Autoregressive Integrated Moving Average (ARIMA) Models

1. Introduction

Time series data are sequences of observations recorded over time. In practice, such data often exhibit **serial dependence**, meaning that current values are correlated with past values. Earlier forecasting methods, such as **exponential smoothing**, capture trends and seasonality but assume that the random component is independent white noise. This assumption is frequently violated, especially when the data have strong autocorrelation.

To address this, we use **Autoregressive Integrated Moving Average (ARIMA)** models, also known as **Box–Jenkins models**. These models explicitly account for the autocorrelation structure, allowing for more efficient and accurate forecasting.

Key ideas:

- **Autoregressive (AR)**: current value depends on its own past values.
- **Moving Average (MA)**: current value depends on past forecast errors (shocks).
- **Integrated (I)**: differencing the data to remove nonstationarity.

ARIMA models are flexible and powerful, encompassing AR, MA, ARMA, and differenced variants to handle both stationary and nonstationary processes

2. Linear Models for Stationary Time Series

A **stationary time series** is one whose statistical properties do not change over time. Formally:

1. The mean is constant over time.
2. The variance is constant over time.
3. The covariance between Yt and Yt+k depends only on lag k not on t.

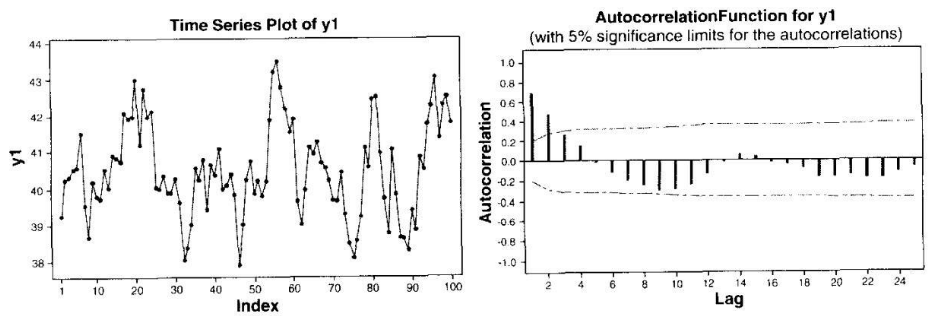
Weak stationarity (based on mean, variance, covariance) is sufficient for most applications.

Why stationarity matters:
ARIMA modeling assumes stationarity. Nonstationary series must be transformed (via differencing or transformations like logs) before applying ARIMA.

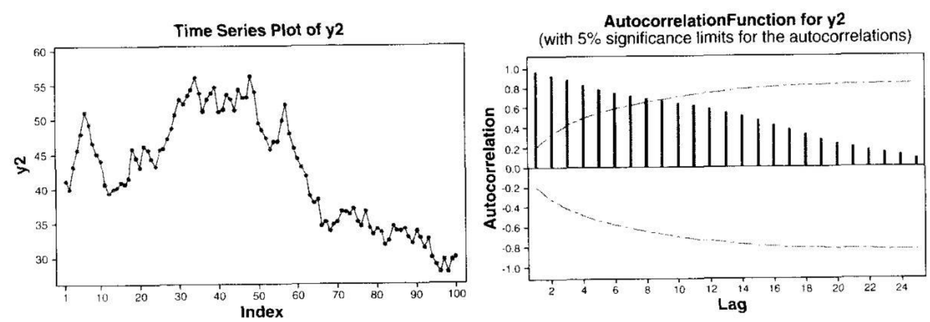
Check for stationarity

- Plot the series (stable mean and variance?).
- Examine the autocorrelation function (ACF). A slowly decaying ACF indicates nonstationarity.
- Conduct formal tests (e.g., Augmented Dickey–Fuller test).

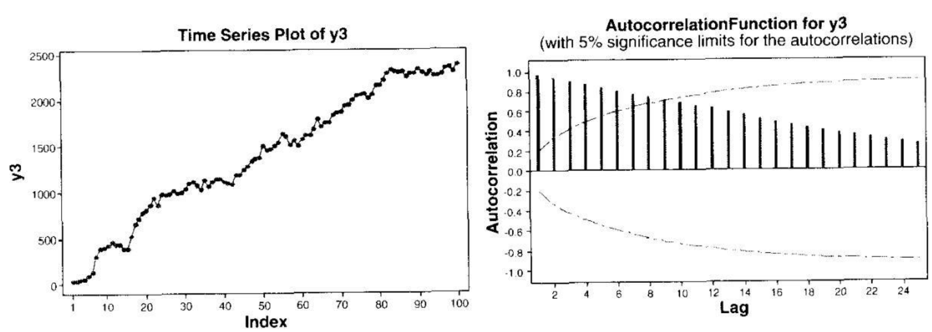
Example



(a) $y_{1,t} = 10 + 0.75y_{1,t-1} + \varepsilon_t$



(b) $y_{2,t} = 2 + 0.95y_{2,t-1} + \varepsilon_t$



(c) $y_{3,t} = 20 + y_{3,t-1} + \varepsilon_t$

FIGURE 5.1 Realizations of (a) stationary, (b) near nonstationary, and (c) nonstationary processes.

The Wold Decomposition

Wold's Theorem states: any weakly stationary, nondeterministic time series can be represented as an **infinite moving average (MA)** of white noise shocks:

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

where ε_t is white noise with mean zero and variance σ^2.

This representation is too general for practical use, so we approximate using **finite-order models**:

- **Moving Average (MA)** of finite order.
- **Autoregressive (AR)** of finite order.
- **Mixed ARMA** processes.

3. Finite-Order Moving Average (MA) Processes

Theory

An MA(q) process expresses Yt as a linear function of current and past q shocks::

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

- Always stationary, regardless of coefficients.
- Mean: E(Yt)=μ.
- Variance: depends on θ values.
- **ACF:** nonzero only up to lag q (cuts off).

This cutoff property makes MA processes identifiable by their sample ACF.

Examples

- **MA(1):**

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

ACF: nonzero only at lag 1.
If $\theta 1>0$, successive observations tend to alternate in sign (negative autocorrelation).
If $\theta 1<0$, successive values tend to move in the same direction briefly (positive autocorrelation).

For the first-order moving average or MA(1) model, we have the autocovariance function as

$$\begin{aligned} \gamma_y(0) &= \sigma^2\left(1+\theta_1^2\right) \\ \gamma_y(1) &= -\theta_1\sigma^2 \\ \gamma_y(k) &= 0, \quad k>1 \end{aligned} \tag{5.12}$$

Similarly, we have the autocorrelation function as

$$\begin{aligned} \rho_y(1) &= \frac{-\theta_1}{1+\theta_1^2} \\ \rho_y(k) &= 0, \quad k>1 \end{aligned} \tag{5.13}$$

From Eq. (5.13), we can see that the first lag autocorrelation in MA(1) is bounded as

$$\left|\rho_y(1)\right| = \frac{|\theta_1|}{1+\theta_1^2} \le \frac{1}{2} \tag{5.14}$$

and the autocorrelation function cuts off after lag 1.

Consider, for example, the following MA(1) model:

$$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

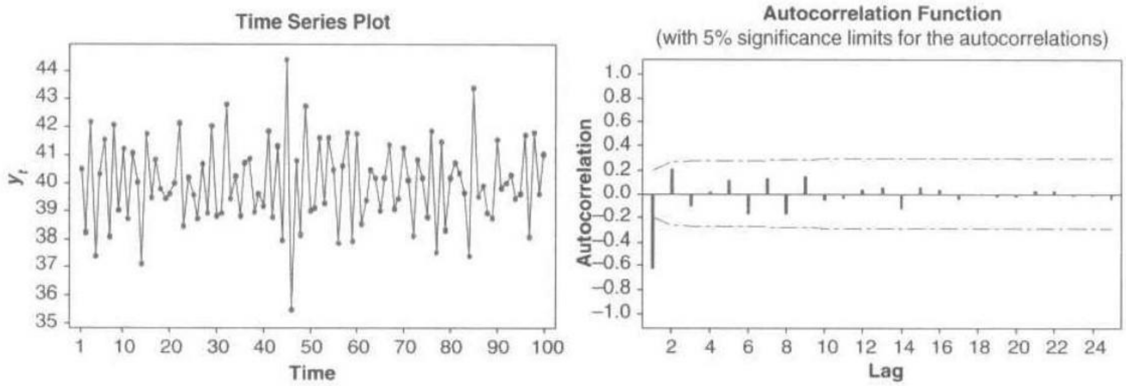


FIGURE 5.3 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$.

We can also consider the following model:

$$y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$$

A realization of this model is given in Figure 5.3. We can see that observations tend to oscillate successively. This suggests a negative autocorrelation as confirmed by the sample ACF plot.

A realization of this model with its sample ACF is given in Figure 5.2. A visual inspection reveals that the mean and variance remain stable while there are some short runs where successive observations tend to follow each other for very brief durations, suggesting that there is indeed some positive autocorrelation in the data as revealed in the sample ACF plot.

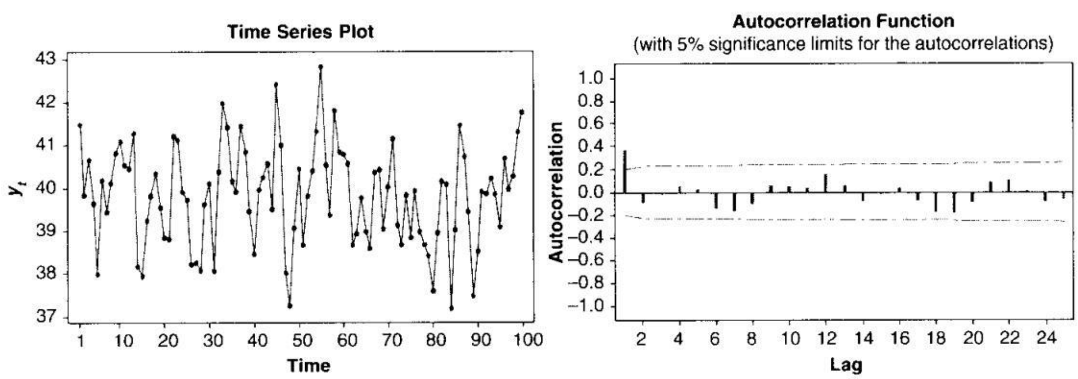


FIGURE 5.2 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$.

- **MA(2):**

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

ACF: nonzero only at lags 1 and 2, then cuts off.

Another useful finite order moving average process is MA(2), given as

$$\begin{aligned} y_t &= \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ &= \mu + \left(1 - \theta_1 B - \theta_2 B^2\right) \varepsilon_t \end{aligned} \tag{5.15}$$

The autocovariance and autocorrelation functions for the MA(2) model are given as

$$\begin{aligned} \gamma_y(0) &= \sigma^2 \left(1 + \theta_1^2 + \theta_2^2\right) \\ \gamma_y(1) &= \sigma^2 (-\theta_1 + \theta_1 \theta_2) \\ \gamma_y(2) &= \sigma^2 (-\theta_2) \\ \gamma_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \rho_y(1) &= \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(2) &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.17}$$

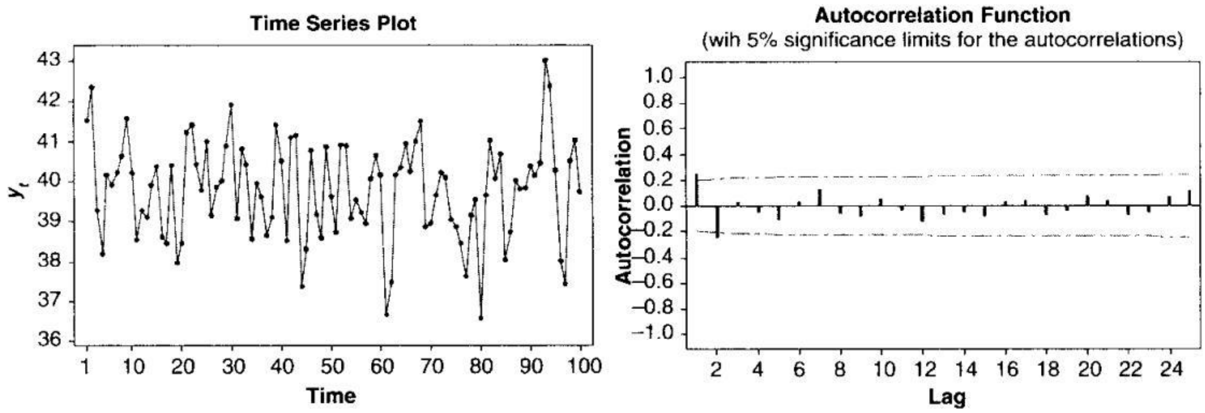


FIGURE 5.4 A realization of the MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$.

Figure 5.4 shows the time series plot and the autocorrelation function for a realization of the MA(2) model:

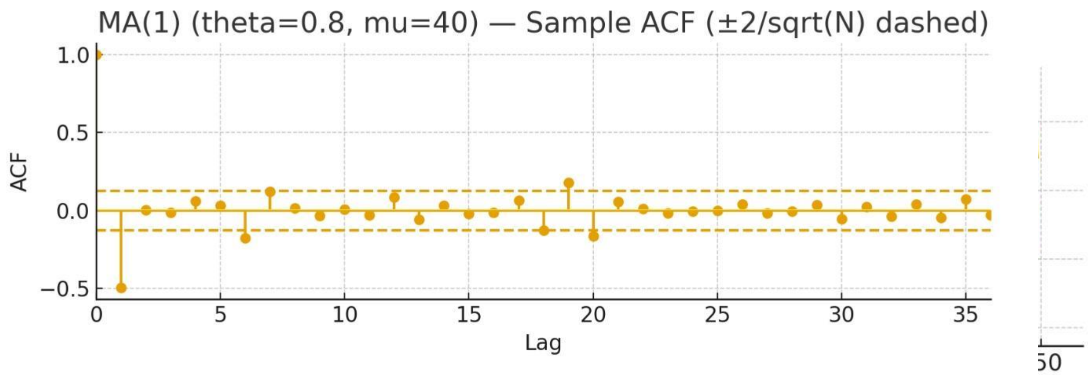
$$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$$

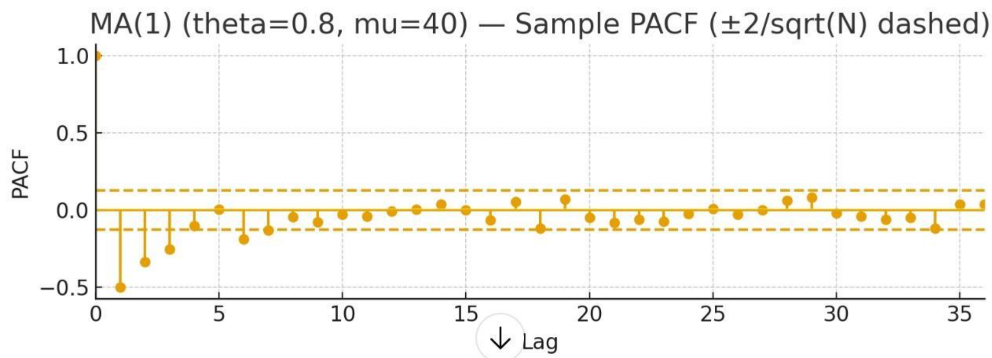
Note that the sample ACF cuts off after lag 2.

Worked Example: MA(1)

- Simulated with θ=0.8,μ=40.
- **Time series**: fluctuates around 40 with short memory.
- **ACF**: significant spike at lag 1, cuts off after.

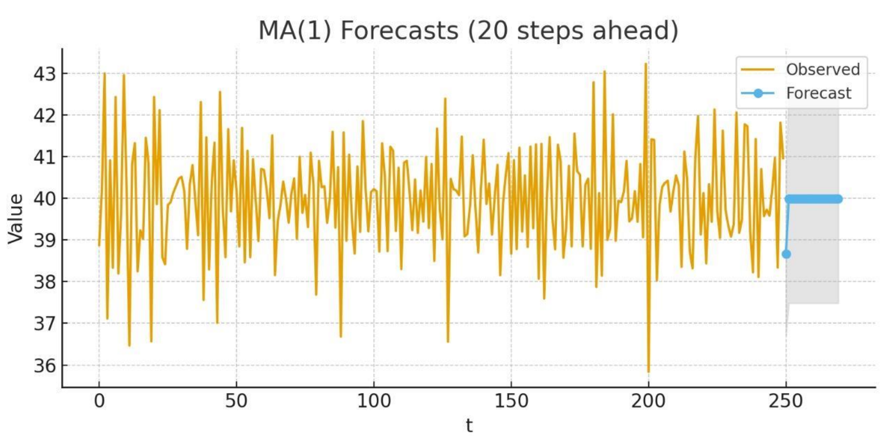
- **PACF**: decays gradually.





Forecasting:

- Forecasts quickly converge to the mean (39.99).
- Fitted parameter: θ ≈ –0.808 (note negative sign convention in statsmodels).



4. Finite-Order Autoregressive (AR) Processes

Theory

An AR(p) process expresses Yt as a function of its own past values plus noise::

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- Stationary if the roots of the characteristic polynomial

$$1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 0$$

lie outside the unit circle.

- Mean: $\mu/(1 - \phi_1 - \cdots - \phi_p)$.
- **ACF:** decays gradually (exponential or damped sinusoid).
- **PACF:** cuts off after lag p.

Examples

- **AR(1):**

$$Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$$

Stationary if $|\phi_1| < 1$.
ACF decays geometrically at rate ϕ_1.

- **AR(2):**

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

ACF can be a mixture of exponential decays or damped oscillations, depending on roots of characteristic equation.

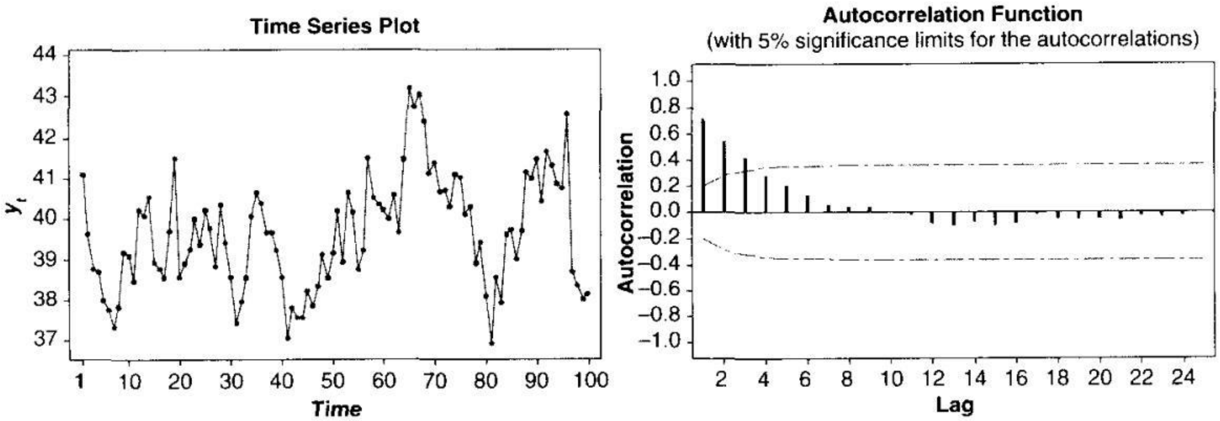


FIGURE 5.5 A realization of the AR(1) process, $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$.

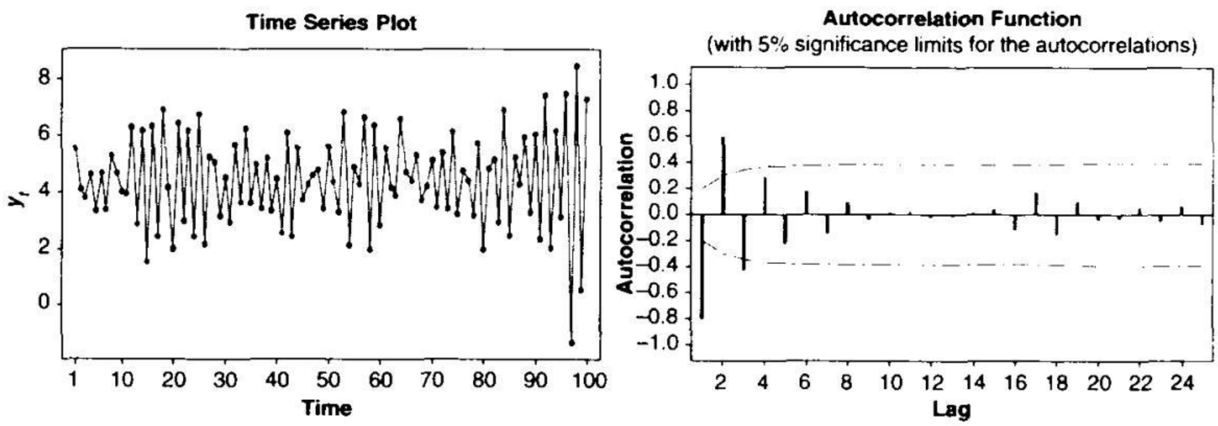
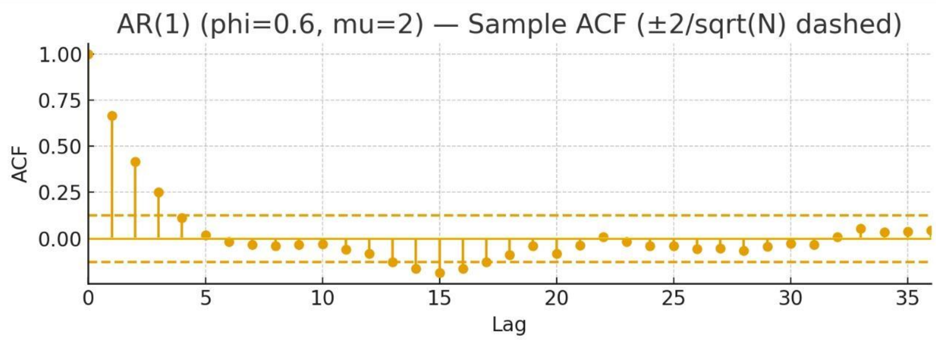


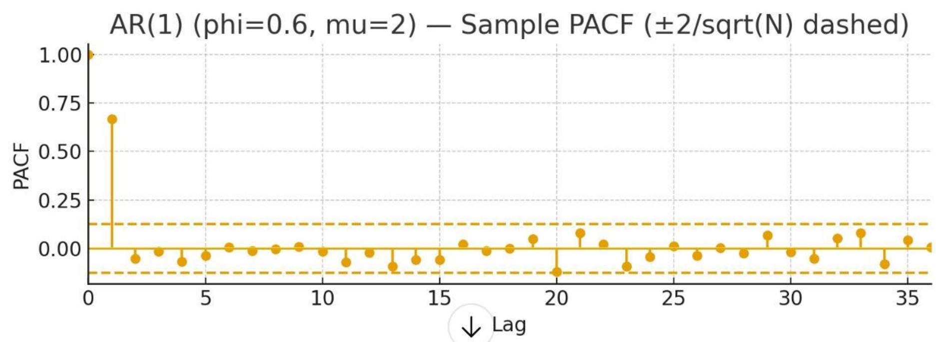
FIGURE 5.6 A realization of the AR(1) process, $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$.

Worked Example: AR(1)

- Simulated with $\phi=0.6, \mu=2$.
- **Time series**: mean-reverting around $\sim 2/(1-0.6) \approx 5$.
- **ACF**: gradual geometric decay.

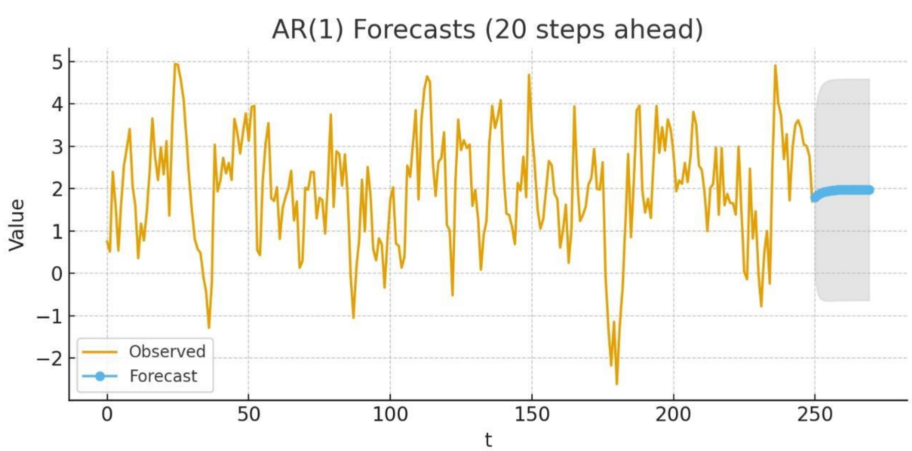
- **PACF**: cutoff after lag 1.





Forecasting:

- Forecasts taper smoothly, approaching long-run mean.
- Estimated φ ≈ 0.665, close to true 0.6.



Partial Autocorrelation Function (PACF)

The PACF measures correlation between Yt and Yt−k after controlling for intermediate lags.

Identification rule:

- **MA(q):** ACF cuts off after lag q, PACF tails off.
- **AR(p):** PACF cuts off after lag p, ACF tails off.
- **ARMA(p,q):** both ACF and PACF tail off.

This is the foundation of Box–Jenkins model identification.

5. Mixed ARMA Processes

An ARMA(p,q) combines AR(p) and MA(q):

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

- ACF and PACF both tail off gradually.
- Identification requires ACF/PACF patterns + criteria like AIC/BIC.

6. Nonstationary Processes and Differencing

Many real-world series are nonstationary due to trend or changing variance.

Differencing operator:

$$\nabla Y_t = Y_t - Y_{t-1}$$

Higher orders: $\nabla^d Y_t$.

If differencing d times yields stationarity, the model is ARIMA(p,d,q).

Examples:

- **Random Walk (ARIMA(0,1,0)):**

$$Y_t = Y_{t-1} + \varepsilon_t$$

Nonstationary, but first difference is white noise.

- **ARIMA(0,1,1):** Equivalent to exponential smoothing model.

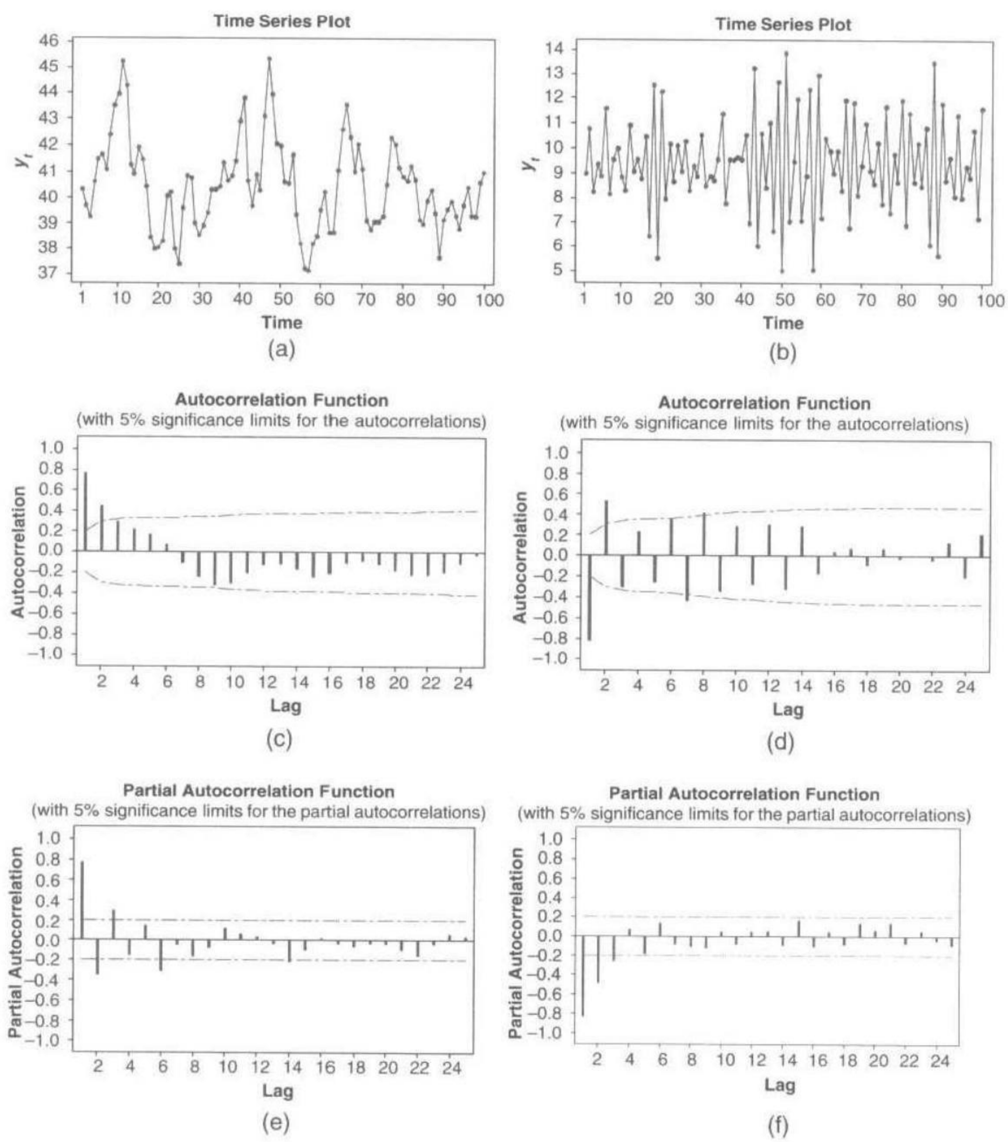


FIGURE 5.10 Two realizations of the ARMA(1,1) model: (a) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$ and (b) $y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$. (c) The ACF of (a), (d) the ACF of (b), (e) the PACF of (a), and (f) the PACF of (b).

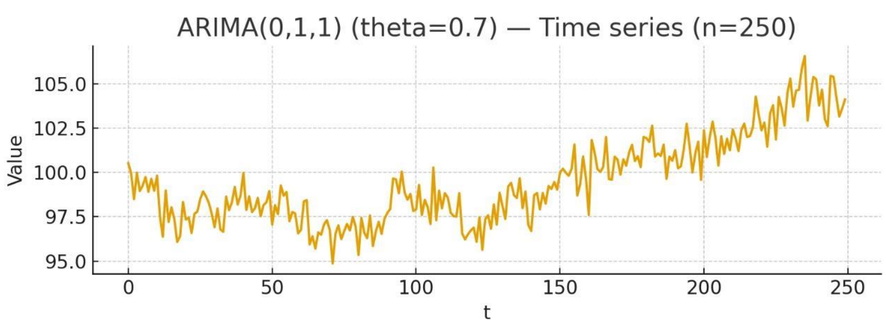
ARIMA Models (Integrated)

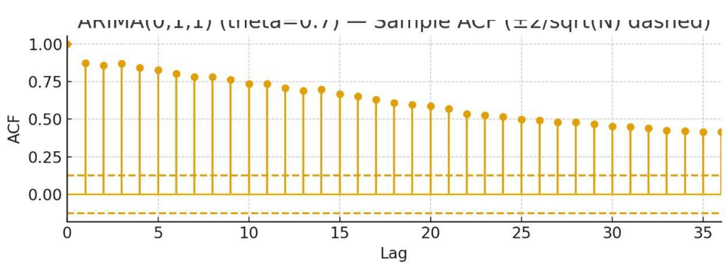
Theory

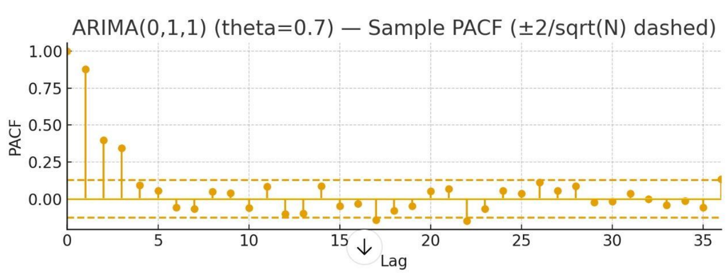
- Many series are nonstationary due to trend.
- Differencing operator: $\nabla Y_t = Y_t - Y_{t-1}$.
- ARIMA(p,d,q) applies ARMA to differenced series.

Worked Example: ARIMA(0,1,1)

- Simulated with $\theta = 0.7$.
- **Time series**: exhibits random wandering (nonstationary).
- **ACF**: large spike at lag 1, slow decay.
- **PACF**: decays gradually.







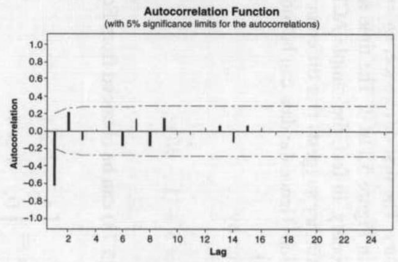
Interpretation: The series behaves like a **random walk with short-term correlation**.

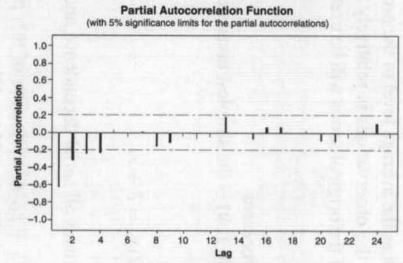
TABLE 5.2 Sample ACFs and PACFs for Some Realizations of MA(1) and MA(2) Models

Model	Sample ACF	Sample PACF

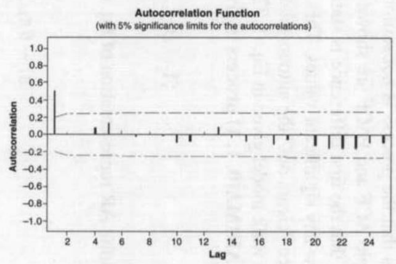
MA(1)

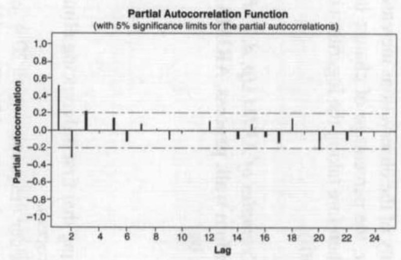
$y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$





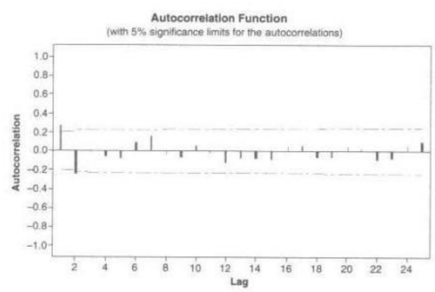
$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$

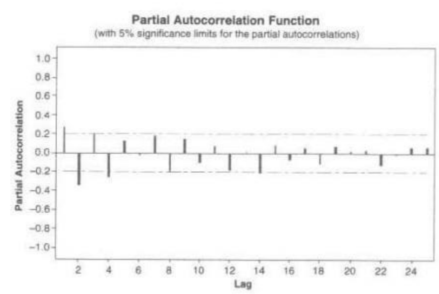




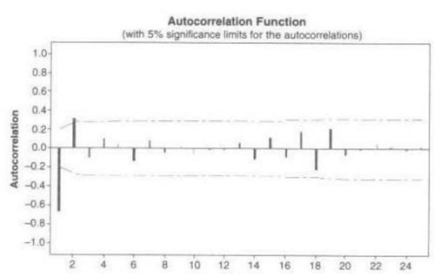
MA(2)

$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$





$y_t = 40 + \varepsilon_t - 1.1\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$



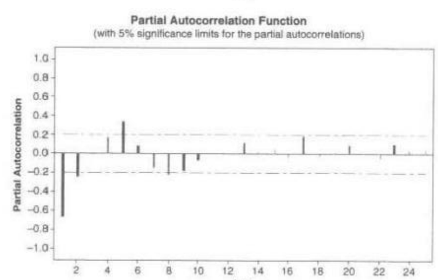


TABLE 5.3 Sample ACFs and PACFs for Some Realizations of AR(1) and AR(2) Models

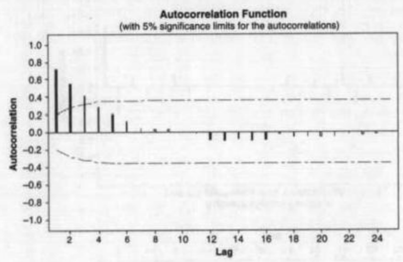
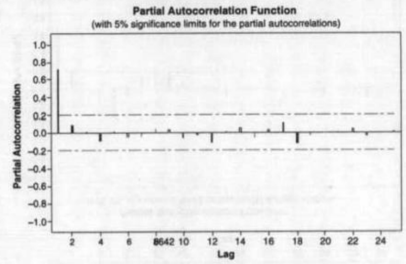
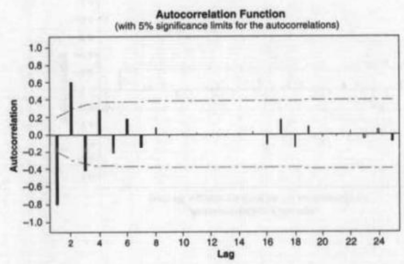
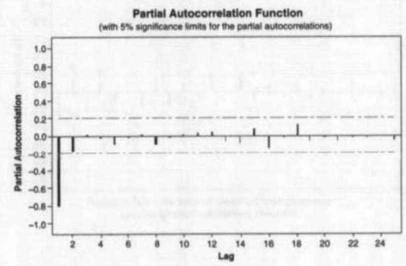
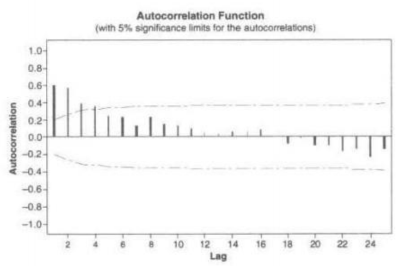
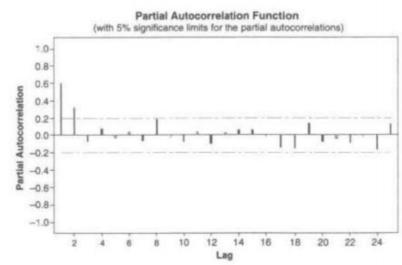
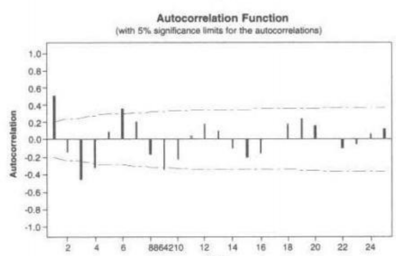
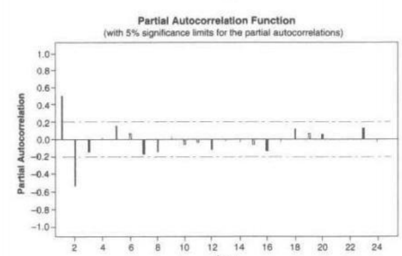
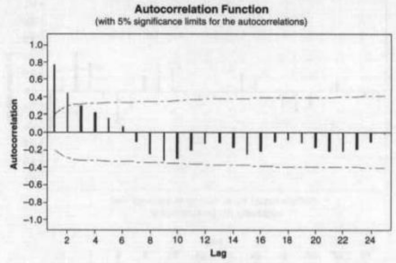
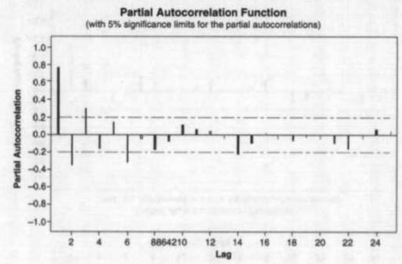
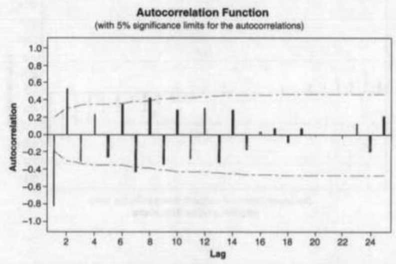
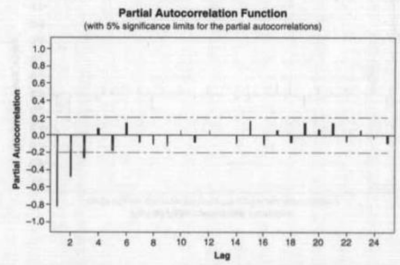
Model	Sample ACF	Sample PACF
AR(1) $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$		
$y_t = 8 - 0.8y_{t-1} + \varepsilon_t$		
AR(2) $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$		
$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$		

TABLE 5.4 Sample ACFs and PACFs for Some Realizations of ARMA(1,1) Models

Model	Sample ACF	Sample PACF
ARMA(1.1) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$		
$y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$		

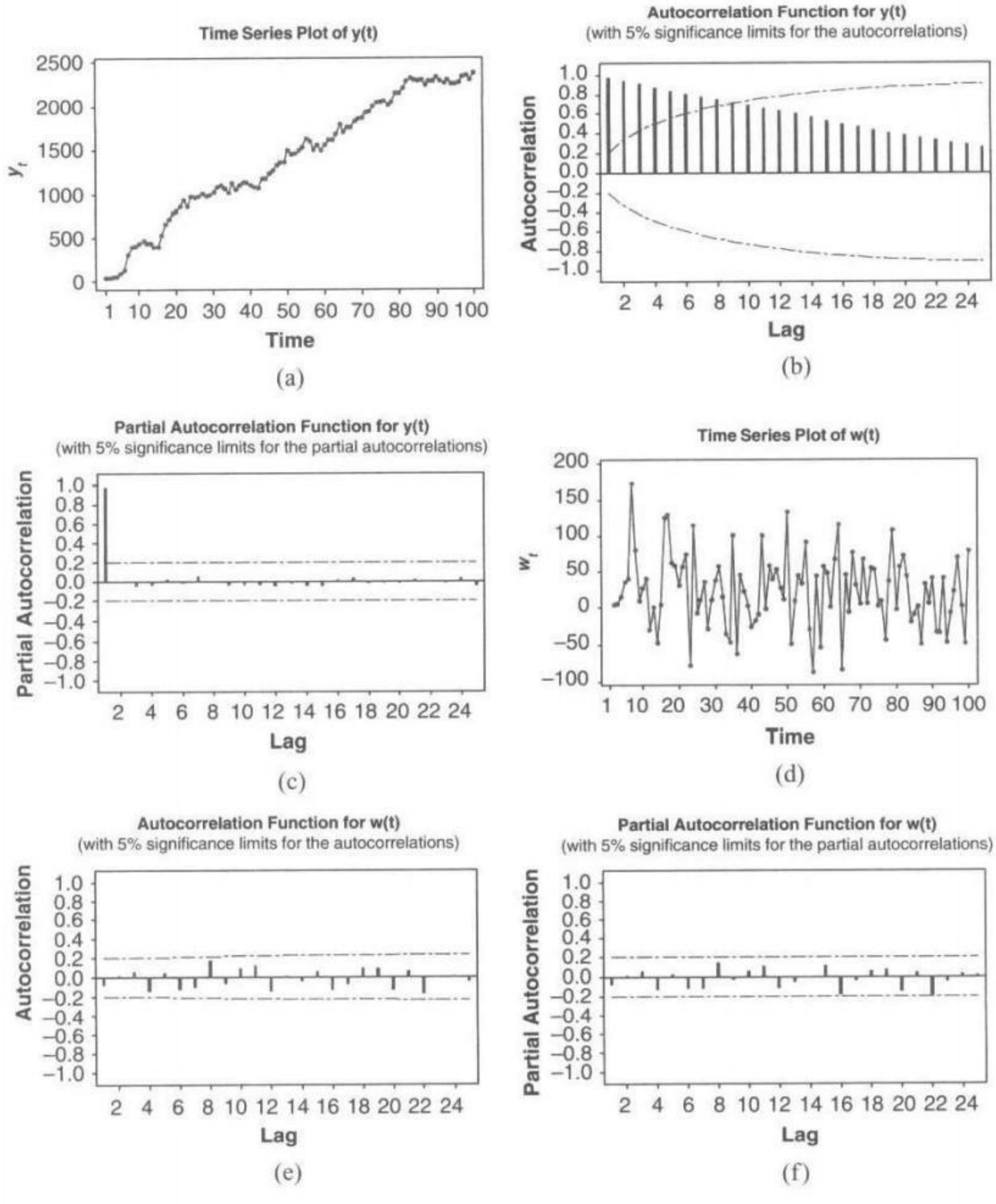


FIGURE 5.11 A realization of the ARIMA(0, 1, 0) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

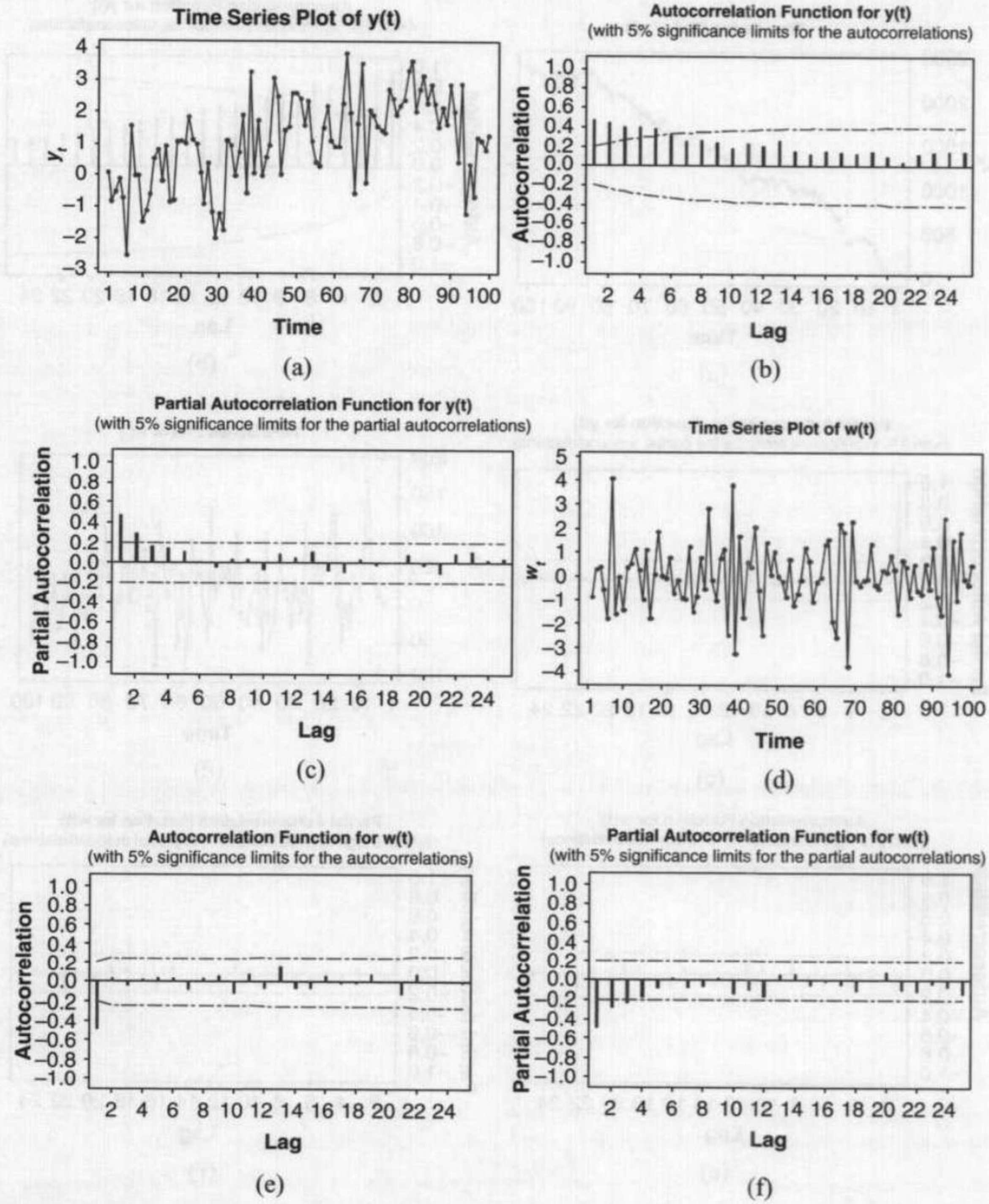


FIGURE 5.12 A realization of the ARIMA(0, 1, 1) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

7. Seasonal ARIMA Models

Theory

To handle seasonality, we apply seasonal AR, differencing, and MA terms at lag sss (season length).

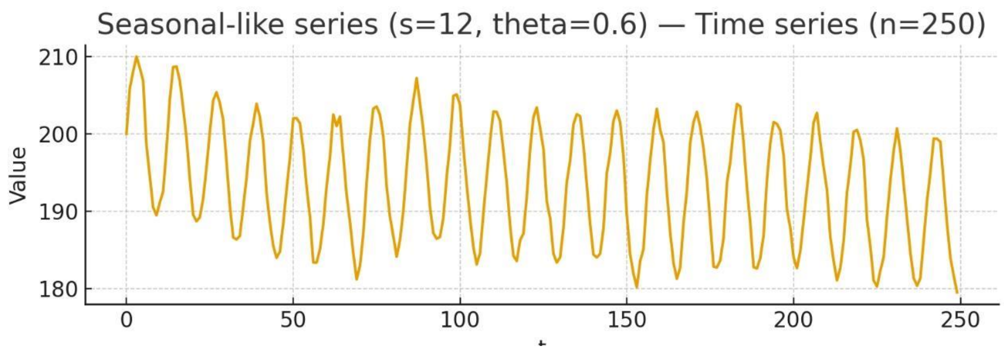
General seasonal ARIMA:

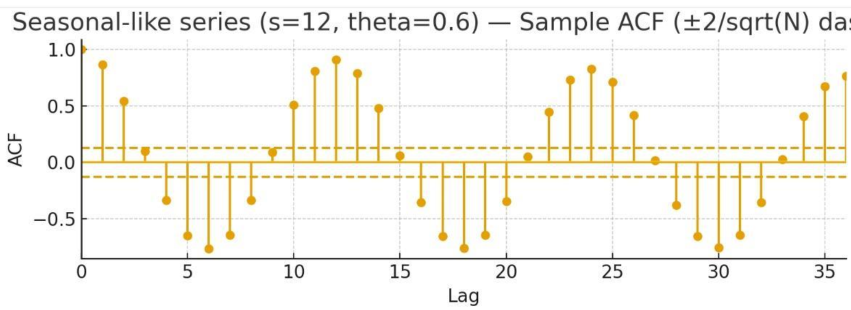
$$\Phi(B^s)\phi(B)\nabla^d\nabla_s^D Y_t = \Theta(B^s)\theta(B)\varepsilon_t$$

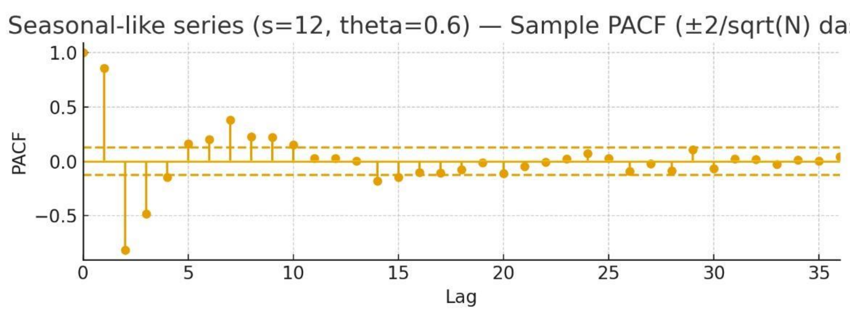
- Handles both trend (differencing) and seasonality (seasonal differencing or seasonal AR/MA).
- Seasonal period s = length of repeating cycle .

Worked Example: Seasonal-like series (s=12)

- Simulated with sinusoidal seasonal component + seasonal MA shocks.
- **Time series**: strong annual-like oscillation.
- **ACF**: seasonal spikes at lags 12, 24, …
- **PACF**: also shows seasonal correlation.







8. Model Building Procedure (Box–Jenkins Approach)

- **Identification**:
  - Plot series, check stationarity.
  - Apply differencing if needed.
  - Use ACF/PACF to suggest AR/MA orders.
- **Estimation**:
  - Estimate parameters using Maximum Likelihood or Least Squares.
- **Diagnostic Checking**:

- Analyze residuals (should be white noise).
- Use Ljung–Box Q-test for independence.

- **Forecasting**:

  - Use fitted ARIMA model for future predictions.

1. **Identification**: Plot series, check stationarity, inspect ACF/PACF.
   - AR(1) → PACF cutoff at lag 1.
   - MA(1) → ACF cutoff at lag 1.
   - ARIMA(0,1,1) → needs differencing, then MA(1) pattern emerges.
   - Seasonal → seasonal differencing reveals stationary ARMA structure.
2. **Estimation**: Parameters estimated via maximum likelihood.
   - We saw AR(1): $\varphi \approx 0.665$.
   - MA(1): $\theta \approx -0.808$.
3. **Diagnostics**: Residuals should look white noise.
   - Not detailed here, but normally check Ljung–Box Q test, residual ACF.
4. **Forecasting**:
   - AR(1): forecasts taper toward mean.
   - MA(1): forecasts quickly flatten to mean.
   - ARIMA(0,1,1): forecasts follow last observed trend + shock.
   - Seasonal: forecasts repeat seasonal structure.

Examples from Applications

- **Bank loan data:** Identified AR(1) process after differencing.
- **Dow Jones stock index:** Modeled using ARIMA with slow ACF decay.
- **U.S. clothing sales:** Seasonal ARIMA model required (seasonal differencing + ARMA terms).

9. Final Notes

ARIMA models:

- Provide a flexible framework for stationary and nonstationary series.
- Explicitly incorporate autocorrelation, unlike exponential smoothing.
- Require careful identification, estimation, and diagnostic checking.

Limitations:

- Purely univariate (do not include explanatory variables).
- Not ideal when strong external predictors are available (then regression with ARIMA errors is preferred).
- Identification can be complex.

Despite these challenges, ARIMA remains a cornerstone of time series analysis, widely used in economics, engineering, business, and many applied sciences.