0324	−6.238	0.0000
G_{t-1}	−0.1653	0.0335	−4.930	0.0000
G_{t-2}	−0.0700	0.0331	−2.115	0.0371

Observations = 96 $R^2 = 0.654$ $\hat{\sigma} = 0.1726$

9.3
Serial Correlation

- When is assumption *TSMR5*, $\text{cov}(e_t, e_s) = 0$ for $t \neq s$ likely to be violated, and how do we assess its validity?
  - When a variable exhibits correlation over time, we say it is **autocorrelated** or **serially correlated**
    - These terms are used interchangeably

FIGURE 9.5 Scatter diagram for G_t and G_{t-1}



- Recall that the population correlation between two variables *x* and *y* is given by:

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\,\text{var}(y)}}$$

- For the Okun's Law problem, we have:

Eq. 9.12

$$\rho_1 = \frac{\operatorname{cov}(G_t, G_{t-1})}{\sqrt{\operatorname{var}(G_t)\operatorname{var}(G_{t-1})}} = \frac{\operatorname{cov}(G_t, G_{t-1})}{\operatorname{var}(G_t)}$$

- The notation ρ_1 is used to denote the population correlation between observations that are one period apart in time
  - This is known also as the **population autocorrelation of order one**.
  - The second equality in Eq. 9.12 holds because $\operatorname{var}(G_t) = \operatorname{var}(G_{t-1})$, a property of time series that are stationary

- The first-order sample autocorrelation for G is obtained from Eq. 9.12 using the estimates:

$$\widehat{\operatorname{cov}(G_t, G_{t-1})} = \frac{1}{T-1}\sum_{t=2}^{T}(G_t - \bar{G})(G_{t-1} - \bar{G})$$

$$\widehat{\operatorname{var}(G_t)} = \frac{1}{T-1}\sum_{t=1}^{T}(G_t - \bar{G})^2$$

- Making the substitutions, we get:

Eq. 9.13

$$r_1 = \frac{\sum_{t=2}^{T}(G_t - \bar{G})(G_{t-1} - \bar{G})}{\sum_{t=1}^{T}(G_t - \bar{G})^2}$$

- More generally, the ***k*-th order sample autocorrelation** for a series *y* that gives the correlation between observations that are *k* periods apart is:

Eq. 9.14

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

- Applying this to our problem, we get for the first four autocorrelations:

Eq. 9.16

$$r_1 = 0.494 \quad r_2 = 0.411 \quad r_3 = 0.154 \quad r_4 = 0.200$$

- How do we test whether an autocorrelation is significantly different from zero?
  - The null hypothesis is H_0: $\rho_k = 0$
  - A suitable test statistic is:

Eq. 9.17

$$Z = \frac{r_k - 0}{\sqrt{\frac{1}{T}}} = \sqrt{T} r_k \sim N(0,1)$$

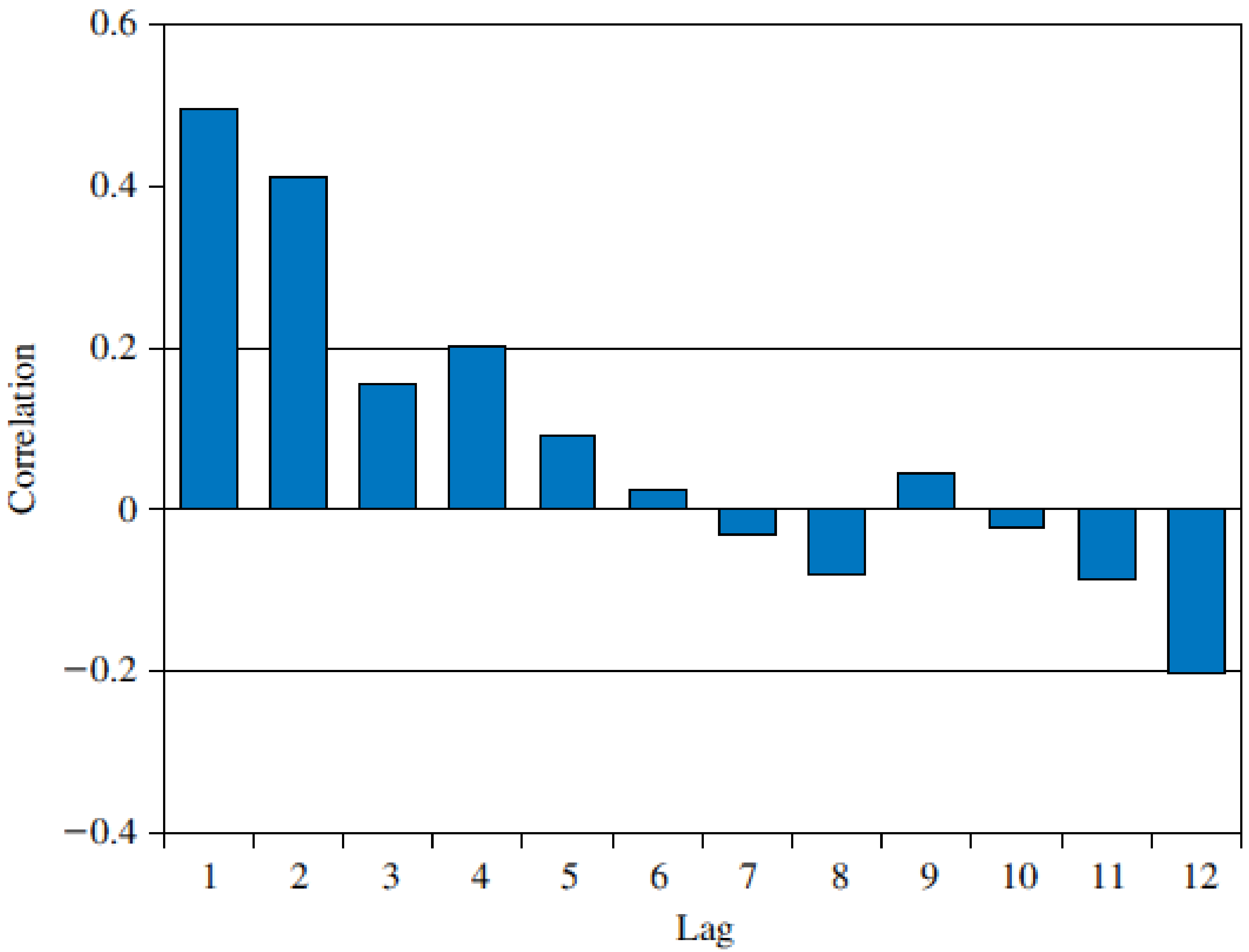
- For our problem, we have:

$$Z_1 = \sqrt{98} \times 0.494 = 4.89, \;\; Z_2 = \sqrt{98} \times 0.414 = 4.10$$
$$Z_3 = \sqrt{98} \times 0.154 = 1.52, \;\; Z_4 = \sqrt{98} \times 0.200 = 1.98$$

  - We reject the hypotheses $H_0\colon \rho_1 = 0$ and $H_0\colon \rho_2 = 0$
  - We have insufficient evidence to reject $H_0\colon \rho_3 = 0$
  - ρ_4 is on the borderline of being significant.
  - We conclude that G, the quarterly growth rate in U.S. GDP, exhibits significant serial correlation at lags one and two

- The **correlogram**, also called the **sample autocorrelation function**, is the sequence of autocorrelations r_1, r_2, r_3, …
  - It shows the correlation between observations that are one period apart, two periods apart, three periods apart, and so on

FIGURE 9.6 Correlogram for *G*



- The correlogram can also be used to check whether the multiple regression assumption $\mathrm{cov}(e_t, e_s) = 0$ for $t \neq s$ is violated

- Consider a model for a Phillips Curve:

Eq. 9.18

$$INF_t = INF_t^E - \gamma(U_t - U_{t-1})$$

– If we initially assume that inflationary expectations are constant over time ($\beta_1 = INF^E{}_t$) set $\beta_2 = -\gamma$, and add an error term:

Eq. 9.19

$$INF_t = \beta_1 + \beta_2 DU_t + e_t$$

FIGURE 9.7 (a) Time series for Australian price inflation

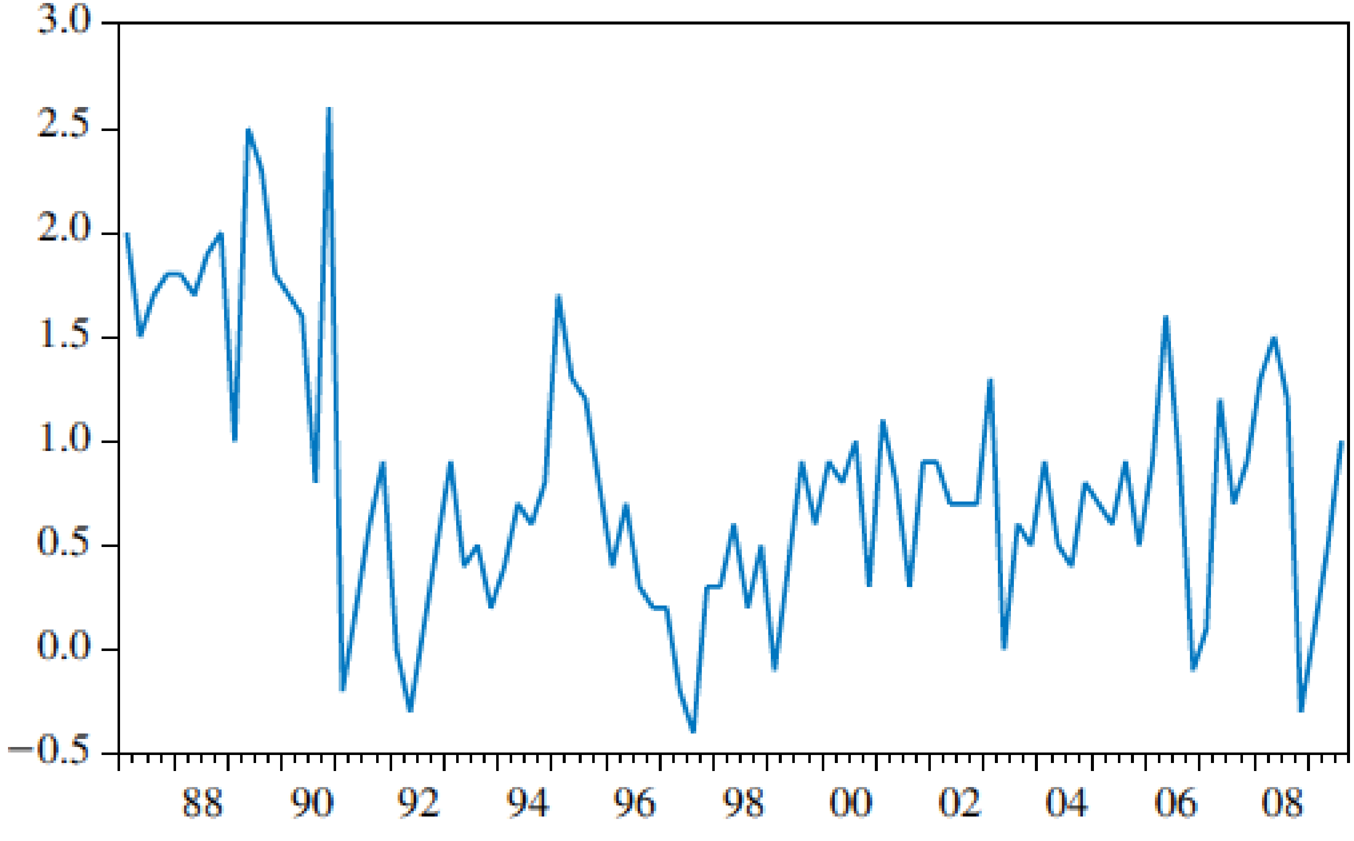
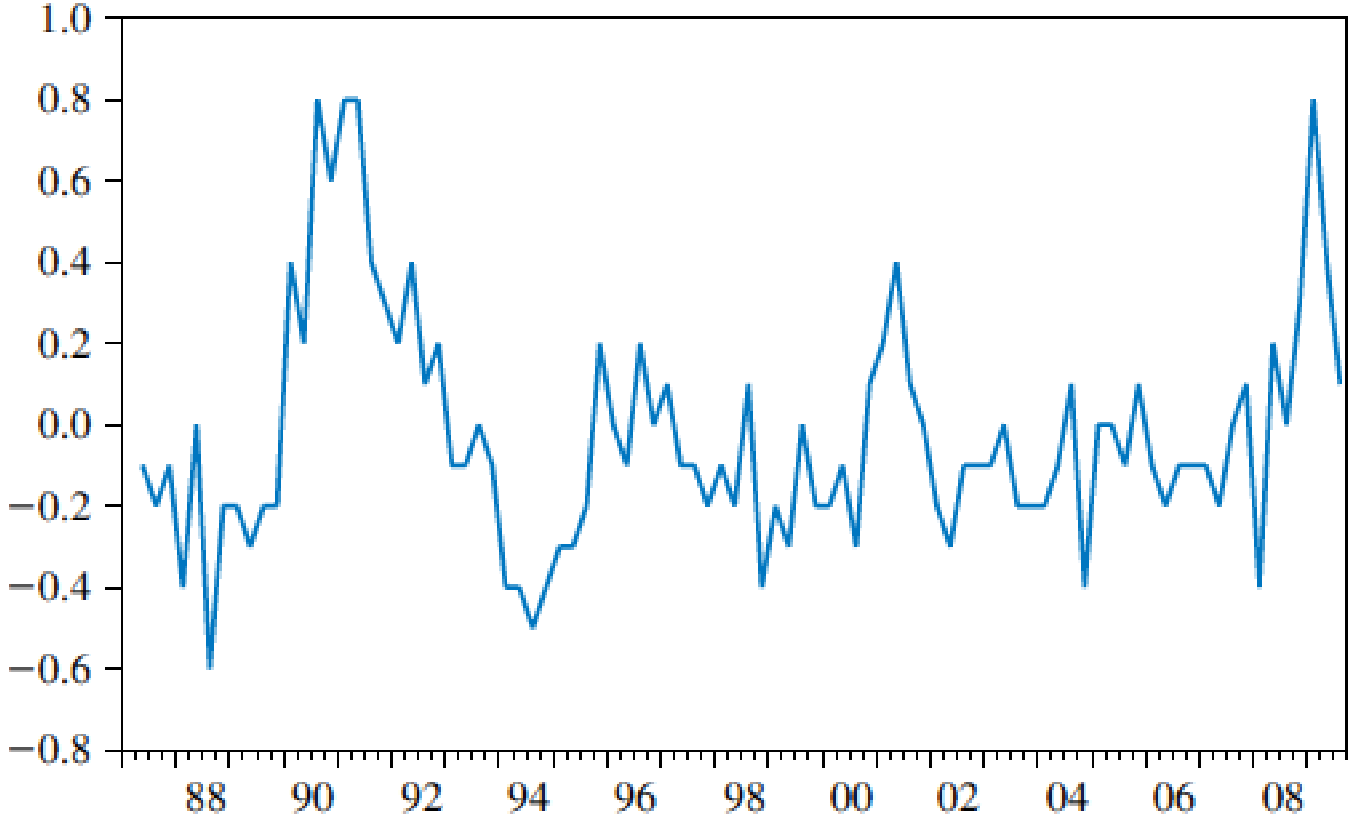


FIGURE 9.7 (b) Time series for the quarterly change in the Australian unemployment rate

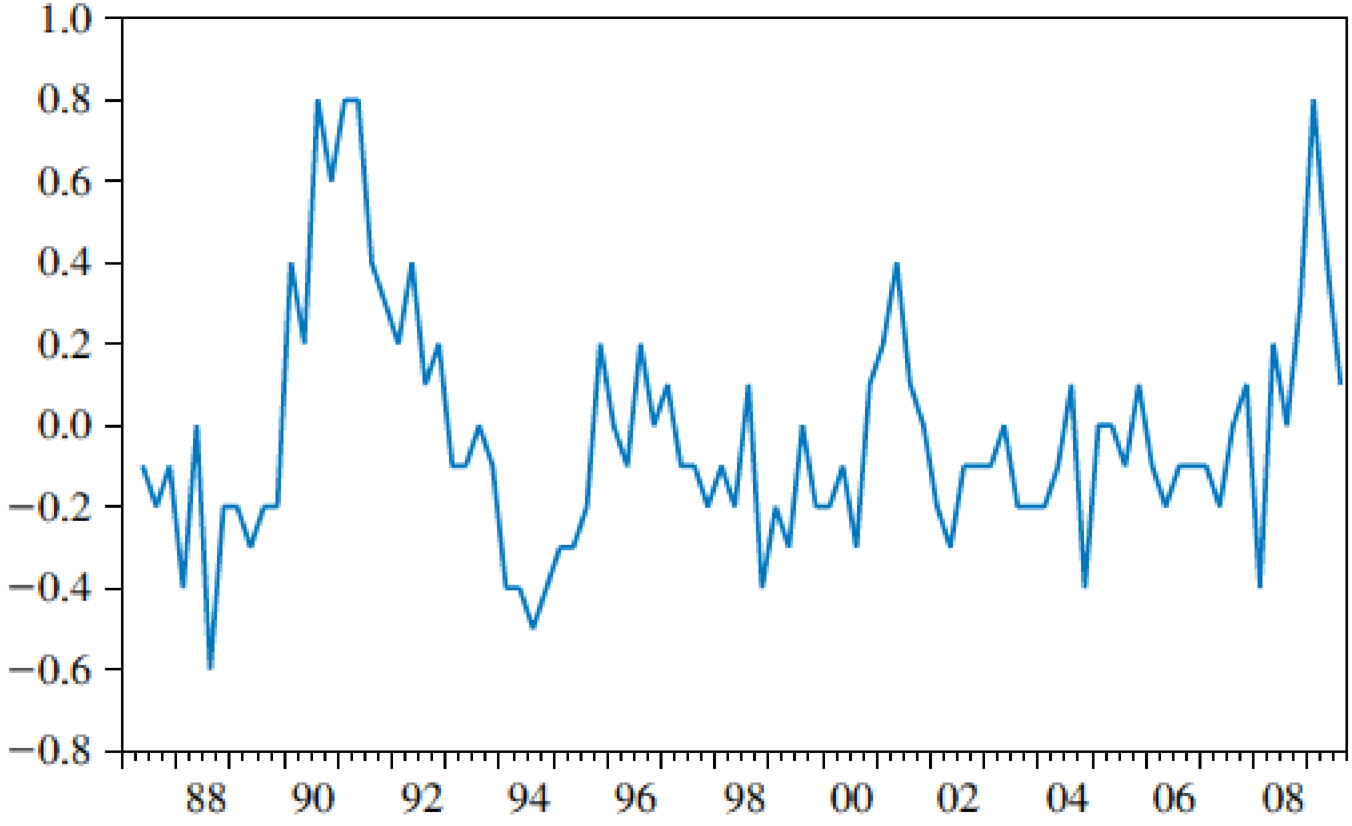


- To determine if the errors are serially correlated, we compute the least squares residuals:

Eq. 9.20

$$\hat{e}_t = INF_t - b_1 - b_2 DU_t$$

FIGURE 9.8 Correlogram for residuals from least-squares estimated Phillips curve



- The k-th order autocorrelation for the residuals can be written as:

Eq. 9.21

$$r_k = \frac{\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^{T} \hat{e}_t^2}$$

  – The least squares equation is:

Eq. 9.22

$$\begin{aligned} \widehat{INF} &= 0.7776 - 0.5279DU \\ (se) \quad &(0.0658) \; (0.2294) \end{aligned}$$

- The values at the first four lags are:

$$r_1 = 0.549 \quad r_2 = 0.456 \quad r_3 = 0.433 \quad r_4 = 0.420$$

9.4

Other Tests for Serially Correlated Errors

- An advantage of this test is that it readily generalizes to a joint test of correlations at more than one lag

- If e_t and e_{t-1} are correlated, then one way to model the relationship between them is to write:

Eq. 9.23

$$e_t = \rho e_{t-1} + v_t$$

  - We can substitute this into a simple regression equation:

Eq. 9.24

$$y_t = \beta_1 + \beta_2 x_t + \rho e_{t-1} + v_t$$

- To derive the relevant auxiliary regression for the autocorrelation LM test, we write the test equation as:

Eq. 9.25

$$y_t = \beta_1 + \beta_2 x_t + \rho \hat{e}_{t-1} + v_t$$

- But since we know that $y_t = b_1 + b_2 x_t + \hat{e}_t$, we get:

$$b_1 + b_2 x_t + \hat{e}_t = \beta_1 + \beta_2 x_t + \rho \hat{e}_{t-1} + v_t$$

Eq. 9.26

- Rearranging, we get:

$$\begin{aligned}\hat{e}_t &= (\beta_1 - b_1) + (\beta_2 - b_2)x_t + \rho\hat{e}_{t-1} + v_t \\ &= \gamma_1 + \gamma_2 x_t + \rho\hat{e}_{t-1} + v\end{aligned}$$

  – If H_0: $\rho = 0$ is true, then LM = T x R^2 has an approximate $\chi^2_{(1)}$ distribution
    - T and R^2 are the sample size and goodness-of-fit statistic, respectively, from least squares estimation of Eq. 9.26

- Then:

$\hat{e}_0$

$$LM = T \times R^2 = 90 \times 0.3066 = 27.59$$

- These value is much larger than 3.84, which is the 5% critical value from a $\chi^2_{(1)}$-distribution
  - We reject the null hypothesis of no autocorrelation
- Alternatively, we can reject H_0 by examining the *p*-value for LM = 27.61, which is 0.000