4. Stochastic Time Series Model

- In Chapter 2 and 3 we consider forecast models of the form

$$z_t = f(\mathbf{x}_t; \boldsymbol{\beta}) + \varepsilon_t$$

- In these forecasting models it is usually assumed that the error ε_t are uncorrelated, which in turn implies that the observations z_t are uncorrelated. However this assumption is rarely met in Practice

- Thus models that capture the correlation structure have to be considered. Such as AR(P), MA(q), ARMA(p,q) and ARIMA(d,p,q) will be studied in this chapter.

4.1 Stochastic Processes

- An observed time series $(z_1, z_2, \ldots, z_n)$ can be thought of as a particular realization of a stochastic process.

- Stochastic processes in general can be described by an n-dimensional probability distribution $p(z_1, z_2, \ldots, z_n)$.

- Assuming joint normality, then there are only n observations but $n+(n+1)n/2$ unknown parameters. To infer such a general probability structure from just one realizations of the stochastic process will be impossible.

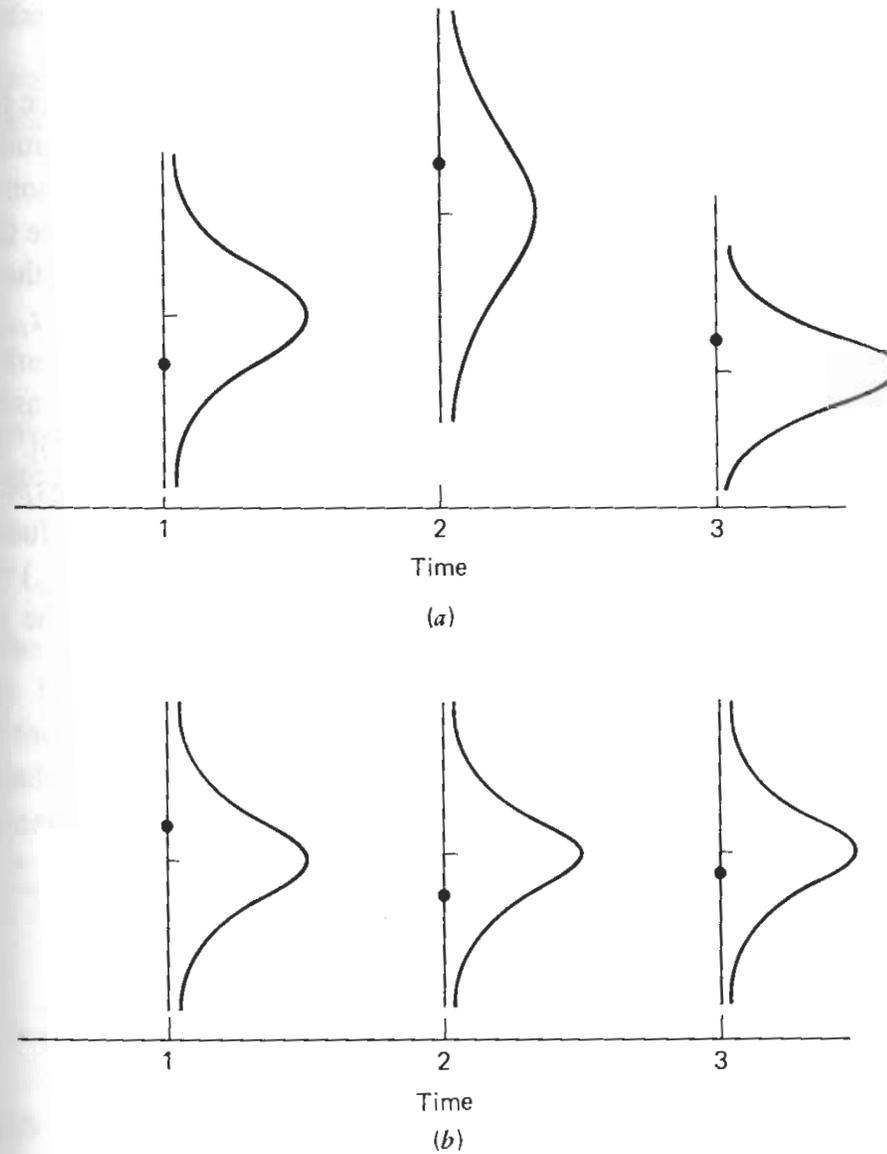- Hence **Stationary Assumption** is used to simplify the probability structure.

(a)

(b)

Figure 5.1. (a) Probability distributions for a general stochastic process ($n = 3$). The covariances $Cov(z_1, z_2)$, $Cov(z_2, z_3)$, $Cov(z_1, z_3)$ may be different. The dots represent possible realizations. (b) Probability distributions for a stationary stochastic process ($n = 3$). Also $Cov(z_1, z_2) = Cov(z_2, z_3)$. The dots represent possible realizations.

Stationary Stochastic Process

- The joint distribution are stable with respect to time:

$$p(z_{t1}, \ldots, z_{t_m}) = p(z_{t1+k}, \ldots, z_{t_m+k})$$

- **Autocorrelation Functions** (**ACF**)

    – The autocovariances

    $$\gamma_k = \mathrm{Cov}(z_t, z_{t-k}) = \mathrm{E}(z_t - \mu)(z_{t-k} - \mu)$$

    – The autocorrelations

    $$\rho_k = \frac{\mathrm{Cov}(z_t, z_{t-k})}{[\mathrm{Var}(z_t) \cdot \mathrm{Var}(z_{t-k})]^{\frac{1}{2}}} = \frac{\gamma_k}{\gamma_0}$$

- **Sample Autocorrelation Function** (SACF)

$$r_k = \frac{\displaystyle\sum_{t=k+1}^{n} (z_t - \bar{z})(z_{t-k} - \bar{z})}{\displaystyle\sum_{t=1}^{n} (z_t - \bar{z})^2}, \quad k = 0, 1, 2, \ldots$$

For uncorrelated observations, the variance of r_k is approximately given by

$$\mathrm{Var}(r_k) \approx \frac{1}{n}$$

Table 5.1. Sample Autocorrelations for the Yield Data

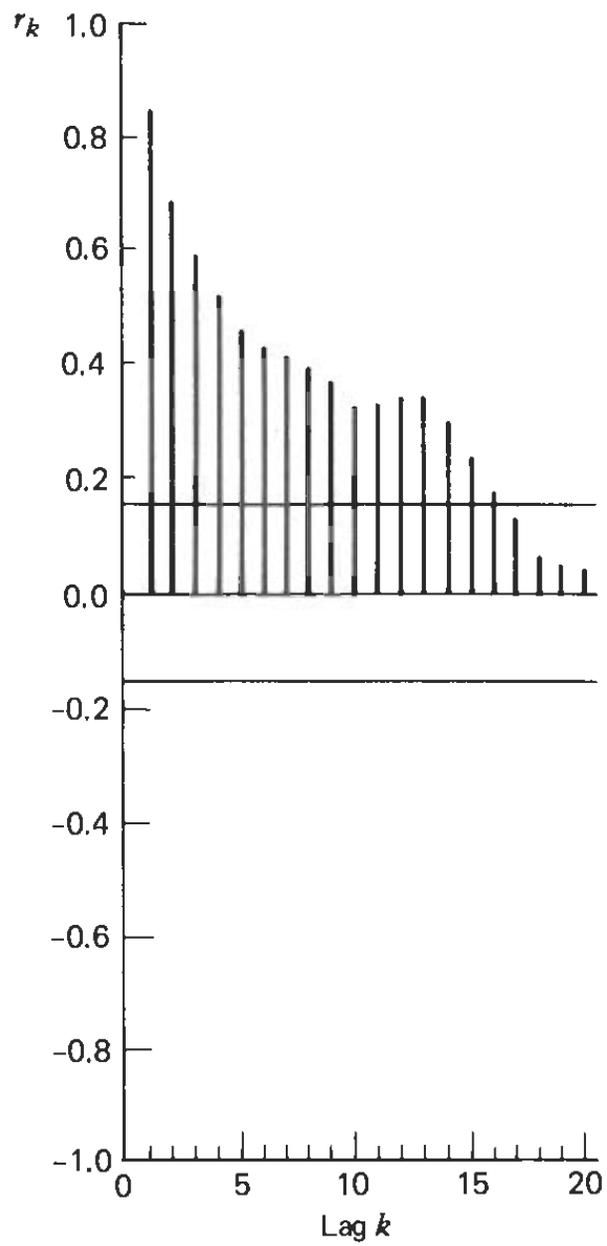| k	1	2	3	4	5	6	7	8	9	10
r_k	0.841	0.683	0.584	0.515	0.457	0.427	0.405	0.386	0.361	0.321
k	11	12	13	14	15	16	17	18	19	20
r_k	0.329	0.338	0.337	0.294	0.231	0.166	0.126	0.062	0.047	0.042

Figure 5.2. SACF for the yield series.

4.2 Stochastic Difference Equation Models

- A time series in which successive values are autocorrelated can be represented as a linear combination (**linear filter**) of a sequence of uncorrelated random variables. The linear filter representation is given by

$$z_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad \psi_0 = 1.$$

The random variable $\{a_t; t = 0, \pm 1, \pm 2, \ldots\}$ are a sequence of uncorrelated random variable from a fixed distribution with mean $\mathrm{E}(a_t) = 0, \mathrm{Var}(a_t) = \mathrm{E}(a_t^2) = \sigma^2$, and $\mathrm{Cov}(a_t, a_{t-k}) = \mathrm{E}(a_k a_{t-k}) = 0$ for all $k \neq 0$.

- $\mathrm{E}z_t = \mu, \quad \gamma_0 = \mathrm{E}(z_t - \mu)^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2,$

$$\gamma_k = \mathrm{E}[(z_t - \mu)(z_{t-k} - \mu)] = \sigma^2 \sum_{j=1}^{\infty} \psi_j \psi_{j+k},$$

$$\rho_k = \left(\sum_{j=0}^{\infty} \psi_j \psi_{j+k} \right) / \sum_{j=0}^{\infty} \psi_j^2.$$

Autoregressive Processes

- **First-Order Autoregressive Process** (AR(1))

$$(z_t - \mu) - \phi(z_{t-1} - \mu) = a_t \Leftrightarrow (1 - \phi B)(z_t - \mu) = a_t$$

We have

$$z_t - \mu = (1 - \phi B)^{-1} a_t = (1 + \phi B + \phi^2 B^2 + \cdots) a_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots$$

Application of the general results for linear filter lead to

$$\gamma_0 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1 - \phi^2}$$

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+k} = \frac{\sigma^2 \phi^k}{1 - \phi^2} = \phi^k \gamma_0, \quad k = 0, 1, 2, \ldots$$

$$\rho_k = \phi^k, \quad k = 0, 1, 2, \ldots$$

Another approach to get the ACF for AR(1). This approach is more suitable for high order situations. Without loss of generality, we assume $\mu = 0$ and $z_t = \phi z_{t-1} + a_t$, then we have

$$\gamma_k = \mathrm{E}(z_t z_{t-k}) = \phi \mathrm{E}(z_{t-1} z_{t-k}) + \mathrm{E}(a_t z_{t-k}) = \phi \gamma_{k-1} + \mathrm{E}(a_t z_{t-k})$$

where

$$k = 0, \quad \mathrm{E}(a_t z_t) = \mathrm{E}[a_t(a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots)] = \sigma^2$$

$$k > 0, \quad \mathrm{E}(a_t z_{t-k}) = 0.$$

Thus

$$
\begin{aligned}
\gamma_0 &= \phi \gamma_{-1} + \sigma^2 = \phi \gamma_1 + \sigma^2 \\
\gamma_k &= \phi \gamma_{k-1}, (k > 1), (\gamma_1 = \phi \gamma_0) \\
\gamma_0 &= \phi^2 \gamma_0 + \sigma^2 = \frac{\sigma^2}{1 - \gamma^2}, \\
\gamma_k &= \phi^k \gamma_0, \\
\rho_k &= \phi^k
\end{aligned}
$$

- **Second-Order Autoregression Process** [AR(2)]

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$$

$$(1 - \phi_1 B - \phi_2 B^2) z_t = a_t$$

and

$$z_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t = (1 + \psi_1 B + \psi_2 B^2 + \cdots) a_t$$

The coefficients $(\psi_1, \psi_2, \ldots)$ could be determined as follows:

$$
\begin{aligned}
1 \quad &= \quad (1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots), \\
B \quad &: \quad \psi_1 - \phi_1 = 0 \Rightarrow \psi_1 = \phi_1; \\
B^2 \quad &: \quad \psi_2 - \phi_1 \psi_2 - \phi_2 \psi_1 = 0 \Rightarrow \psi_2 = \phi_1^2 + \psi_2; \\
B^3 \quad &: \quad \psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 = 0 \Rightarrow \psi_3 = \phi_1^3 + 2\phi_1 \phi_2; \\
\psi_j \quad &= \quad \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2}, j \geq 2.
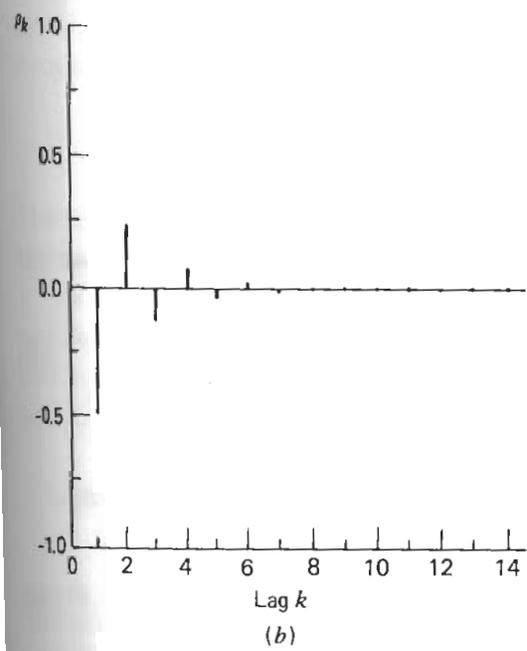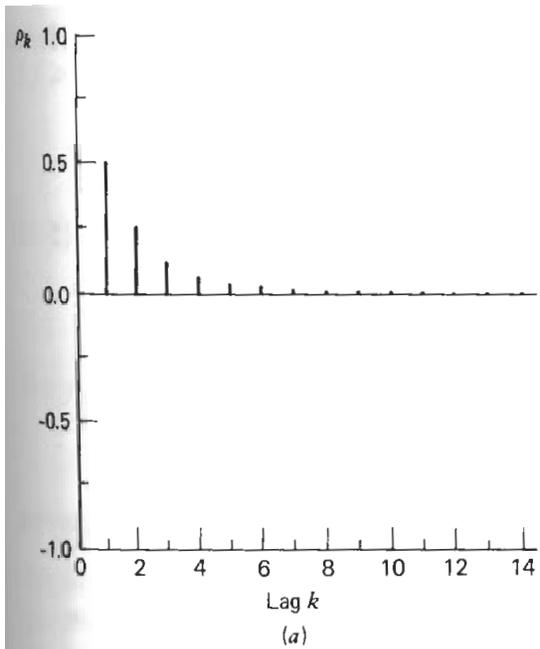\end{aligned}
$$

Figure 5.3. ACF for two AR(1) processes: (a) $\phi = .5$; (b) $\phi = -.5$.

12

- **Stationarity**

$$AR(p) : (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) = (1 - G_1 B) \cdots (1 - G_p B).$$

The stationarity relates to the conditions: $|G_i^{-1}| > 1, i = 1, 2, \ldots, p..$
Particulary for **AR(2)**, the stationarity corresponds to

$$\phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1, -1 < \phi_2 < 1.$$

- **ACF** of **AR(2)**

$$
\begin{aligned}
\mathrm{E}(z_t z_{t-k}) &= \phi \mathrm{E}(z_{t-1} z_{t-k}) + \phi \mathrm{E}(z_{t-2} z_{t-k}) + \mathrm{E}(z_{t-k} a_t). \\
\gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2, \\
\gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, k \geq 1. \\
\rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}. \\
(1 - \rho_1 B - \rho_2 B^2) \rho_k &= 0. \ k = 1, 2, \ldots
\end{aligned}
$$

Actually the general solution of the difference equation:

$$(1 - \phi_1 B - \phi_2 B^2)\rho_k = 0$$

is

$$
\rho_k = \begin{cases}
A_1 G_1^k + A_2 G_2^k & (G_1 \neq G_2), \\
(A_1 + A_2 k)G^k, & (G_1 = G_2 = G), \\
\{A_1 \sin(\theta k) + A_2 \cos(\theta k)\}(p^2 + q^2)^{-k/2}, & p \pm qi = G_{1,2}^{-1}, \\
& \theta = \arcsin(p^2/\sqrt{p^2 + q^2})
\end{cases}
$$

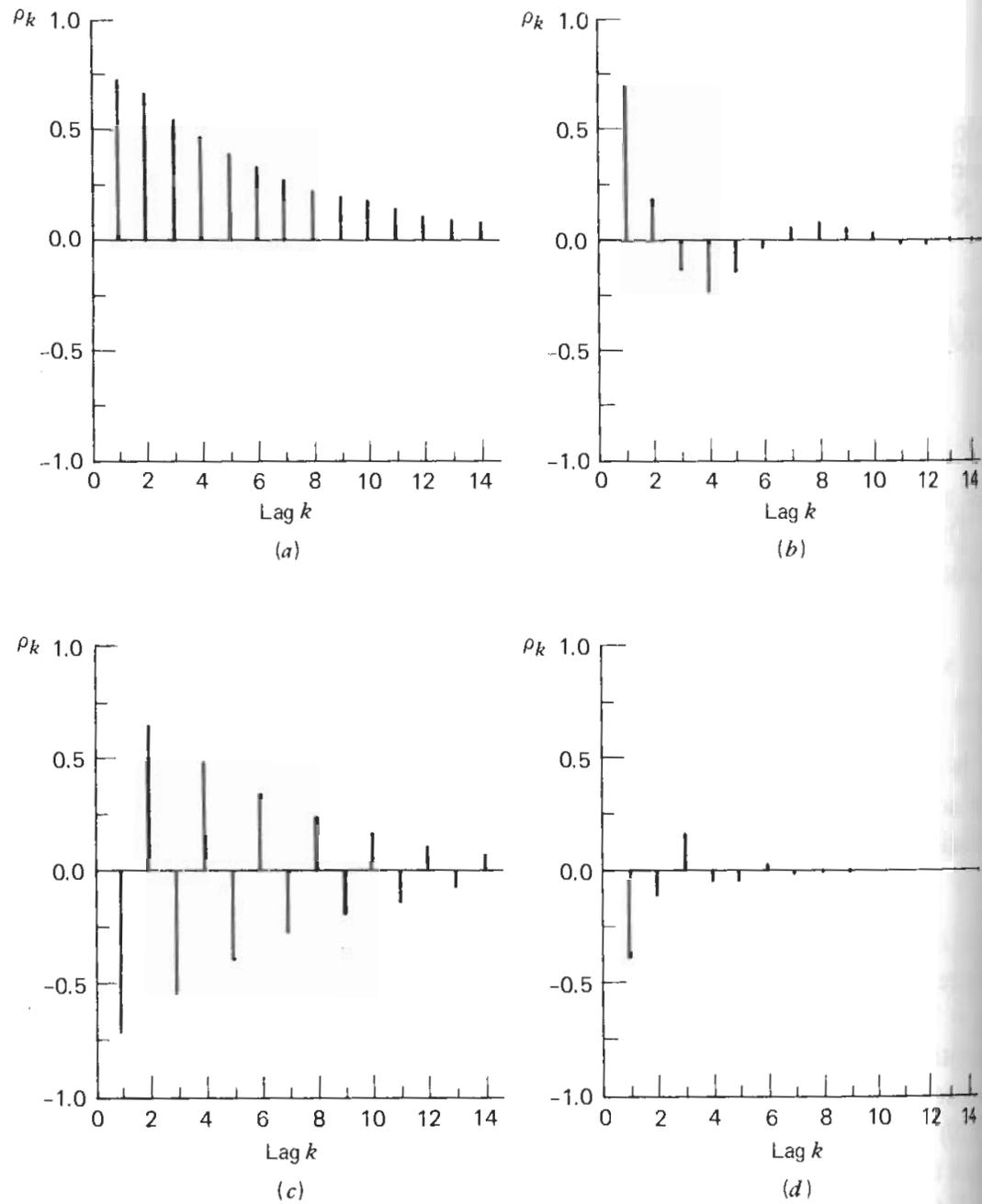So, when stationary, the correlogram shows damped exponentials/sinusoidal waves.

Figure 5.4. ACF for various stationary AR(2) processes: (a) $\phi_1 = .5$, $\phi_2 = .3$; (b) $\phi_1 = 1.0$, $\phi_2 = -.5$; (c) $\phi_1 = -.5$, $\phi_2 = .3$; (d) $\phi_1 = -.5$, $\phi_2 = -.3$.

- **The general AR(p) model**

$$\gamma_0 = \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2 = \frac{\sigma^2}{1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p} = \frac{\sigma^2}{\phi(B)\rho_0}$$

$$\begin{aligned}
\gamma_k &= \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p}, \ k > 0 \\
\rho_k &= \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}.
\end{aligned}$$

- Letting $k = 1, 2, \ldots, p$, we have the **Yule-walker equations:**

$$\boldsymbol{\rho} = P\boldsymbol{\phi},$$

$$P = \begin{bmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
\rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1
\end{bmatrix}$$

- **Moments Methods:**

$$\boldsymbol{\phi} = P^{-1}\hat{\boldsymbol{\rho}}$$

Partial Autocorrelations

When considering two random variables' correlation coefficient, it would be fair if we could get ride of the influence of other random variables related to them. In the **AR(p)** model: assuming the stability

$$z_t = \phi_{k1} z_{k-1} + \phi_{k2} z_{t-2} + \cdots + \phi_{kk} z_{t-k} + a_t \Rightarrow \phi z_t = a_t$$

implies $\phi(B)$ has all the roots outside the unit circle. Therefore $|\phi_{kk}| < 1$.

This is defined as the partial autocorrelation between z_t and z_{t-k}.

- AR(1): $z_t = \phi_{11} z_{t-1} + a_t \Rightarrow \phi_{11} = \rho_1$.

- AR(2): $\rho_1 = \phi_{21} + \rho_1 \phi_{22}$, $\rho_2 = \rho_1 \phi_{21} + \phi_{22} \Rightarrow \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$

- AR(p): Using Yule-walker Equations

$$
\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}
$$

$$
\phi_{kk} = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix} \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}^{-1}
$$

- **The basic features of AR(p):**

  - **ACF:** Combinations of damped exponentials/sinusodal waves;
  - **PACF:** zero for lags larger than p.

- **Sample Partial Autocorrelation Functions** (SPAF)

  1. Use the sample correlation coefficient r_j in place of ρ_j and solve the Yule-Walker equation for $\hat{\rho}_{kk}$. This scheme is impractical if lag k is large.
  2. Recursive forms:

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j},$$

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j}, \ j = 1, 2, \ldots, k-1.$$

  3. Standard Errors of $\hat{\phi}_{kk}$
     If the data follow an AR(p) process, then for lags greater than p the variance of $\hat{\phi}_{kk}$ can be approximated by

$$\mathrm{Var}(\hat{\phi}) \approx n^{-1} \quad \text{for } k > p.$$

Approximate standard error for $\hat{\phi}_{kk}$ are given by $n^{-1/2}$.

Table 5.1. Sample Autocorrelations for the Yield Data

k	1	2	3	4	5	6	7	8	9	10
r_k	0.841	0.683	0.584	0.515	0.457	0.427	0.405	0.386	0.361	0.321
k	11	12	13	14	15	16	17	18	19	20
r_k	0.329	0.338	0.337	0.294	0.231	0.166	0.126	0.062	0.047	0.042

Table 5.2. Sample Partial Autocorrelations for the Yield Data

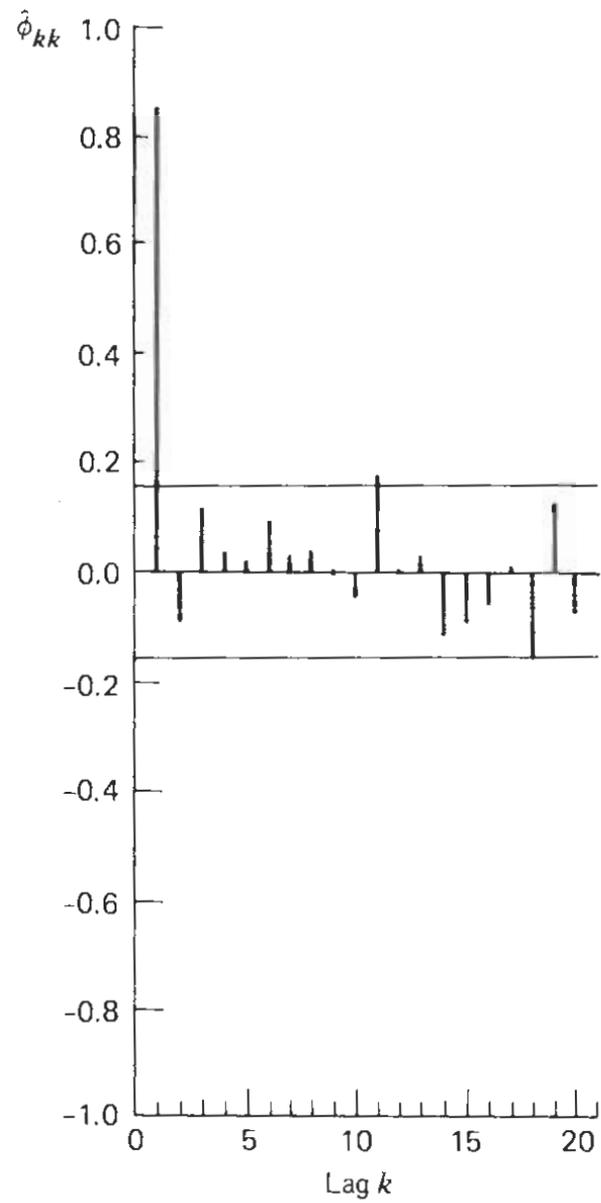| k	1	2	3	4	5	6	7	8	9	10
$\hat{\phi}_{kk}$	0.841	−0.083	0.111	0.036	0.018	0.091	0.025	0.035	0.003	−0.044
k	11	12	13	14	15	16	17	18	19	20
$\hat{\phi}_{kk}$	0.168	0.001	0.027	−0.110	−0.080	−0.057	0.007	−0.152	0.122	−0.071

Figure 5.5. SPACF for the yield series.

Moving Average MA(q) Model

$$z_t - \mu = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t = \theta(B)a_t$$

- **Autocovariances and Autocorrelations**

$$
\begin{aligned}
\gamma_0 &= (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2 \\[2mm]
\gamma_k &= (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma^2, \ \ k = 1, 2, \ldots, q \\[2mm]
\gamma_k &= 0, \ \ k > q \\[2mm]
\rho_k &= \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, \ \ k = 1, 2, \ldots, q \\[2mm]
\rho_k &= 0, k > q
\end{aligned}
$$

- **Partial Autocorrelation Function for MA Process**

The ACF of a moving average process of order q cuts off after lag q ($\rho_k = 0, k > q$). However, its PACF is infinite in extent (it tails off).

In General, the PACF is dominated by combinations of damped exponentials and/or damped sine waves.

- **Example: MA(1) Model:** $z_t - \mu = a_t - \theta a_{t-1}$

$$\gamma_0 = (1 + \theta^2)\sigma^2$$

$$\gamma_1 = -\theta\sigma^2$$

$$\rho_1 = \frac{-\theta}{1 + \theta^2}; (|\theta| < \frac{1}{2}), \rho_k = 0, \ k > 1$$

Invertibility condition: since both $(\theta, 1/\theta)$ are roots of the equation

$$\theta^2 \rho_1 + \theta + \rho_1 = 0,$$

so we restrict ourselves to the situation with $|\theta| < 1$ so that the model could invert to the **AR** from

$$z_t - \mu = \pi_1(z_{t-1} - \mu) + \pi_2(z_{t-2} - \mu) + \cdots + a_t$$

where $\sum_{j=1}^{\infty} |\pi_j| < +\infty$.
For **MA(1)** Model

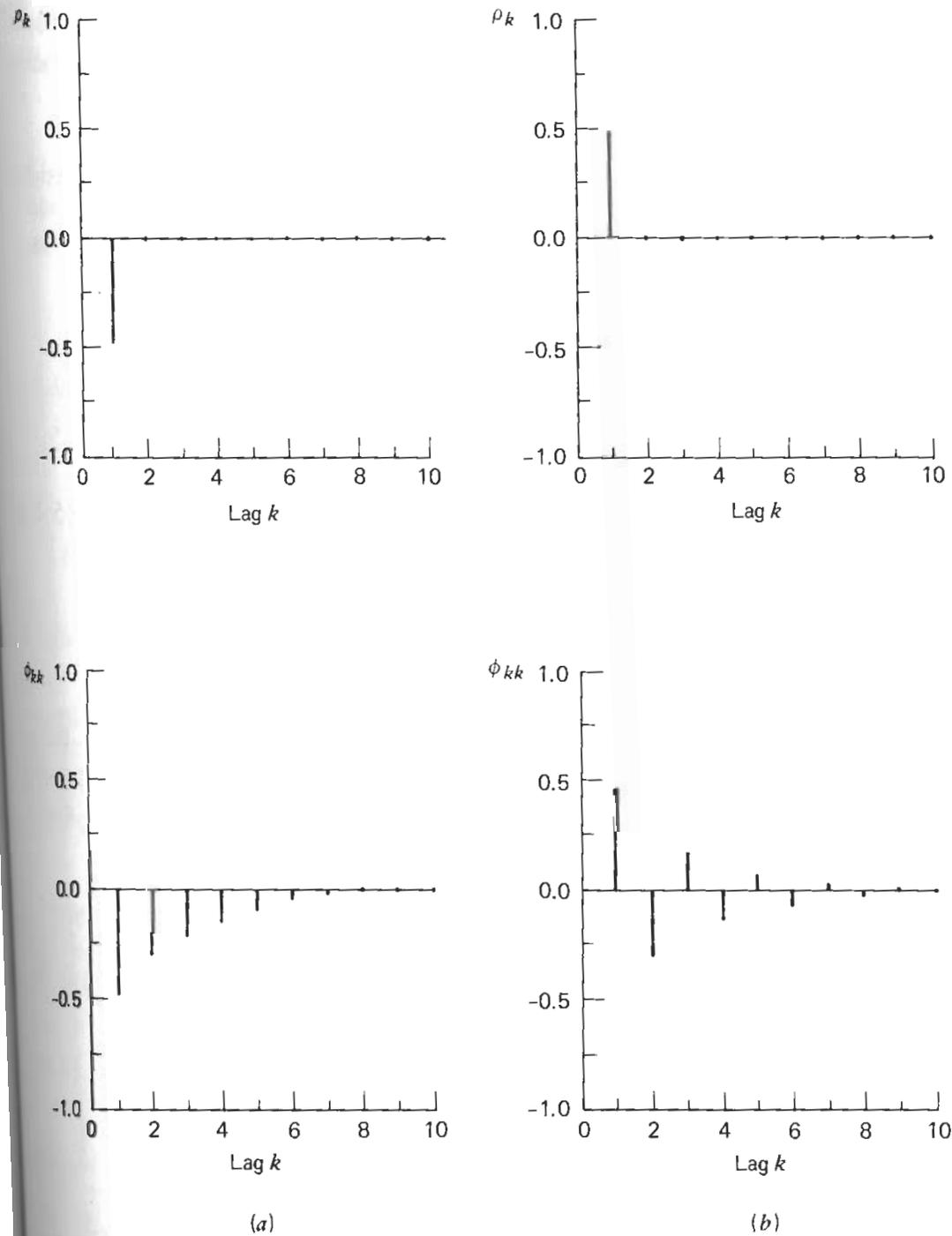$$\pi_j = -\theta^j, (|\theta| < 1), \quad \phi_{kk} = \frac{-\theta^k(1 - \theta^2)}{1 - \theta^{2(k+1)}}, \ k > 0.$$

Figure 5.6. ACF (top) and PACF (bottom) for two MA(1) processes: (a) $\theta = .7$; (b) $\theta = -.7$.

Autoregressive Moving Average (ARMA(p,q)) Processes

$$(1 - \phi_1 B - \cdots \phi_p B^p)(z_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t$$

or

$$
\begin{aligned}
z_t - \mu &= \phi_1(z_{t-1} - \mu) + \cdots + \phi_p(z_{t-p} - \mu) \\
&+ a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\end{aligned}
$$

- $k > q - p$, ACF dominated by AR part

- $k > p - q$, PACF dominated by MA part

Summary

The ACF and PACF of a mixed ARMA(p,q) process are both infinite in extent and tail off (die down) as the lag k increases. Eventually (for $k > q - p$), the ACF is determined from the autoregressive part of the model. The PACF is eventually (for $k > p - q$) determined from moving average part of model.

- **ARMA(1,1) Process**

$$(1 - \phi B)(z_t - \mu) = (1 - \theta B)a_t$$

  - **ACF:**

$$
\begin{aligned}
\gamma_0 &= \frac{1 + \theta^2 - 2\phi\theta}{1 - \phi^2}\sigma^2, \\
\gamma_1 &= \frac{(1 - \phi\theta)(\phi - \theta)}{1 - \phi^2}\sigma^2, \\
\rho_1 &= \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta} \\
\rho_k &= \phi\rho_{k-1}, \ k > 1.
\end{aligned}
$$

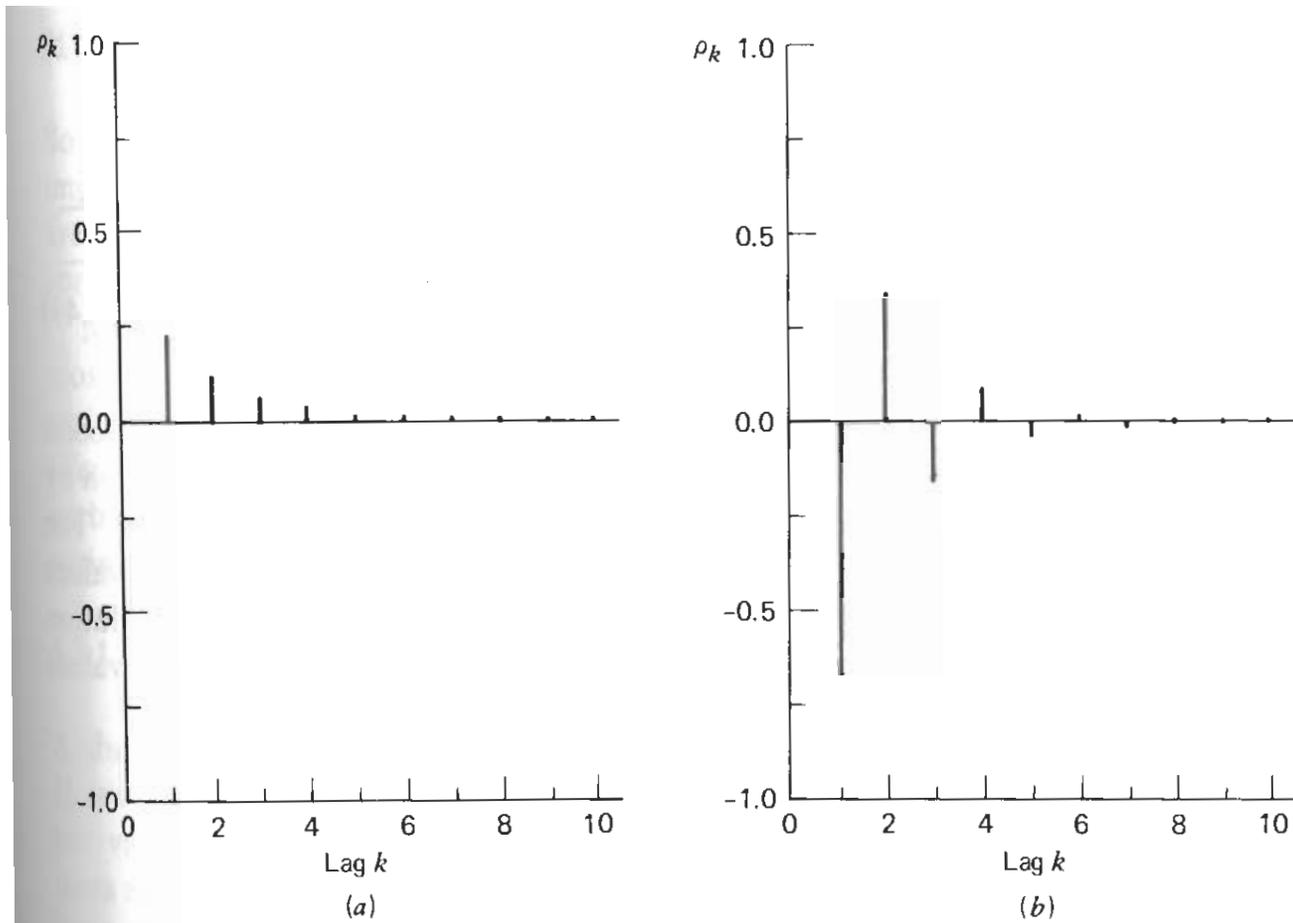 - **PACF:** $\phi_{11} = \rho_1$, then similar to **MA(1)**

Figure 5.7. ACF for two ARMA$(1, 1)$ processes: (a) $\phi = .5, \theta = .3$; (b) $\phi = -.5, \theta = .3$.

4.3 Nonstationary processes

- **Nonstationary, Differencing, and Transformations**

M1: $z_t = \beta_0 + \beta_1 t + a_t$
M2: $(1 - B)z_t = \beta_1 + a_t$ or $z_t = Z_{t-1} + \beta_1 + a_t$.
M3: $(1 - B)^2 z_t = a_t$ or $z_t = 2z_{t-1} - z_{t-2} + a_t$.

- Nonstationary Variance

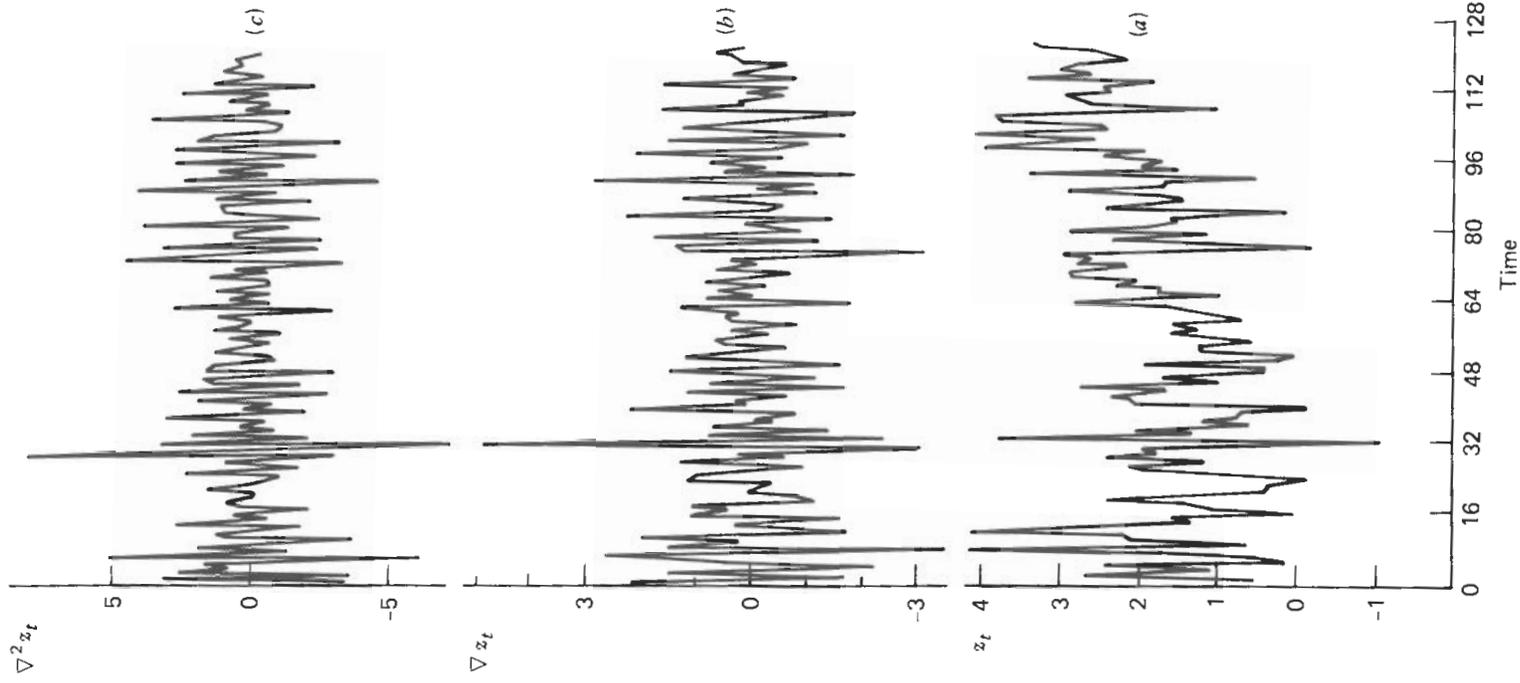- Examples: 1. Growth rates. 2. Demand for repair parts.

Figure 5.8. Growth rates of Iowa nonfarm income: (a) original series z_t; (b) first differences ∇z_t; (c) second differences $\nabla^2 z_t$.
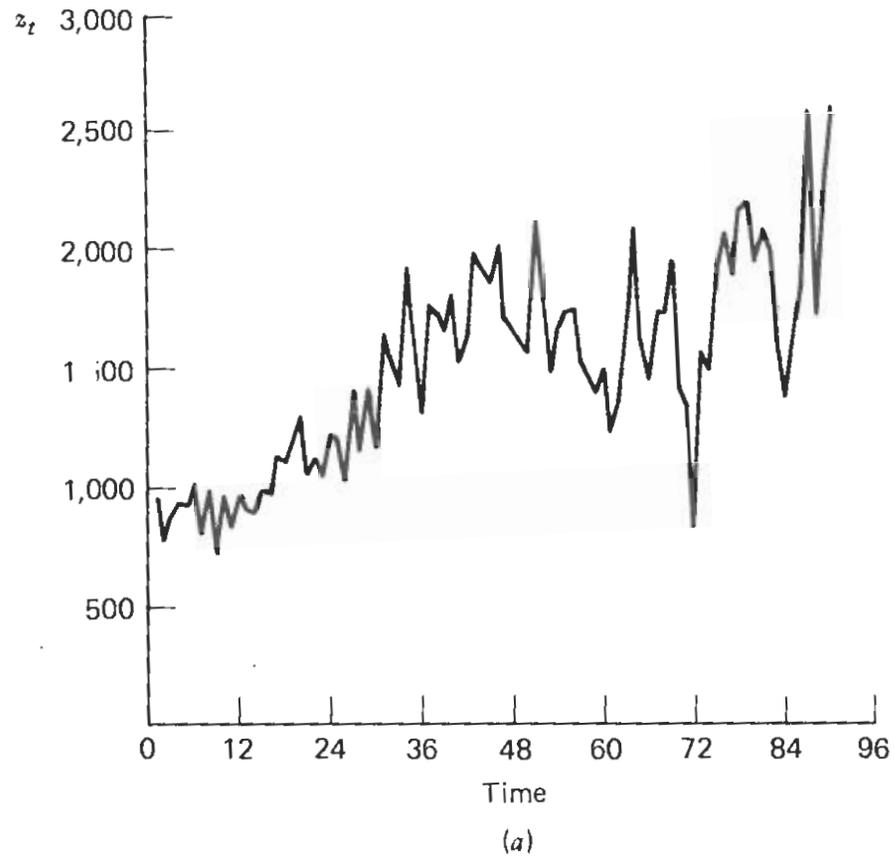
Figure 5.9. Demand for repair parts: (a) original series z_t; (b) $\ln z_t$; (c) first differences $\nabla \ln z_t$; (d) second differences $\nabla^2 \ln z_t$.
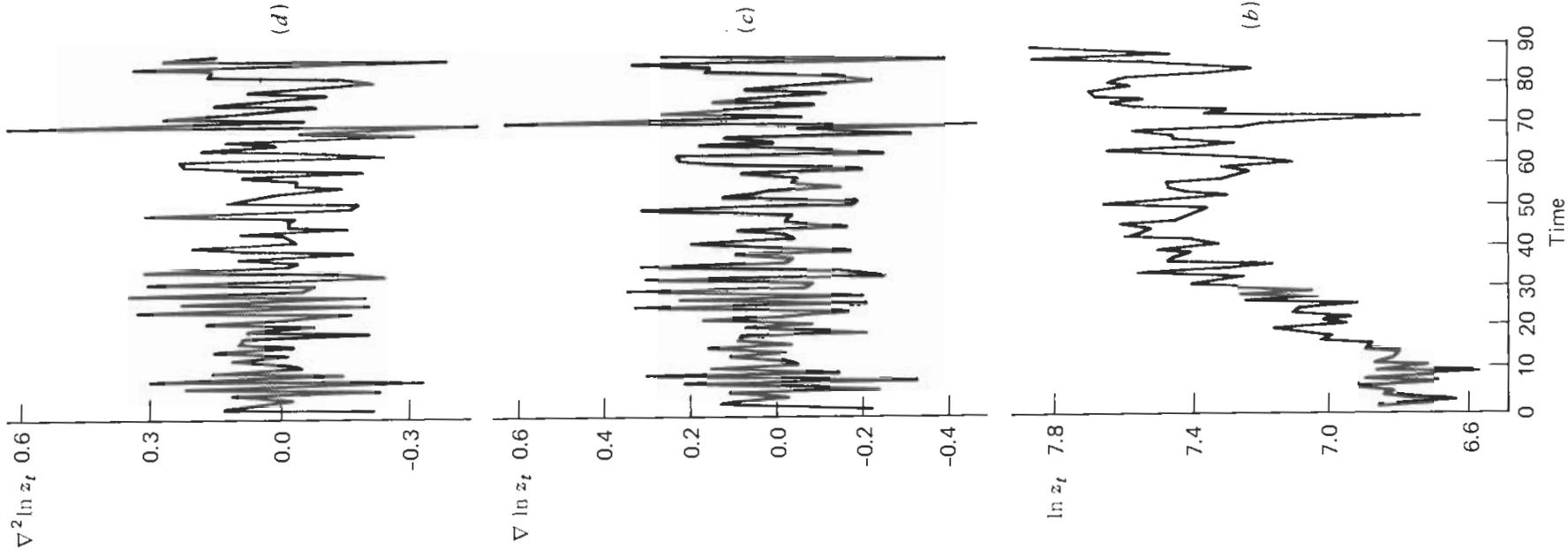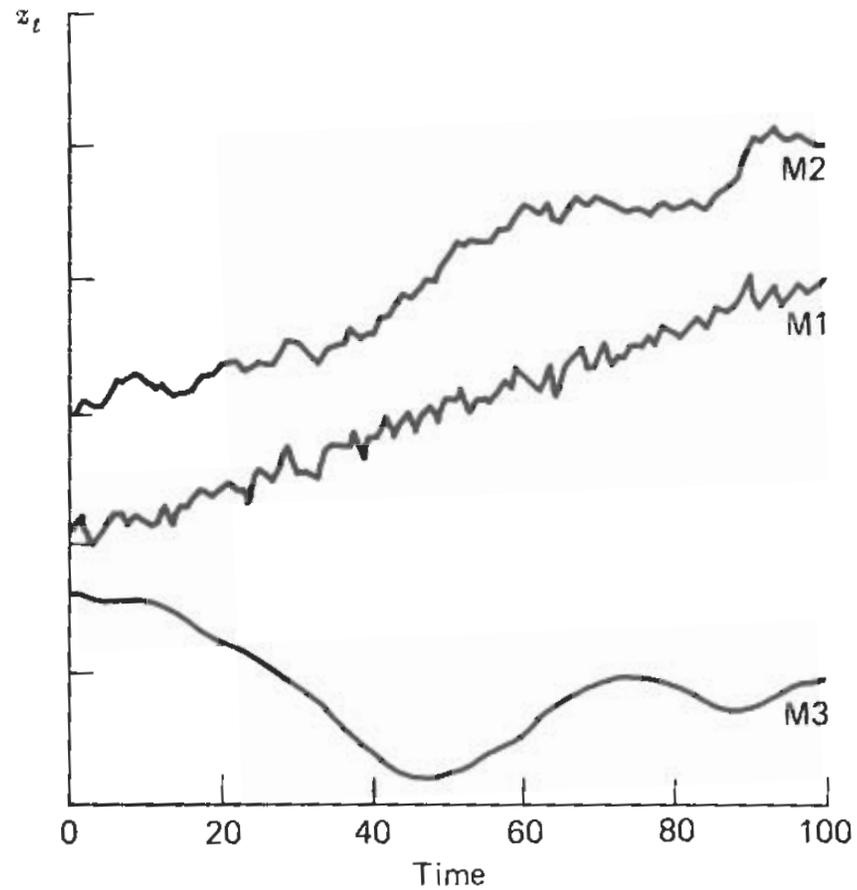
Figure 5.9. (*Continued*).

31

₂ 5.10. Sample paths for three nonstationary processes. M1: $z_t = \beta_0 + \beta_1 t + a_t$. M2: $ɜ)z_t = \beta_1 + a_t$. M3: $(1 - B)^2 z_t = a_t$.

An Alternative Argument for Differencing

Let us assume that data z_t are described by a nonstationary level

$$\mu_t = \mathrm{E}(z_t) = \beta_0 + \beta_1 t$$

and that we wish to entertain an ARMA model for a function of z_t that has a level independent of time.

- Consider the model $\phi_{p+1}(B)z_t = \theta_0 + \theta(B)a_t$

- θ_0 represent the level of $\phi_{p+1}(B)$. To achieve a θ_0 that is constant, independent of time,
$$\phi_{p+1}(B) = (1 - B)\phi(B)$$
  and

$$\theta_0 = \phi_{p+1}(B)(\beta_0 + \beta_1 t) = \phi(B)(1 - B)(\beta_0 + \beta_1) = \phi(B)\beta_1$$

- If $\mu_t = \beta_0 + \beta_1 t + \cdots + \beta_d t^d$, we have to choose the autoregressive operator in $\phi_{p+d}(B)z_t = \theta_0 + \theta(B)a_t$ such that

$$\phi_{p+d}(B) = (1 - B)d\phi(B)$$

and then

$$
\begin{aligned}
\theta_0 &= \phi_{p+d}(B)(\beta_0 + \beta_1 t + \cdots + \beta_d t^d) \\
&= \phi(B)(1 - B)^d(\beta_0 + \beta_1 t + \cdots + \beta_d t^d) \\
&= \phi(B)\beta_d d!.
\end{aligned}
$$

Autoregressive Integrated Moving Average (ARIMA) Models

- Transform a nonstationary series z_t into a stationary one by considering relevant differences

$$w_t = \nabla^d z_t = (1 - B)^d z_t.$$

- Then use the ARMA models to describe w_t. The corresponding model can be written as

$$\phi(B) w_t = \theta_0 + \theta(B) a_t$$

or

$$\phi(B)(1 - B)^d z_t = \theta_0 + \theta(B) a_t$$

- ARIMA(p,d,q) The order p, d, q are usually small. The model is called "integrated", since z_t can be thought of as the summation (integration) of the stationary series w_t. For example, if $d = 1$

$$z_t = (1 - B)^{-1} w_t = \sum_{k=-\infty}^{t} w_k$$

Choice of the proper degree of differencing

- Plot of the series

- **Examination of the SACF**
  A stationary AR(p) process requires that all G_i in

$$(1 - \phi_1 B - \cdots - \phi_p B^p) = (1 - G_1 B) \cdots (1 - G_p B)$$

are such that $|G_i| < 1$ $(i = 1, 2, \ldots, p)$, Now assume one of them, say G_1, approaches 1, that is $G_1 = 1 - \delta$, where δ is a small positive number. Then the autocorrelations

$$\rho_k = A_1 G_1^k + \cdots + A_p G_p^k \cong A_1 G_1^k.$$

So the exponential decay will be show and almost linear [i.e. $A_1 G_1^k = A_1 (1 - \delta)^k \cong A_1 (1 - \delta_k)$]

- Consider the ARMA(1,1) model

$$(1 - \phi B)z_t = (1 - \theta B)a_t$$

The ACF of this process is given by

$$\begin{aligned} \rho_1 &= \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta} \\ \rho_k &= \phi^k \rho_1 \end{aligned}$$

- If we let $\phi \rightarrow 1$, then $\rho_k \rightarrow 1$ for any finite $k > 0$. Again the autocorrelation decay very slowly when ϕ is near 1.

- It should be pointed out that the slow decay in the SACF can start at values of r_1 considerably smaller that 1.

Dangers of Overdifferencing:

Although further differences of a stationary series will again be stationary, overdifferencing can lead to serious difficulties.

- Effect of overdifferencing on the autocorrelation structure.

    - Consider the stationary MA(1) process $z_t = (1 - \theta B)a_t$.
      The autocorrelation function of this process is given by

    $$\rho_k = \begin{cases} \frac{-\theta}{1+\theta^2}, & k = 1 \\ 0, k > 1 \end{cases}$$

    - The first difference of the process is

    $$(1 - B)z_t = (1 - (1 + \theta)B + \theta B^2)a_t$$

    - and the autocorrelation of this difference is given by

    $$\rho_k = \begin{cases} \frac{-(1+\theta)^2}{2(1+\theta+\theta^2)}, & k = 1 \\ \frac{-\theta^2}{2(1+\theta+\theta^2)}, & k = 2 \\ 0, k > 2 \end{cases}$$

- Effect of overdifferencing on variance.

  - Consider the same MA(1) process. Its variance is given by

  $$\gamma_0(z) = (1 + \theta)^2 \sigma^2$$

  - The variance of the overdifferenced series $w_t = (1 - B)z_t$, which follows an MA(2) process, is given by

  $$\gamma_0(w) = 2(1 + \theta + \theta^2)\sigma^2$$

  - Hence

  $$\gamma_0(w) - \gamma_0(z) = (1 + \theta^2)\sigma^2 > 0$$

- Considered the stationary process AR(1) process $(1 - \phi B)z_t = a_t$ with variance

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}$$

- The first difference w_t follows the ARMA(1,1) process $(1 - \phi B)w_t = (1 - B)a_t$. The variance of this process is given by

$$\gamma_0(w) = \frac{2(1 - \phi)\sigma^2}{1 - \phi^2}$$

- and

$$\gamma_0(w) - \gamma_0(z) = \frac{(1 - 2\phi)\sigma^2}{1 - \phi^2}$$

Hence when $\phi < \frac{1}{2}$, overdifferencing will increasing the variance.

Regression and ARIMA Models

- ARIMA(0,1,1) model $(1 - B)z_t = (1 - \theta B)a_t$

  - Since $\psi(B) = (1 - \theta B)/(1 - B) = 1 + (1 - \theta)B + (1 - \theta)B^2 + \cdots$, we can related the variable z_{t+l} at time $t + l$ to a fixed origin t and express it as

  $$z_{t+l} = \beta_0^{(t)} + e_t(l)$$

  - where $e_t(l) = a_{t+l} + (1 - \theta)[a_{t+l-1} + \cdots + a_{t+1}]$ is a function of the random shocks that enter the system after time t, and $\beta_0^{(t)} = (1 - \theta) \sum_{j=-\infty}^{t} a_j$ depends on all the shock (observations) up to and including time t.
  - The level $\beta_0^{(t)}$ can be updated according to

  $$\beta_0^{(t+1)} = \beta_0^{(t)} + (1 - \theta)a_{t+1}$$

  - In the special case when $\theta = 1$, $e_t(l) = a_{t+l}$ and $\beta_0^{(l+1)} = \beta_0^{(t)} = \beta_0$. In this case, the model is equivalent to the constant mean model $z_{t+l} = \beta_0 + a_{t+l}$.

- ARIMA(0,2,2) model $(1 - B)^2 z_{t+l} = (1 - \theta_1 B - \theta_2 B^2) a_{t+l}$

  - Using an argument similar to that followed for ARIMA(0,1,1) model, the ARIMA(0,2,2) can be written as

$$z_{t+l} = \beta_0^{(t)} + \beta_1^{(t)} l + e_t(l)$$

  where $e_t(l)$ are correlated and given by

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}$$

  - where the ψ weights may be obtained from

$$(1 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - B)^2 = 1 - \theta_1 B - \theta_2 B^2$$

  These ψ weights can be given as $\psi_j = (1 + \theta_2) + j(1 - \theta_1 - \theta_2), j \geq 1$.

– Furthermore, the trend parameter $\beta_0^{(t)}$ and $\beta_1^{(t)}$ adapt with time according to

$$
\begin{aligned}
\beta_0^{(t+1)} &= \beta_0^{(t)} + \beta_1^{(t)} + (1 + \theta_2)a_{t+1} \\
\beta_1^{(t+1)} &= \beta_1^{(t)} + (1 - \theta_1 - \theta_2)a_{t+1}
\end{aligned}
$$

– Again the model above can be thought of as a generalization of the deterministic trend regression model

$$
z_{t+l} = \beta_0 + \beta_1 l + a_{t+l}
$$

– In the special case when $\theta_1 \to 2$ and $\theta_2 \to -1$, the two models above are equivalent
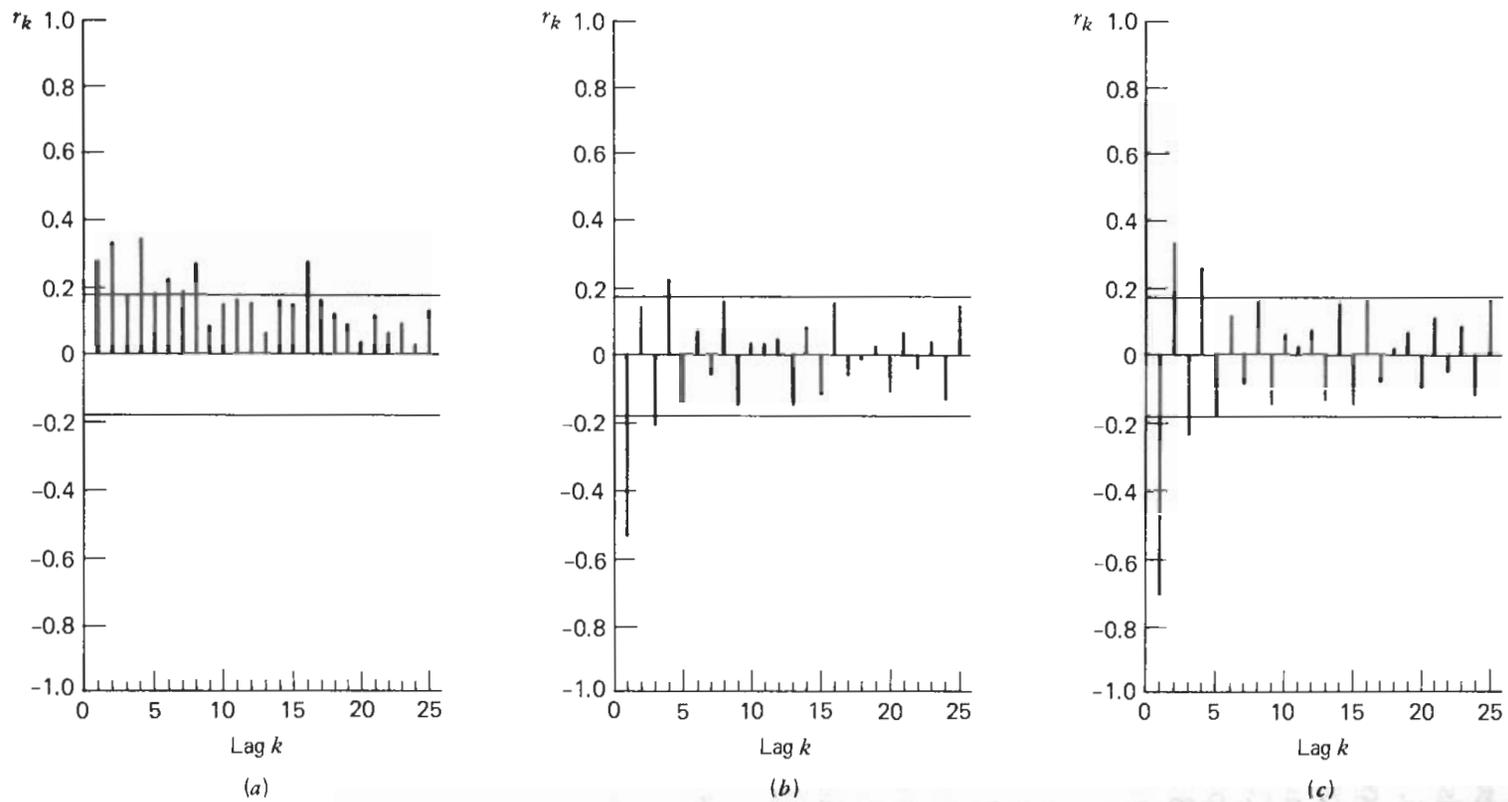
Figure 5.11. SACF for growth rates of Iowa nonfarm income: (*a*) original series z_t; (*b*) first differences ∇z_t; (*c*) second differences $\nabla^2 z_t$.

44
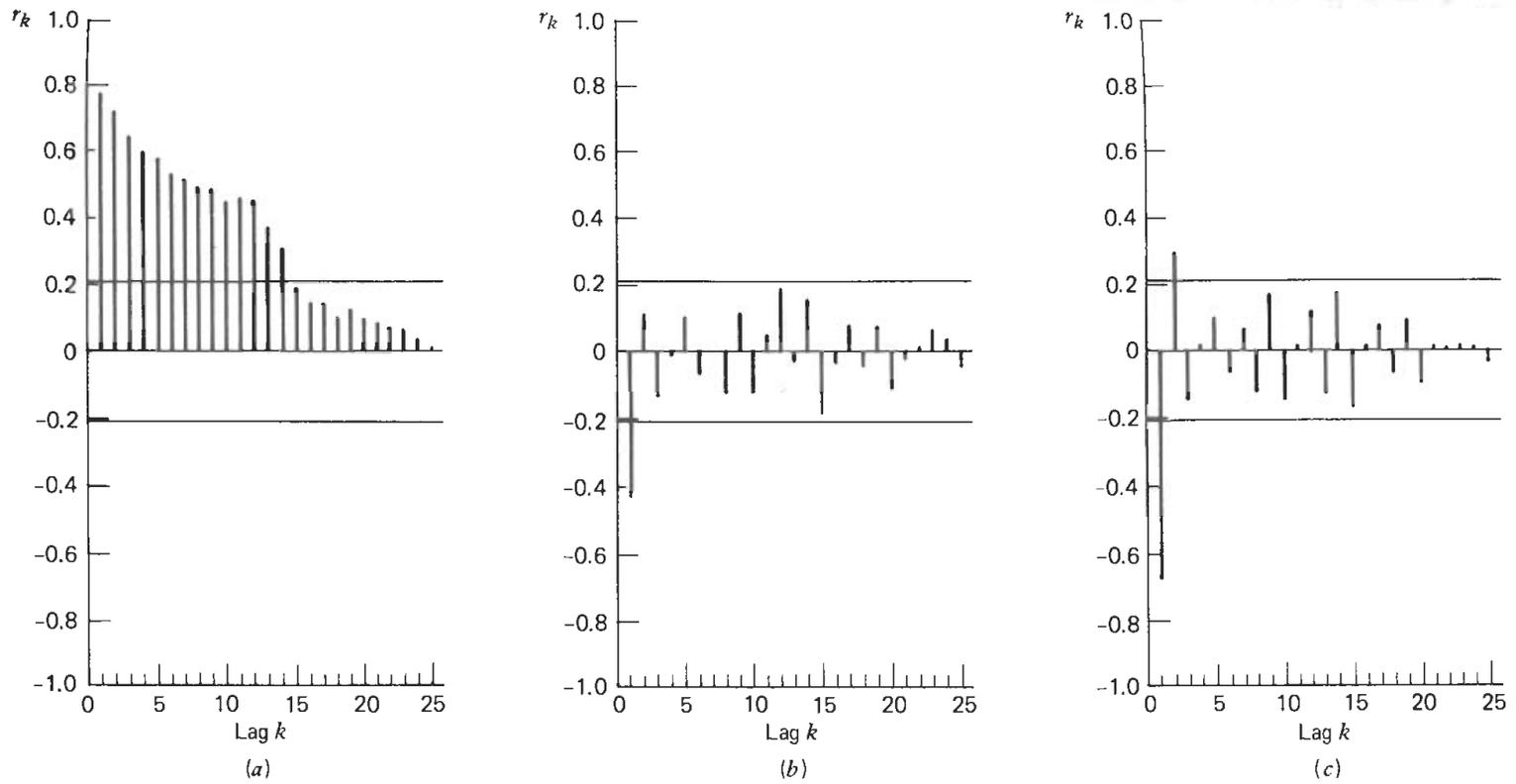
Figure 5.12. SACF for demand for repair parts: (a) $\ln z_t$; (b) first differences $\nabla \ln z_t$; (c) second differences $\nabla^2 \ln z_t$.

4.4 Forecasting

- ARIMA model $\phi(B)(1-B)^d z_t = \theta(B)a_t$

- Autoregressive representation

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-2} + \cdots + a_t$$

where the π weights are given by

$$1 - \pi B - \pi B^2 - \cdots = \frac{\phi(B)(1-B)^d}{\theta(B)}$$

- l-step forecast

$$z_l = \eta_0 z_n + \eta_1 z_{n-1} + \eta_2 z_{n-2} + \cdots$$

  or

$$z_l = \xi_0 a_n + \xi_1 a_{n-1} + \xi_2 a_{n-2} + \cdots$$

Minimum mean square error (MMSE) forecasts: The mean square error $\mathrm{E}[z_{n+l} - z_n(l)]^2$ is minimized.

- Using the ψ weights in $\psi(B) = \theta(B)\phi^{-1}(B)(1-B)^{-d}$ to write the model in its linear filter representation

$$z_{n+l} = a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} + \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots$$

- Then the mean square error is expressed as

$$
\begin{aligned}
\mathrm{E}[z_{n+l} - z_n(l)] &= \mathrm{E}[a_{n+l} + a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} + (\psi_l - \xi_0)a_n \\
&\quad + (\psi_{l+1} - \xi_1)a_{n-1} + \cdots]^2 \\
&= (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma^2 + \sum_{j=0}^{\infty}(\psi_{l+j} - \xi_j)^2 \sigma^2.
\end{aligned}
$$

- The minimum mean square error (MMSE) forecast

$$z_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots$$

- The forecast error is given by

$$e_n(l) = z_{n+l} - z_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1}$$

and its variance by

$$\text{Var}[e_n(l)] = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)$$

-

$$E(a_{n+j}|z_n, z_{n-1}, \ldots) = \left\{ \begin{array}{ll} a_{n+j} & j \leq 0 \\ 0 & j > 0 \end{array} \right.$$

Hence

$$\begin{aligned} E(z_{n+l}|z_n, z_{n-1}, \ldots) &= E[(a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} \\ &\quad + \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots)|z_n, z_{n-1}, \ldots] \\ &= \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots \end{aligned}$$

Examples:

- AR(1) process: $z_t - \mu = \phi(z_{t-1} - \mu) + a_t$

- AR(2) process: $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$

- ARIMA(0,1,1) process: $z_t = z_{t-1} + a_t - \theta a_{t-1}$

- ARIMA(1,1,1) process: $(1 - \phi B)(1 - B)z_t = \theta_0 + (1 - \theta B)a_t$ or

$$z_t = \theta_0 + (1 + \phi)z_{t-1} - \phi z_{t-2} + a_t - \theta a_{t-1}$$

- ARIMA(0,2,2) process: $(1 - B)^2 z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$ or

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Prediction Limits

- It is usually not sufficient to calculate point forecast alone. The uncertainty associated with a forecast should be expressed, and the prediction limits should always be computed.

- It is possible, but not very easy, to calculate a simultaneous "prediction region" for a set of future observations. Therefore, we give probability limits only for individual forecasts.

- Assuming that a_t's (and hence the z's) are normally distributed, we can calculate $100(1 - \lambda)$ percent prediction limits form

$$z_n(l) \pm u_{\lambda/2}\{V(e_n(l))\}^{1/2}$$

where $u_{\lambda/2}$ is the $100(1 - \lambda/2)$ percentage point of standard normal distribution.

- Because $z_n(l)$ and $V(e_n(l))$ are unknown exactly, and need estimate. Hence the estimated prediction limits are given by

$$\hat{z}_n(l) \pm u_{\lambda/2}\{\hat{V}(e_n(l))\}^{1/2}$$

- Consider AR(1) process for Yield data, we found $\hat{z}_{156}(1) = 0.56$, $\hat{z}_{156}(2) = 0.62$, $\hat{z}_{156}(3) = 0.68$, $\hat{V}[e_{156}(1)] = 0.024$, $\hat{V}[e_{156}(2)] = 0.041$, $\hat{V}[e_{156}(1)] = 0.054$. Thus the 95 percent prediction limits for $z_{156+l}, l = 1, 2, 3$ are given by

$$z_{157} : \hat{z}_{156}(1) \pm 1.96\{\hat{V}[e_{156}(1)]\}^{1/2} = .56 \pm 1.96(.15) = .56 \pm .29$$

$$z_{158} : \hat{z}_{156}(2) \pm 1.96\{\hat{V}[e_{156}(2)]\}^{1/2} = .62 \pm 1.96(.20) = .62 \pm .39$$

$$z_{159} : \hat{z}_{156}(3) \pm 1.96\{\hat{V}[e_{156}(3)]\}^{1/2} = .68 \pm 1.96(.23) = .68 \pm .45$$

Forecast Updating

- Suppose we are at time n and we are predicting $l + 1$ steps ahead. The forecast is given by

$$z_n(l + 1) = \psi_{l+1}a_n + \psi_{l+2}a_{n-1} + \cdots$$

- After z_{n+1} has become available, we need to update our prediction of z_{n+l+1}. On the other hand $z_{n+1}(l)$ can be written as

$$z_{n+l} = \psi_l a_{n+1} + \psi_{l+1}a_n + \psi_{l+2}a_{n-1} + \cdots$$

- we find that

$$z_{n+1}(l) = z_n(l + 1) + \psi_l a_{n+1} = z_n(l + 1) + \psi_l(z_{n+1} - z_n(l))$$

- The updated forecast is a linear combination of the previous forecast of z_{n+l+1} made at time n and the most recent one-step-ahead forecast error $e_n(l) = z_{n+1} - z_n(l) = a_{n+1}$.

Use Yield data example as shown as before. We already obtained the forecast for z_{157}, z_{158}, z_{159} from time origin $n = 156$. Suppose now that $z_{157} = 0.74$ is observed and we need to update the forecasts for z_{158} and z_{159}.

- Since the model under consideration is AR(1) with the parameter estimate $\hat{\phi} = 0.85$, the estimated ψ weights are given by $\hat{\psi}_j = \hat{\phi}^j = 0.85^j$.

- Then the updated forecasts are given by

$$z_{158} : \hat{z}_{156}(2) + \hat{\psi}_1[z_{157} - \hat{z}_{156}(1)] = 0.62 + 0.85(0.74 - 0.56) = 0.77$$

$$z_{159} : \hat{z}_{156}(3) + \hat{\psi}_2[z_{157} - \hat{z}_{156}(1)] = 0.62 + 0.85^2(0.74 - 0.56) = 0.81$$

- The updated 95 percent prediction limits for z_{158} and z_{159} become

$$z_{158} : \hat{z}_{157}(1) \pm 1.96\{\hat{V}[e_{157}(1)]\}^{1/2} = .77 \pm .29$$

$$z_{159} : \hat{z}_{157}(2) \pm 1.96\{\hat{V}[e_{157}(2)]\}^{1/2} = .81 \pm .39$$

4.5 Model Specification

Step 1. **Variance-stabilizing transformations.** If the variance of the series changes with the level, then a logarithmic transformation will often stabilize the variance. If the logarithmic transformation still not stabilize the variance, a more general approach and employ the class of power transformations can be adopted.

Step 2. **Degree of differencing.** If the series or its appropriate transformation is not mean stationary, then the proper degree of differencing has to determined. For the series and its differences, we can examine

1. Plots of the time series
2. Plot of the SACF
3. Sample variances of the successive differences

Step 3. **Specification of p and q.** Once we obtain a stationary difference we must specify the orders of the autoregressive (p) and moving average (q) polynomials. The orders can be specified by matching the patterns in the sample autocorrelations and partial autocorrelations with the theoretical patterns of known models (See Table 5.3).

Step 4. **Inclusion of a trend parameter.** If the series requires differencing, we should check whether it is necessary to include a deterministic trend θ_0 in the model. This can be done by comparing the sample mean $\bar{w}$ of the stationary difference with its standard error $s(\bar{w})$. This standard error can be approximated by

$$s(\bar{w}) \cong \left[\frac{c_0}{n}(1 + 2r_1 + 2r_2 + \cdots + 2r_k) \right]^{1/2}$$

where c_0 is the sample variance and $r_1, \cdots, r_k$ are the first K significant sample autocorrelations of the stationary differences. This approximation is appropriate only if $s(\bar{w}) > 0$.

Table 5.3. Properties of the ACF and the PACF for Various ARMA Models

Model	ACF	PACF
$(1, d, 0)$ AR(1)	Exponential or oscillatory decay	$\phi_{kk} = 0$ for $k > 1$
$(2, d, 0)$ AR(2)	Exponential or sine wave decay	$\phi_{kk} = 0$ for $k > 2$
$(p, d, 0)$ AR(p)	Exponential and/or sine wave decay	$\phi_{kk} = 0$ for $k > p$
$(0, d, 1)$ MA(1)	$\rho_k = 0$ for $k > 1$	Dominated by damped exponential
$(0, d, 2)$ MA(2)	$\rho_k = 0$ for $k > 2$	Dominated by damped exponential or sine wave
$(0, d, q)$ MA(q)	$\rho_k = 0$ for $k > q$	Dominated by linear combination of damped exponentials and/or sine waves
$(1, d, 1)$ ARMA(1, 1)	Tails off. Exponential decay from lag 1	Tails off. Dominated by exponential decay from lag 1
(p, d, q) ARMA(p, q)	Tails off after $q - p$ lags. Exponential and/or sine wave decay after $q - p$ lags	Tails off after $p - q$ lags. Dominated by damped exponentials and/or sine waves after $p - q$ lags

4.6^* Model Estiamtion

After specifying the form of the model, it is necessary to estimate its parameters. Let $(z_1, z_2, \ldots, z_N)$ represent the vector of the original observations and $\mathbf{w} = (w_1, \ldots, w_n)'$ the vector of the $n = N - d$ stationary differences.

Maximum Likelihood Estimates

- The ARIMA model can be written as

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p}$$

- The joint probability density function of $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$ is given by

$$p(\mathbf{a}|\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^{n} a_t^2\right]$$

- The joint probability density function of $\mathbf{w}$ can be written down, at least in principle. It is of the form

$$L(\phi, \theta, \sigma^2 | \mathbf{w}) = g_1(\phi, \theta, \sigma^2) \exp\left[-\frac{1}{2\sigma^2} S(\phi, \theta)\right]$$

where g_1 is a function of the parameter (ϕ, θ, σ^2) and

$$S(\phi, \theta) = \sum_{t=1-p-q}^{n} \mathrm{E}^2(u_t | \mathbf{w})$$

$\mathrm{E}(u_t | \mathbf{w}) = \mathrm{E}(u_t | \mathbf{w}, \phi, \theta, \sigma^2)$ is the conditional expectation of u_t given $\mathbf{w}, \phi, \theta$ and σ^2, and

$$u_t = \begin{cases} a_t, & t = 1, 2, \ldots, n \\ g_2(\mathbf{a}_*, \mathbf{w}_*), & t \leq 0 \end{cases}$$

where g_2 is a linear function of the initial unobservable values $\mathbf{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$ and $w_* = (w_{1-p}, \ldots, w_{-1}, w_0)'$.

Exact likelihood function for three important models

1. AR(1): $a_t = w_t - \phi w_{t-1}$

$$L(\phi, \sigma^2|\mathbf{w}) = (2\pi\sigma^2)^{-n/2}(1-\phi^2)^{1/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[(1-\phi^2)w_1^2 + \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2\right]\right\}$$

2. MA(1): $a_t = \theta a_{t-1} + w_t$

$$L(\theta, \sigma^2|\mathbf{w}) = (2\pi\sigma^2)^{-n/2}\left[\frac{1-\theta^2}{1-\theta^{2(n+1)}}\right]^{1/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{t=0}^{n}\mathrm{E}^2(a_t|\mathbf{w})\right]$$

3. ARMA(1,1): $a_t = \theta a_{t-1} + w_t - \phi w_{t-1}$

$$L(\phi, \theta, \sigma^2|\mathbf{w}) = (2\phi\sigma^2)^{-n/2}|\mathbf{Z'Z}|^{-1/2}\exp\left[-\frac{1}{2\sigma^2}S(\phi, \theta)\right]$$

where $$|\mathbf{Z'Z}| = \frac{(1-\theta^2)(1-\phi^2) + (1-\theta^{2n})(\theta-\phi)^2}{(1-\theta^2)(1-\phi^2)}$$

and

$$S(\phi, \theta) = \mathrm{E}^2(a_0|\mathbf{w}) + \frac{1-\phi^2}{(\theta-\phi)^2}[\mathrm{E}(w_0|\mathbf{w}) - \mathrm{E}(a_0|\mathbf{w})]^2 + \sum_{t=1}^{n}\mathrm{E}^2(a_t|\mathbf{w})$$

Maximum likelihood Estimate (MLE) of parameters (ϕ, θ, σ^2) for the ARIMA model can be obtained by maximizing the likelihood function. In general, closed-form solutions cannot be found. However, various algorithms are available to compute the MLE's or close approximations.

1. AR(1): Difficult to get closed-form solution, and iterative methods have to be employed.

2. MA(1): Need to calculate $\mathrm{E}(a_t|\mathbf{w}), t = 1, 2, \ldots, n$. These can be calculated using the recurrence relation

$$\mathrm{E}(a_t|\mathbf{w}) = \theta \mathrm{E}(a_{t-1}|\mathbf{w}) + w_t, t = 1, 2, \ldots, n$$

provide $\mathrm{E}(a_0|\mathbf{w})$. The difficulties of obtaining the MLE are thus twofold: (1) Calculate $\mathrm{E}(a_0|\mathbf{w})$ and (2) the likelihood function is a complicated function of θ.

3. ARMA(1,1) Similar as MA(1), to calculate the likelihood function, we need $\mathrm{E}(a_0|\mathbf{w})$ and $\mathrm{E}(w_0|\mathbf{w})$. From these we can calculate

$$\mathrm{E}(a_t|\mathbf{w}) = \theta \mathrm{E}(a_{t-1}|\mathbf{w}) + w_t - \phi \mathrm{E}(w_{t-1}|\mathbf{w})$$

In summary, there are two major difficulties with maximum likelihood estimation approach.

1. The present of the function $g_1(\phi, \theta, \sigma^2)$ in likelihood function makes the maximization difficult.

2. This approach requires the calculation of the conditional expectations $\mathrm{E}(\mathbf{u}_*|\mathbf{w})$, where $\mathbf{u}_* = (u_{1-p-q}, \ldots, u_{-1}, u_0)'$. This in turn requires the calculation of

$$\mathrm{E}(a_t|\mathbf{w}), t = 1 - q, \ldots, -1, 0 \quad \text{and} \quad \mathrm{E}(w_t|\mathbf{w}), t = 1 - p, \ldots, -1, 0$$

Unconditional Least Squares Estimates

Many approximations have been considered in the literature to ease these difficulties. A common approximation is to ignore the function $g_1(\phi, \theta, \sigma^2)$ and maximize $\exp[-(1/2\sigma^2)S(\phi, \theta)]$, or equivalently, minimize $S(\phi, \theta)$. The resulting estimates are called *least squares estimates* (LSE).

- This approximation is satisfactory if the parameters are not close to the invertibility boundaries. In this case it can be shown that the likelihood function is dominated by the exponential part and the removal of g_1 has only negligible effect.

- Even in the minimization of $S(\phi, \theta)$ there are two major difficulties.

  1. the evaluation of initial expectations as shown before.
  2. $S(\phi, \theta)$ can be a complicated (not necessarily quadratic) function of the parameters.

Evaluation of the Initial Expectation

There are two methods available for the evaluation of $\mathrm{E}(u_t|\mathbf{w})(t \leq 0)$, the method of least squares and backforecasting.

Method of least squares.

For the ARMA(1,1) model, it is true in general that $\mathrm{E}(\mathbf{u}_*|\mathbf{w}) = \hat{\mathbf{u}}_*$ where $\hat{\mathbf{u}}_*$ is the least squares estimator of $\mathbf{u}_*$ in

$$\mathbf{Lw} = -\mathbf{Zu}_* + \mathbf{u}$$

where $\mathbf{L}$ and $\mathbf{Z}$ are $(n+p+q) \times n$ and $(n+p+q) \times (p+q)$ matrices involving the parameters ϕ and θ (For detail See Appendix 5 in the textbook).

For given values of (ϕ, θ), we can calculate $\hat{\mathbf{u}}_*$, and hence $S(\phi, \theta)$. Then we can search for a minimum value of $S(\phi, \theta)$ in the space of (ϕ, θ)

Method of backforecasting.

- The MMSE forecast of z_{n+l} made at time n is given by the conditional expectation $\mathrm{E}(z_{n+l} | z_n, z_{n-1}, \ldots)$. This result may be used to compute the unknown quantities $\mathrm{E}(a_t | \mathbf{w})(t = 1 - q, \ldots, -1, 0)$ and $\mathrm{E}(w_t | \mathbf{w})(t = 1 - p, \ldots, -1, 0)$.

- The MMSE forecast of a_t and $w_t (t \leq 0)$ can be found by considering the "backward" model

$$ e_t = \theta_1 e_{t+1} + \theta_2 e_{t+2} + \cdots + \theta_q e_{t+q} + w_t - \phi_1 w_{t+1} - \cdots \phi_p w_{t+p}. $$

The ACF of this backward model is same as the original model. Furthermore, e_t is a white-noise sequence with $\mathrm{Var}(e_t) = \sigma^2$.

- This method of evaluating $\mathrm{E}(a_t | \mathbf{w})$ and $\mathrm{E}(w_t | \mathbf{w})$ is called *backforecasting* (or backcasting).

The method of backforecasting with a simple MA(1) process $a_t = \theta a_{t-1} + w_t$

- This process can be expressed in the backward form as

$$e_t = \theta e_{t+1} + w_t$$

- When the process is going backwards, then $e_0, e_{-1}, \ldots$ are random shocks that are "future" to the observations $w_n, w_{n-1}, \ldots, w_1$. Hence,

$$\mathrm{E}(e_t|\mathbf{w}) = 0, \ t \leq 0$$

- Furthermore, since the MA(1) process has a memory of only one period, the correlation between a_{-1} and $w_1, w_2, \ldots, w_n$ is zero, and

$$\mathrm{E}(a_t|\mathbf{w}) = 0, \ t \leq -1$$

- It is easily seen that

$$E(w_i | \mathbf{w}) = w_i, i = 1, 2, \ldots, n$$

- However the computation of $E(w_0 | \mathbf{w})$ is different.

$$E(e_0 | \mathbf{w}) = \theta E(e_1 | \mathbf{w}) + E(w_0 | \mathbf{w}) \Rightarrow E(w_0 | \mathbf{w}) = -\theta E(e_1 | \mathbf{w})$$

since $E(e_0 | \mathbf{w}) = 0$.

- Start at the end of series and assume that $E(e_{n+1} | \mathbf{w}) = 0$, under this assumption

$$
\begin{aligned}
E(e_n | \mathbf{w}) &= \theta E(e_{n+1} | \mathbf{w}) + w_n \\
E(e_{n-1} | \mathbf{w}) &= \theta E(e_n | \mathbf{w}) + w_{n-1} \\
&\vdots \\
E(e_1 | \mathbf{w}) &= \theta E(e_2 | \mathbf{w}) + w_1
\end{aligned}
$$

- with the value of $\mathrm{E}(w_0|\mathbf{w})$, same as before we can compute $\mathrm{E}(a_t|\mathbf{w})(t = 1, 2, \ldots, n)$

$$
\begin{aligned}
\mathrm{E}(a_0|\mathbf{w}) &= \theta \mathrm{E}(a_{-1}|\mathbf{w}) + \mathrm{E}(w_0|\mathbf{w}) \\
\mathrm{E}(a_1|\mathbf{w}) &= \theta \mathrm{E}(a_{0|\mathbf{w}}) + w_1 \\
&\vdots \\
\mathrm{E}(a_n|\mathbf{w}) &= \theta \mathrm{E}(a_{n-1}|\mathbf{w}) + w_n
\end{aligned}
$$

- The sum of squares $S(\theta)$ for any θ can be obtained by

$$
S(\theta) = \sum_{t=0}^{n} \mathrm{E}^2(a_t|\mathbf{w}),
$$

and its minimum can be found.

Table 5.4. Computation of $S(\theta)$ Through Backforecasting

t	z_t	$w_t = \nabla z_t$	$E(w_t|\mathbf{w})$	$E(e_t|\mathbf{w})$	$E(a_t|\mathbf{w})$
0	1.84		-0.88	0	-0.88
1	3.42	1.58	1.58	1.10	0.88
2	2.62	-0.80	-0.80	-0.60	-0.10
3	3.02	0.40	0.40	0.25	0.32
4	2.76	-0.26	-0.26	-0.19	0.00
5	2.16	-0.60	-0.60	0.09	-0.60
6	2.32	0.16	0.16	0.86	0.32
7	2.59	0.27	0.27	0.88	0.01
8	3.24	0.65	0.65	0.76	0.66
9	3.38	0.14	0.14	0.14	0.67
10				0	

Conditional Least Squares Estimates

Computationally simpler estimates can be obtained by minimizing the conditional sum of squares $S_c(\phi, \theta) = \sum_{t=p+1}^{n} a_t^2$, where the starting values $a_p, a_{p-1}, \ldots, a_{p+1-q}$ in the calculation of the appropriate a_t's are set equal to zero. The resulting estimates are called ***conditional least squares estimates*** (CLSE).

- In the AR(p) case, the CLSE is obtained by minimizing

$$S_c(\phi) = \sum_{t=p+1}^{n} a_t^2 = \sum_{t=p+1}^{n} (w_t - \phi_1 w_{t-1} - \cdots - \cdots - \phi_p w_{t-p})^2$$

which leads to the least square estimate

$$\hat{\phi}_c = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y}$$

where

$$\mathbf{y} = (w_{p+1}, w_{p+2}, \ldots, w_n)'$$

and

$$\boldsymbol{X} = \begin{bmatrix} w_p & w_{p-1} & \cdots & w_1 \\ w_{p+1} & w_p & \cdots & w_2 \\ \vdots & \vdots & & \vdots \\ w_{n-1} & w_{n-2} & \cdots & w_{n-p} \end{bmatrix}$$

- The special case of AR(1) process

Nonlinear Estimation

If the model contains moving average terms, then the conditional and unconditional sums of squares are not quadratic functions of the parameters (ϕ, θ). This is because $\mathrm{E}(u_t | \mathbf{w})$ in $S(\phi, \theta)$ is nonlinear in the parameters. Hence nonlinear least squares procedures must be used to minimize $S(\phi, \theta)$ or $S_c(\phi, \theta)$.

The nonlinear estimation procedure and drive the unconditional least squares estimates of the parameters in the ARMA(1,1) model

$$a_t = \theta a_{t-1} + w_t - \phi w_{t-1}$$

Step 1. Computation at the starting point. For given $\mathbf{w}$ and a starting point $(\phi, \theta) = (\phi_0, \theta_0)$, using backforecasting or the method of least squares, we can get

$$\hat{u}_{-10} = \mathrm{E}(a_0 | \mathbf{w}, \phi_0, \theta_0)$$

and

$$\hat{u}_{00} = \frac{(1 - \phi_0)^{1/2}}{\theta_0 - \phi_0}[\mathrm{E}(w_0 | \mathbf{w}, \phi_0, \theta_0) - \mathrm{E}(a_0 | \mathbf{w}, \phi_0, \theta_0)]$$

The remaining elements in $\hat{\mathbf{u}}_0 = (\hat{u}_{-10}, \hat{u}_{00}, \hat{u}_{10}, \ldots, \hat{u}_{n0})'$ can be obtained from $\hat{u}_{t0} = \mathrm{E}(a_t | \mathbf{w}, \phi_0, \theta_0)$ for $t = 1, 2, \ldots, n$.

Step 2. Linearization. Regard $\hat{u}_t = \mathrm{E}(a_t | \mathbf{w}, \phi, \theta)(t = -1, 0, 1, \cdots, n)$ as a function of $\boldsymbol{\beta} = (\phi, \theta)'$ and linearize this function at the starting point $\boldsymbol{\beta}_0 = (\phi_0, \theta_0)'$ in a first-order Taylor series expansion:

$$
\begin{aligned}
\hat{u}_t &\cong \hat{u}_{t0} + (\phi - \phi_0)\left.\frac{\partial \hat{u}_t}{\partial \phi}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + (\theta - \theta_0)\left.\frac{\partial \hat{u}_t}{\partial \theta}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= \hat{u}_{t0} - x_{1,t}(\phi - \phi_0) - x_{2,t}(\theta - \theta_0)
\end{aligned}
$$

where

$$x_{1,t} = -\left.\frac{\partial \hat{u}_t}{\partial \phi}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \quad \text{and} \quad x_{2,t} = \left.\frac{\partial \hat{u}_t}{\partial \theta}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

So we have

$$\hat{u}_{t0} = (x_{1,t}, x_{2,t}) \begin{bmatrix} \phi - \phi_0 \\ \theta - \theta_0 \end{bmatrix} + \hat{u}_t$$

Considering this equation for $t = -1, 0, \ldots, n$, we obtain

$$\hat{\mathbf{u}}_0 = \boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \hat{\mathbf{u}}$$

where $\hat{\mathbf{u}} = (\hat{u}_{-1}, \hat{u}_0, \ldots, \hat{u}_n)'$ and

$$\boldsymbol{X} = \begin{bmatrix} x_{1,-1} & x_{2,-1} \\ x_{1,0} & x_{2,0} \\ \vdots & \vdots \\ x_{1,n} & x_{2,n} \end{bmatrix}$$

The least squares estimate of $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ may be obtained from $\hat{\boldsymbol{\delta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\mathbf{u}}_0$. Hence we obtain a modified estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_0 + \hat{\boldsymbol{\delta}}$$

Step 3. Repeat step 1 and step 2 with $\hat{\beta}_1$ as the new point of expansion in the Taylor series and continue the process until convergence occurs. Convergence is usually determined by one of the following criteria:

1. The relative change in the sum of squares $\sum_{t=-1}^{n} \hat{u}_t^2$ from two consecutive iteration is less than a prespecified small value.

2. The largest relative change in the parameter estimate of two consecutive iterations is less than a prespecified small value.

When convergence occurs, we have an estimate $\hat{\beta}$ based on local linear least squares, and this will be taken as the unconditional least squares estimate of $\hat{\beta}$. The conditional least squares estimate can be obtained in a similar fashion by replacing the vector $\hat{\mathbf{u}}$ by $\mathbf{a} = (a_2, a_3, \ldots, a_n)$ and setting $a_1 = 0$.

Standard Errors of the Estimates

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\boldsymbol{X}'_{\hat{\beta}} \boldsymbol{X}_{\hat{\beta}})^{-1}$$

where the estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n}$$

Possible Difficulties in Nonlinear Estimation

- In some cases the convergence to the final estimates can be rather slow. This can be due to poorly chosen Starting values of $\boldsymbol{\beta}_0$ or to misspecified models with too many parameters.

- In some rare case the likelihood function can have more than one maximum, and the estimates can converge to a local maximum rather than a global one.

4.7 Model Checking

- Analysis of residuals $\hat{a}_t$

  1. the mean of the residuals should be close to zero,
  2. the variance of the residuals should be approximately constant,
  3. the autocorrelation of the residuals should be negligible.

- The sample autocorrelation of the residual

$$r_{\hat{a}}(k) = \frac{\sum\limits_{t=k+1}^{n} (\hat{a}_t - \bar{\hat{a}})(\hat{a}_{t-k} - \bar{\hat{a}})}{\sum\limits_{t=1}^{n} (\hat{a}_t - \bar{\hat{a}})^2}$$

  The standard error of $r_{\hat{a}}(k)$ is usually approximated as

$$s[r_{\hat{a}(k)}] \cong n^{-1/2}$$

- For small k, the true standard error can be much smaller. In general, the true standard error depend on

  1. the form of the fitted model,
  2. the true parameter values,
  3. the value of k

Portmanteau Test

- Under null hypothesis of model adequacy, the large-sample distribution of $\mathbf{r}_{\hat{a}}$ is multivariate normal and that

$$Q^* = n \sum_{k=1}^{K} r_{\hat{a}}^2(k)$$

  has a large-sample chi-square distribution with $K - p - q$ degrees of freedom.

- For small sample size, this test is quite conservative; that is, the chance of incorrectly rejecting the null hypothesis of model adequacy is smaller than the chosen significance level. The modified test statistics is

$$Q = n(n+2) \sum_{k=1}^{K} \frac{r_{\hat{a}}^2(k)}{n-k}$$

4.8 Examples

Yield data

This series consists of monthly differences between the yield on mortgages and the yield on government loans in the Netherlands from January 1961 to December 1973. The observations for January to March 1974 are also available, but they are held back to illustrating the forecasting and updating.

- Figures: Time series plot and SACF

- AR(1) model

$$z_t - \mu = \phi(z_{t-1} - \mu) + a_t$$

The maximum likelihood estimates are

$$\hat{\mu} = 0.97(0.08), \quad \hat{\phi} = 0.85(0.04), \quad \hat{\sigma}^2 = 0.024$$
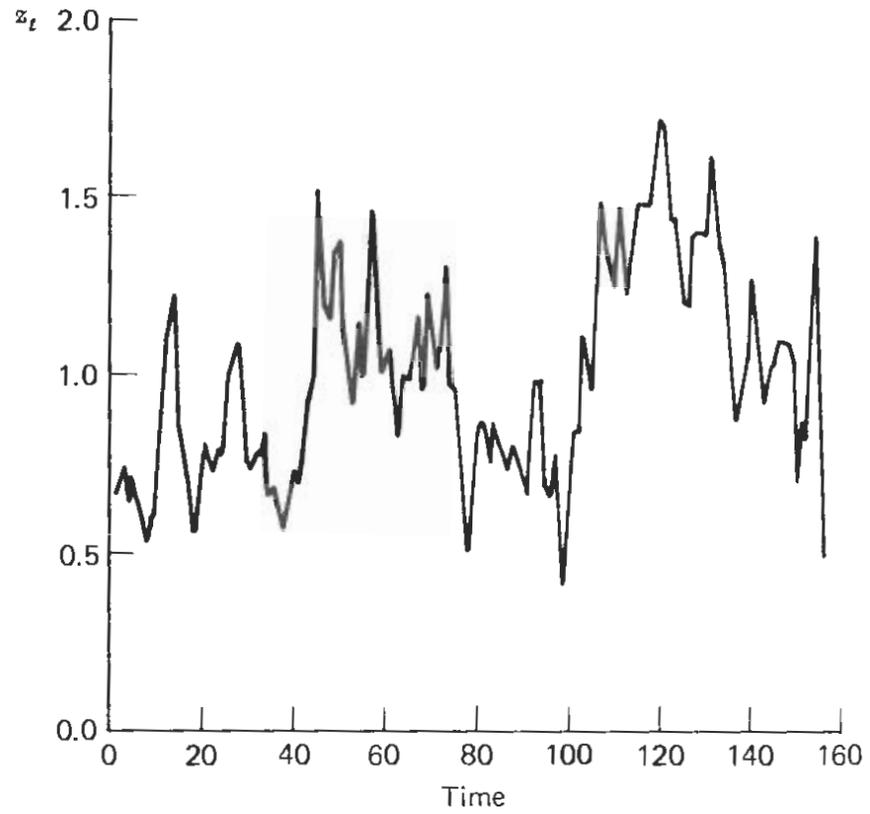
Figure 5.13. Yield series.

- SACF of the residuals

- the Q statistics. Suppose take $K = 24$. Then

$$Q = 156{\times}158 \left(\frac{(.09)^2}{155} + \frac{(-.10)^2}{154} + \cdots + \frac{(.06)^2}{132} \right) = 22.4 < \chi^2_{0.05}(22) = 33.9$$
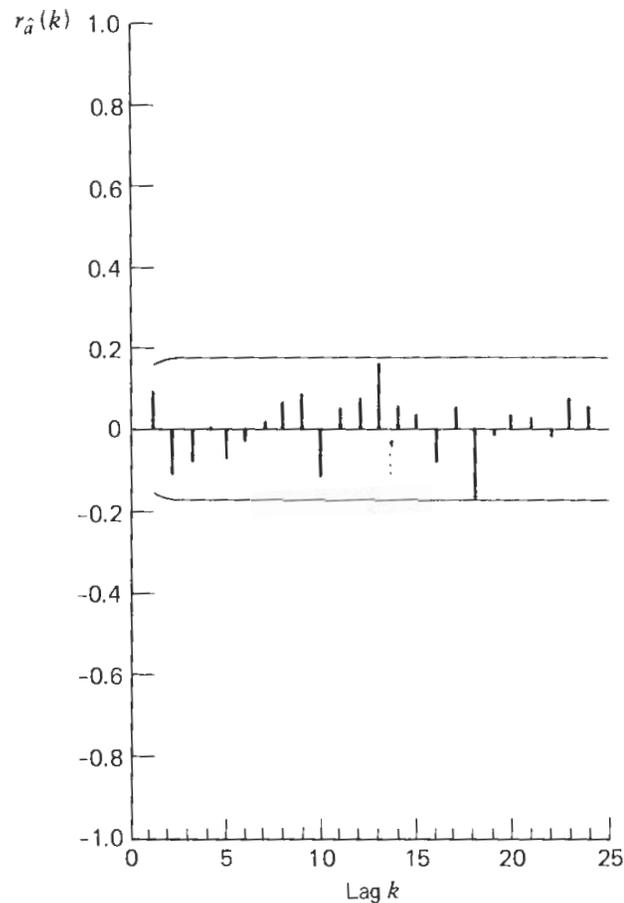

Figure 5.14. SACF of the residuals—yield series.

81

Forecasting

- $\hat{z}_{156}(1) = \hat{\mu} + \hat{\phi}(z_{156} - \hat{\mu}) = .97 + .85(.49 - .97) = .56$

Prediction limits

$$\hat{z}_{156}(1) \pm 1.96\hat{\sigma} = 0.56 \pm 1.96(.024)^{1/2} = 0.56 \pm 0.29$$

- $\hat{z}_{156}(2) = \hat{\mu} + \hat{\phi}^2(z_{156} - \hat{\mu}) = .97 + .85^2(.49 - .97) = .62$

Prediction limits

$$\hat{z}_{156}(2) \pm 1.96\hat{\sigma}(1 + \hat{\psi}_1^2)^{1/2} = 0.62 \pm 1.96(.024)^{1/2}(1 + .85^2)^{1/2} = 0.62 \pm 0.39$$

- $\hat{z}_{156}(3) = \hat{\mu} + \hat{\phi}^3(z_{156} - \hat{\mu}) = .97 + .85^3(.49 - .97) = .68$

Prediction limits

$$
\begin{aligned}
\hat{z}_{156}(3) \pm 1.96\hat{\sigma}(1 + \hat{\psi}_1^2 + \hat{\psi}_2^2)^{1/2} &= 0.68 \pm 1.96(.024)^{1/2}(1 + .85^2 + .85^4)^{1/2} \\
&= 0.68 \pm 0.45
\end{aligned}
$$

Suppose $z_{157} = 0.74$ has now become available. Forecasts for z_{158} and z_{159} are updated according to

- $z_{158} : \hat{z}_{157}(1) = \hat{z}_{156}(2) + \hat{\psi}_1[z_{157} - \hat{z}_{156}(1)] = 0.62 + .85(.74 - .56) = .77$

- $z_{159} : \hat{z}_{157}(2) = \hat{z}_{156}(3) + \hat{\psi}_2[z_{157} - \hat{z}_{156}(1)] = 0.62 + .85^2(.74 - .56) = .81$

- The prediction interval for z_{158} interval for z_{158} is 0.77 ± 0.29, and that for z_{159} is 0.81 ± 0.39.

Growth Rates

- Quarterly growth of Iowa nonfarm income from the second quarter of 1948 to the fourth quarter of 1978 ($N = 123$) are shown in Figure 5.8a

- The plot of the series indicates that no variance-stabilizing transformation is necessary.

- z_t is nonstationary and the first difference ∇z_t should be analyzed. The second difference is also stationary but would lead to overdifferencing. This is also indicated by the increase in the sample variance $[s^2(\nabla z_t) = 1.58, s^2(\nabla^2 z_t) = 4.82]$.

- The SACF of ∇z_t in Figure 5.11b indicates a cutoff after lag 1. The SPACF in Figure 5.15 shows a rough exponential decay. These patterns lead us to consider the model

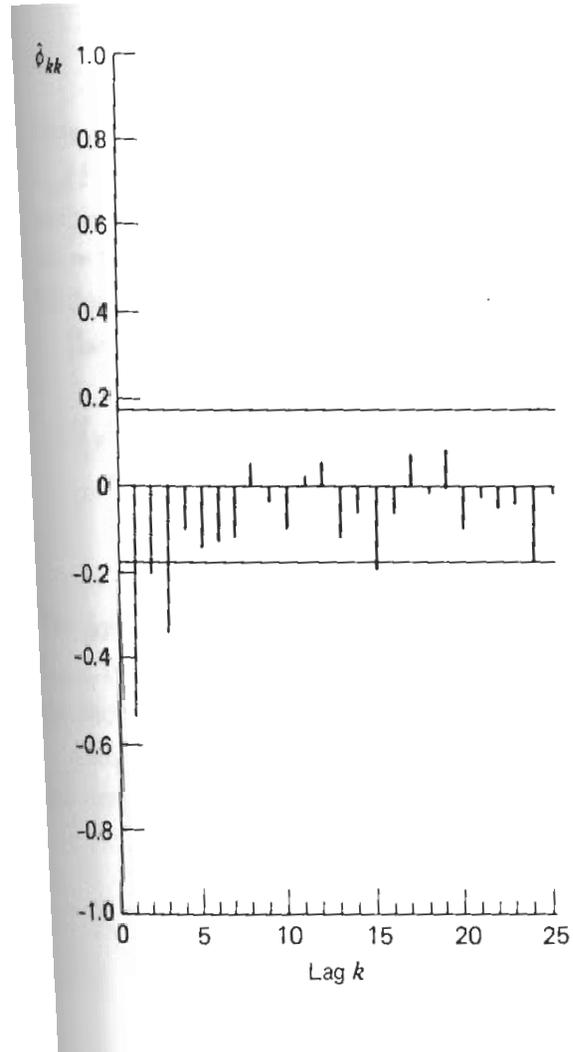$$\nabla z_t = \theta_0 + (1 - \theta B)a_t$$

Figure 5.15. SPACF of ∇z_t—growth rates.

- The ML estimates and their standard errors are

$$\hat{\theta}_0 = 0.012(.008), \quad \hat{\theta} = 0.92(.04), \quad \hat{\sigma}^2 = 0.93$$

- Dropping the insignificant θ_0 from the model and reestimating the parameters leads to

$$\hat{\theta} = 0.88(.04) \quad \hat{\sigma}^2 = 0.95$$

- The estimated residual autocorrelations together with their probability limits are present in Figure 5.16. Q statistic based on the first $K = 24$ autocorrelations $(n = 123 - 1 = 122)$

$$Q = 122 \times 124 \left(\frac{(-.01)^2}{121} + \frac{(.08)^2}{120} + \cdots + \frac{(-.09)^2}{98} \right) = 20.5 < \chi^2_{0.05}(23) = 35.2$$

Hence we take

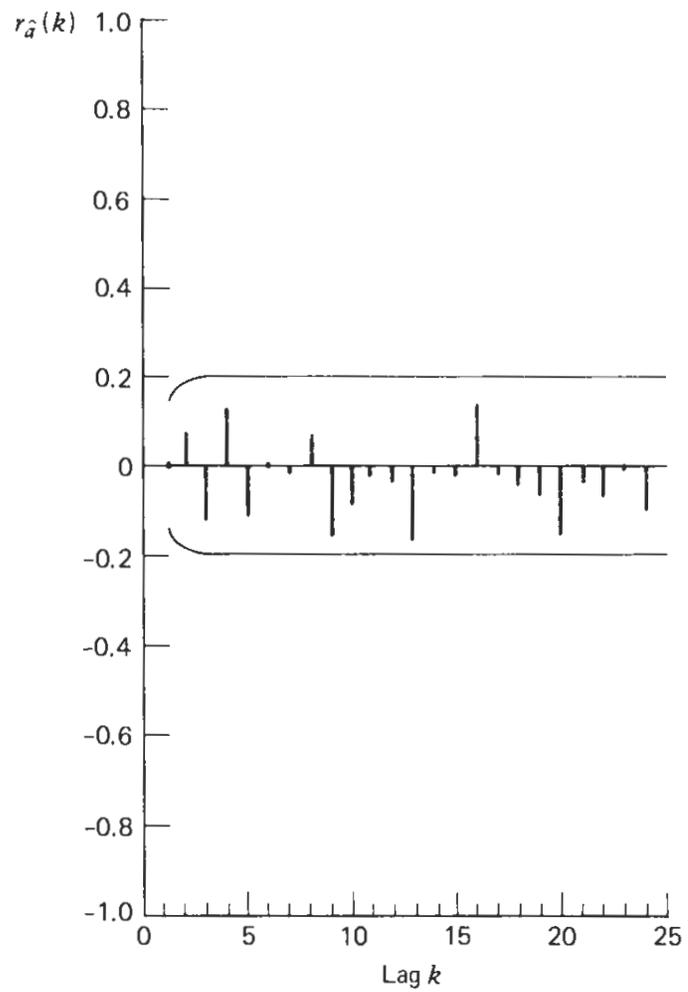$$z_t = z_{t-1} + a_t - 0.88a_{t-1}$$

as the forecasting model.

Figure 5.16. SACF for the residuals—growth rates.

Foreasting

- The predictions $\hat{z}_{123}(\ell)$ are described below

$$\hat{z}_{123}(\ell) = \hat{z}_{123}(1) = z_{123} - \theta \hat{a}_{123} = 3.38 - .88(.718) = 2.75$$

Note that $\hat{a}_{123} = .718$ is the last residual.

- The estimated variance of the ℓ-step-ahead forecast error is

$$\hat{V}[e_{123}(\ell)] = [1 + (\ell - 1)(1 - \hat{\theta})^2]\hat{\sigma}^2 = [1 + (\ell - 1)(1 - 0.88)^2](.95)$$

- after a new observation $z_{124} = 1.55$ has become available, update the forecast

$$\hat{z}_{124}(\ell) = \hat{z}_{123}(\ell+1) + \hat{\psi}_\ell[z_{124} - \hat{z}_{123}] = 2.75 + (1 - 0.88)(1.55 - 2.75) = 2.61$$

Furthermore, after $z_{125} = 2.93$ has become available,

$$\hat{z}_{125}(\ell) = \hat{z}_{124}(\ell+1) + \hat{\psi}_\ell[z_{125} - \hat{z}_{124}] = 2.61 + (1 - 0.88)(2.93 - 2.61) = 2.65$$

Table 5.5. Forecasts and Prediction Limits—Growth Rates

l	$\hat{z}_{123}(l)$	$\hat{V}[e_{123}(l)]$	95% Prediction Limits
1	2.75	0.95	2.75 ± 1.91
2	2.75	0.96	2.75 ± 1.92
3	2.75	0.97	2.75 ± 1.93
4	2.75	0.98	2.75 ± 1.94

Demand for Pepair Parts This series consists of the monthly demand for repair parts for a large Midwestern production company from January 1972 to June 1979 ($N = 90$). The series is shown in Figure 5.9a.

- A logarithmic transformation need to stabilize the variance and that first differences $\nabla ln Z_t$ are stationary. Further differencing is not necessary since it results in an increase in the sample variance $[s^2(\nabla z_t) = .034, s^2(\nabla^2 ln z_t) = .099]$

- The sample autocorrelation and partial autocorrelation are shown in Figure 5.12b and 5.17. They indicate a cutoff in the SACF after lag 1 and a rough exponential decay in the SPACF. Thus we consider the ARIMA(0,1,1) model

$$(1 - B)y_t = (1 - \theta B)a_t$$

where $y_t = \ln z_t$. The ML estiamtes are

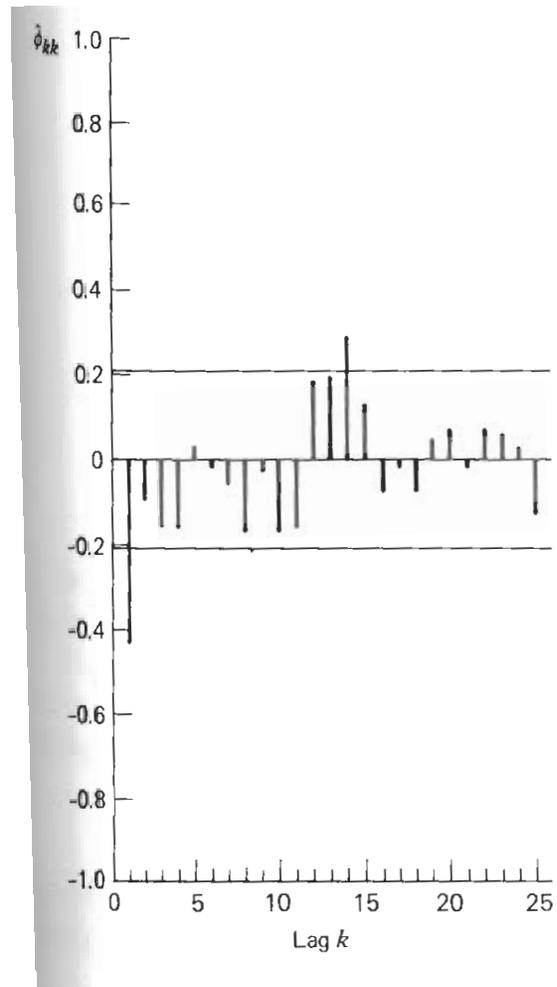$$\hat{\theta} = 0.57(.09), \hat{\sigma}^2 = 0.27$$

Figure 5.17. SPACF of $\nabla \ln z_t$ — demand for repair parts.

- A constant θ_0 was originally included in the model, but it was found to be insignificant. The estimated residual autocorrelations , together with their probability limits are shown in Figure 5.18. All autocorrelation except $r_{\hat{a}}(12)$ are small. The large autocorrelation $r_{\hat{a}}(12)$ could be due to "seasonal variation" in the data

- The Q statistic based on the first 24 autocorrelation is given by $Q = 37$ and is slightly bigger than $\chi^2_{0.05}(23) = 35.2$. This is due mainly to one large autocorrelation at lag 12.

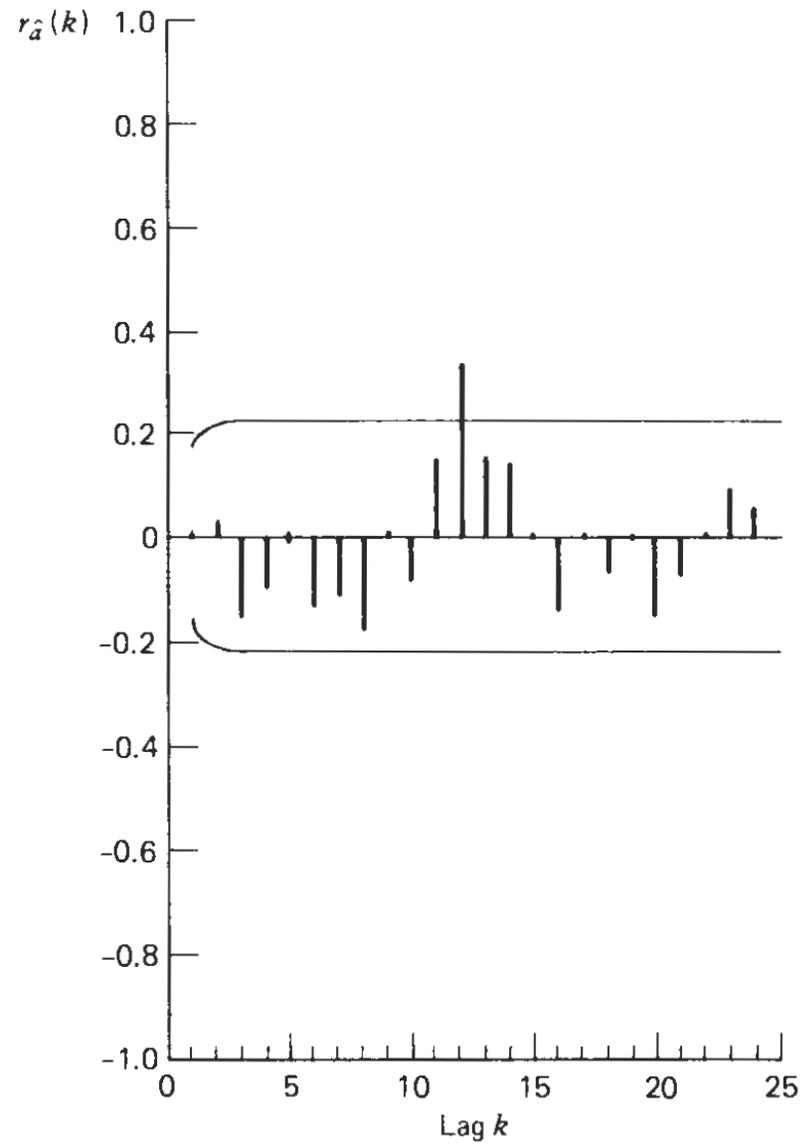- It may need to Consider a modification that takes account of the seasonal component.

Figure 5.18. SACF of the residuals—demand for repair parts.

Forecasting

- The forecasting model is given in terms of the transformed $y_t = \ln z_t$. The MMSE forecasts $\hat{y}(\ell)$ and the associated prediction interval c_1, c_2 for $y_{n+\ell}$ can be derive as usual.

- However, to derive the forecasts of original observation $z_{n+\ell} = \exp(y_{n+\ell})$, we have to back transform the forecasts and the prediction interval. They are given by

$$\hat{z}_n(\ell) = \exp[\hat{y}_n(\ell)] \quad \text{and} \quad [\exp(c_1), \exp(c_2)]$$

- The forecasts of the transformed series $y_{n+\ell}$ are MMSE forecasts. However if the transformation is nonlinear, as in the case the logarithmic transformation, the MMSE will not preserved for $\hat{z}_n(\ell) = \exp[\hat{y}_n(\ell)]$.

- Nevertheless, the forecast $\hat{z}_n(\ell)$ will be the median of the predictive distribution of $z_{n+\ell}$.

- For the present example, the prediction $\hat{y}_{90}(\ell)$ and the corresponding 95 percent prediction limits are given by

$$\hat{y}_{90}(\ell) = \hat{y}_{90}(1) = y_{90} - \hat{\theta}\hat{a}_{90} = 7.864 - .57(.234) = 7.731$$

and

$$\hat{y}_{90}(\ell) \pm 1.96[1+(\ell-1)(1-\hat{\theta})^2]^{1/2}\hat{\sigma} = 7.731 \pm 1.96[1+(\ell-1)(1-.57)^2]^{1/2}(.027)^{1/2}.$$

- Forecasts and 95 percent prediction interval s for y_{91}, y_{92} and z_{91}, z_{92} are given in Table 5.6.

- The prediction intervals are quite large. This is due partially to the large uncertainty $(\hat{\sigma}^2)$ in the past data and partly to the unexplained correlation at lag 12.

Table 5.6. Forecasts and Prediction Limits—Demand for Repair Parts[a]

l	95% Prediction Limits for $y_{90+l} = \ln z_{90+l}$	Approximate 95% Prediction Limits for z_{90+l}
1	7.731 ± 0.322 $= (7.409, 8.053)$	$[\exp(7.409), \exp(8.053)]$ $= (1651, 3143)$
2	7.731 ± 0.351 $= (7.380, 8.082)$	$[\exp(7.380), \exp(8.082)]$ $= (1604, 3236)$

[a] $\hat{y}_{90}(l) = 7.731$; $\hat{z}_{90}(l) = \exp(7.731) = 2278$.