CHAPTER 5

High-Frequency Data Analysis and Market Microstructure

High-frequency data are observations taken at fine time intervals. In finance, they often mean observations taken daily or at a finer time scale. These data have become available primarily due to advances in data acquisition and processing techniques, and they have attracted much attention because they are important in empirical study of market microstructure. The ultimate high-frequency data in finance are the transaction-by-transaction or trade-by-trade data in security markets. Here time is often measured in seconds. The Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE) contains all equity transactions reported on the *Consolidated Tape* from 1992 to the present, which includes transactions on the NYSE, AMEX, NASDAQ, and the regional exchanges. The Berkeley Options Data Base provides similar data for options transactions from August 1976 to December 1996. Transactions data for many other securities and markets, both domestic and foreign, are continuously collected and processed. Wood (2000) provides some historical perspective of high-frequency financial study.

High-frequency financial data are important in studying a variety of issues related to the trading process and market microstructure. They can be used to compare the efficiency of different trading systems in price discovery (e.g., the open out-cry system of the NYSE and the computer trading system of NASDAQ). They can also be used to study the dynamics of bid and ask quotes of a particular stock (e.g., Hasbrouck, 1999; Zhang, Russell, and Tsay, 2001b). In an order-driven stock market (e.g., the Taiwan Stock Exchange), high-frequency data can be used to study the order dynamics and, more interesting, to investigate the question of "who provides the market liquidity." Cho, Russell, Tiao, and Tsay (2003) use intraday 5-minute returns of more than 340 stocks traded on the Taiwan Stock Exchange to study the impact of daily stock price limits and find significant evidence of magnet effects toward the price ceiling.

However, high-frequency data have some unique characteristics that do not appear in lower frequencies. Analysis of these data thus introduces new challenges

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

to financial economists and statisticians. In this chapter, we study these special characteristics, consider methods for analyzing high-frequency data, and discuss implications of the results obtained. In particular, we discuss nonsynchronous trading, bid–ask spread, duration models, price movements that are in multiples of tick size, and bivariate models for price changes and time durations between transactions associated with price changes. The models discussed are also applicable to other scientific areas such as telecommunications and environmental studies.

5.1 NONSYNCHRONOUS TRADING

We begin with nonsynchronous trading. Stock tradings such as those on the NYSE do not occur in a synchronous manner; different stocks have different trading frequencies, and even for a single stock the trading intensity varies from hour to hour and from day to day. Yet we often analyze a return series in a fixed time interval such as daily, weekly, or monthly. For daily series, price of a stock is its *closing* price, which is the last transaction price of the stock in a trading day. The actual time of the last transaction of the stock varies from day to day. As such we incorrectly assume daily returns as an equally spaced time series with a 24-hour interval. It turns out that such an assumption can lead to erroneous conclusions about the predictability of stock returns even if the true return series are serially independent.

For daily stock returns, nonsynchronous trading can introduce (a) lag-1 cross-correlation between stock returns, (b) lag-1 serial correlation in a portfolio return, and (c) in some situations negative serial correlations of the return series of a single stock. Consider stocks A and B. Assume that the two stocks are independent and stock A is traded more frequently than stock B. For special news affecting the market that arrives near the closing hour on one day, stock A is more likely than B to show the effect of the news on the same day simply because A is traded more frequently. The effect of the news on B will eventually appear, but it may be delayed until the following trading day. If this situation indeed happens, return of stock A appears to lead that of stock B. Consequently, the return series may show a significant lag-1 cross-correlation from A to B even though the two stocks are independent. For a portfolio that holds stocks A and B, the prior cross-correlation would become a significant lag-1 serial correlation.

In a more complicated manner, nonsynchronous trading can also induce erroneous negative serial correlations for a single stock. There are several models available in the literature to study this phenomenon; see Campbell, Lo, and MacKinlay (1997) and the references therein. Here we adopt a simplified version of the model proposed in Lo and MacKinlay (1990). Let r_t be the continuously compounded return of a security at the time index t. For simplicity, assume that $\{r_t\}$ is a sequence of independent and identically distributed random variables with mean $E(r_t) = \mu$ and variance $\text{Var}(r_t) = \sigma^2$. For each time period, the probability that the security is not traded is π, which is time-invariant and independent

of r_t. Let r_t^o be the observed return. When there is no trade at time index t, we have $r_t^o = 0$ because there is no information available. Yet when there is a trade at time index t, we define r_t^o as the cumulative return from the previous trade (i.e., $r_t^o = r_t + r_{t-1} + \cdots + r_{t-k_t}$, where k_t is the largest non-negative integer such that no trade occurred in the periods $t - k_t, t - k_t + 1, \ldots, t - 1$). Mathematically, the relationship between r_t and r_t^o is

$$r_t^o = \begin{cases} 0 & \text{with probability } \pi \\ r_t & \text{with probability } (1-\pi)^2 \\ r_t + r_{t-1} & \text{with probability } (1-\pi)^2\pi \\ r_t + r_{t-1} + r_{t-2} & \text{with probability } (1-\pi)^2\pi^2 \\ \vdots & \vdots \\ \sum_{i=0}^{k} r_{t-i} & \text{with probability } (1-\pi)^2\pi^k \\ \vdots & \vdots \end{cases} \tag{5.1}$$

These probabilities are easy to understand. For example, $r_t^o = r_t$ if and only if there are trades at both t and $t-1$, $r_t^o = r_t + r_{t-1}$ if and only if there are trades at t and $t-2$, but no trade at $t-1$, and $r_t^o = r_t + r_{t-1} + r_{t-2}$ if and only if there are trades at t and $t-3$, but no trades at $t-1$ and $t-2$, and so on. As expected, the total probability is 1 given by

$$\pi + (1-\pi)^2[1 + \pi + \pi^2 + \cdots] = \pi + (1-\pi)^2\frac{1}{1-\pi} = \pi + 1 - \pi = 1.$$

We are ready to consider the moment equations of the observed return series $\{r_t^o\}$. First, the expectation of r_t^o is

$$\begin{aligned} E(r_t^o) &= (1-\pi)^2 E(r_t) + (1-\pi)^2\pi E(r_t + r_{t-1}) + \cdots \\ &= (1-\pi)^2\mu + (1-\pi)^2\pi 2\mu + (1-\pi)^2\pi^2 3\mu + \cdots \\ &= (1-\pi)^2\mu[1 + 2\pi + 3\pi^2 + 4\pi^3 + \cdots] \\ &= (1-\pi)^2\mu\frac{1}{(1-\pi)^2} = \mu. \end{aligned} \tag{5.2}$$

In the prior derivation, we use the result $1 + 2\pi + 3\pi^2 + 4\pi^3 + \cdots = 1/(1-\pi)^2$. Next, for the variance of r_t^o, we use $\text{Var}(r_t^o) = E[(r_t^o)^2] - [E(r_t^o)]^2$ and

$$\begin{aligned} E(r_t^o)^2 &= (1-\pi)^2 E[(r_t)^2] + (1-\pi)^2\pi E[(r_t + r_{t-1})^2] + \cdots \\ &= (1-\pi)^2[(\sigma^2 + \mu^2) + \pi(2\sigma^2 + 4\mu^2) + \pi^2(3\sigma^2 + 9\mu^2) + \cdots] \quad (5.3) \\ &= (1-\pi)^2\{\sigma^2[1 + 2\pi + 3\pi^2 + \cdots] + \mu^2[1 + 4\pi + 9\pi^2 + \cdots]\} \quad (5.4) \\ &= \sigma^2 + \mu^2\left[\frac{2}{1-\pi} - 1\right]. \quad (5.5) \end{aligned}$$

In Eq. (5.3), we use

$$E\left(\sum_{i=0}^{k} r_{t-i}\right)^2 = \text{Var}\left(\sum_{i=0}^{k} r_{t-i}\right) + \left[E\left(\sum_{i=0}^{k} r_{t-i}\right)\right]^2 = (k+1)\sigma^2 + [(k+1)\mu]^2$$

under the serial independence assumption of r_t. Using techniques similar to that of Eq. (5.2), we can show that the first term of Eq. (5.4) reduces to σ^2. For the second term of Eq. (5.4), we use the identity

$$1 + 4\pi + 9\pi^2 + 16\pi^3 + \cdots = \frac{2}{(1-\pi)^3} - \frac{1}{(1-\pi)^2},$$

which can be obtained as follows. Let

$$H = 1 + 4\pi + 9\pi^2 + 16\pi^3 + \cdots \quad \text{and} \quad G = 1 + 3\pi + 5\pi^2 + 7\pi^3 + \cdots.$$

Then $(1-\pi)H = G$ and

$$\begin{aligned} (1-\pi)G &= 1 + 2\pi + 2\pi^2 + 2\pi^3 + \cdots \\ &= 2(1 + \pi + \pi^2 + \cdots) - 1 = \frac{2}{(1-\pi)} - 1. \end{aligned}$$

Consequently, from Eqs. (5.2) and (5.5), we have

$$\text{Var}(r_t^o) = \sigma^2 + \mu^2\left[\frac{2}{1-\pi} - 1\right] - \mu^2 = \sigma^2 + \frac{2\pi\mu^2}{1-\pi}. \tag{5.6}$$

Consider next the lag-1 autocovariance of $\{r_t^o\}$. Here we use $\text{Cov}(r_t^o, r_{t-1}^o) = E(r_t^o r_{t-1}^o) - E(r_t^0)E(r_{t-1}^o) = E(r_t^o r_{t-1}^o) - \mu^2$. The question then reduces to finding $E(r_t^o r_{t-1}^o)$. Notice that $r_t^o r_{t-1}^o$ is zero if there is no trade at t, no trade at $t-1$, or no trade at both t and $t-1$. Therefore, we have

$$r_t^o r_{t-1}^o = \begin{cases} 0 & \text{with probability } 2\pi - \pi^2 \\ r_t r_{t-1} & \text{with probability } (1-\pi)^3 \\ r_t(r_{t-1} + r_{t-2}) & \text{with probability } (1-\pi)^3\pi \\ r_t(r_{t-1} + r_{t-2} + r_{t-3}) & \text{with probability } (1-\pi)^3\pi^2 \\ \vdots & \vdots \\ r_t(\sum_{i=1}^{k} r_{t-i}) & \text{with probability } (1-\pi)^3\pi^{k-1} \\ \vdots & \vdots \end{cases} \tag{5.7}$$

Again the total probability is unity. To understand the prior result, notice that $r_t^o r_{t-1}^o = r_t r_{t-1}$ if and only if there are three consecutive trades at $t-2$, $t-1$,

and t. Using Eq. (5.7) and the fact that $E(r_t r_{t-j}) = E(r_t)E(r_{t-j}) = \mu^2$ for $j > 0$, we have

$$E(r_t^o r_{t-1}^o) = (1-\pi)^3 \left\{ E(r_t r_{t-1}) + \pi E[r_t(r_{t-1} + r_{t-2})] + \pi^2 E\left[r_t \left(\sum_{i=1}^{3} r_{t-i} \right) \right] + \cdots \right\}$$
$$= (1-\pi)^3 \mu^2 [1 + 2\pi + 3\pi^2 + \cdots] = (1-\pi)\mu^2.$$

The lag-1 autocovariance of $\{r_t^o\}$ is then

$$\text{Cov}(r_t^o, r_{t-1}^o) = -\pi\mu^2. \tag{5.8}$$

Provided that μ is not zero, the nonsynchronous trading induces a *negative* lag-1 autocorrelation in r_t^o given by

$$\rho_1(r_t^o) = \frac{-(1-\pi)\pi\mu^2}{(1-\pi)\sigma^2 + 2\pi\mu^2}.$$

In general, we can extend the prior result and show that

$$\text{Cov}(r_t^o, r_{t-j}^o) = -\mu^2\pi^j, \quad j \geq 1.$$

The magnitude of the lag-1 ACF depends on the choices of μ, π, and σ and can be substantial. Thus, when $\mu \neq 0$, the nonsynchronous trading induces negative autocorrelations in an observed security return series.

The previous discussion can be generalized to the return series of a portfolio that consists of N securities; see Campbell, Lo, and MacKinlay (1997, Chapter 3). In the time series literature, effects of nonsynchronous trading on the return of a single security are equivalent to that of random temporal aggregation on a time series, with the trading probability π governing the mechanism of aggregation.

5.2 BID–ASK SPREAD

In some stock exchanges (e.g., NYSE), market makers play an important role in facilitating trades. They provide market liquidity by standing ready to buy or sell whenever the public wishes to buy or sell. By market liquidity, we mean the ability to buy or sell significant quantities of a security quickly, anonymously, and with little price impact. In return for providing liquidity, market makers are granted monopoly rights by the exchange to post different prices for purchases and sales of a security. They buy at the *bid* price P_b and sell at a higher ask price P_a. (For the public, P_b is the sale price and P_a is the purchase price.) The difference $P_a - P_b$

is call the *bid–ask spread*, which is the primary source of compensation for market makers. Typically, the bid–ask spread is small—namely, one or two ticks.

The existence of a bid–ask spread, although small in magnitude, has several important consequences in time series properties of asset returns. We briefly discuss the bid–ask bounce—namely, the bid–ask spread introduces *negative* lag-1 serial correlation in an asset return. Consider the simple model of Roll (1984). The observed market price P_t of an asset is assumed to satisfy

$$P_t = P_t^* + I_t \frac{S}{2}, \tag{5.9}$$

where $S = P_a - P_b$ is the bid–ask spread, P_t^* is the time-t fundamental value of the asset in a frictionless market, and $\{I_t\}$ is a sequence of independent binary random variables with equal probabilities (i.e., $I_t = 1$ with probability 0.5 and $= -1$ with probability 0.5). The I_t can be interpreted as an order-type indicator, with 1 signifying buyer-initiated transaction and -1 seller-initiated transaction. Alternatively, the model can be written as

$$P_t = P_t^* + \begin{cases} +S/2 & \text{with probability 0.5,} \\ -S/2 & \text{with probability 0.5.} \end{cases}$$

If there is no change in P_t^*, then the observed process of price changes is

$$\Delta P_t = (I_t - I_{t-1})\frac{S}{2}. \tag{5.10}$$

Under the assumption of I_t in Eq. (5.9), $E(I_t) = 0$ and $\text{Var}(I_t) = 1$, and we have $E(\Delta P_t) = 0$ and

$$\text{Var}(\Delta P_t) = S^2/2, \tag{5.11}$$
$$\text{Cov}(\Delta P_t, \Delta P_{t-1}) = -S^2/4, \tag{5.12}$$
$$\text{Cov}(\Delta P_t, \Delta P_{t-j}) = 0, \quad j > 1. \tag{5.13}$$

Therefore, the autocorrelation function of ΔP_t is

$$\rho_j(\Delta P_t) = \begin{cases} -0.5 & \text{if } j = 1, \\ 0 & \text{if } j > 1. \end{cases} \tag{5.14}$$

The bid–ask spread thus introduces a negative lag-1 serial correlation in the series of observed price changes. This is referred to as the *bid–ask bounce* in the finance literature. Intuitively, the bounce can be seen as follows. Assume that the fundamental price P_t^* is equal to $(P_a + P_b)/2$. Then P_t assumes the value P_a or P_b. If the previously observed price is P_a (the higher value), then the current observed price is either unchanged or lower at P_b. Thus, ΔP_t is either 0 or $-S$. However, if the previous observed price is P_b (the lower value), then ΔP_t is either 0 or S. The negative lag-1 correlation in ΔP_t becomes apparent. The bid–ask spread does not introduce any serial correlation beyond lag 1, however.

A more realistic formulation is to assume that P_t^* follows a random walk so that $\Delta P_t^* = P_t^* - P_{t-1}^* = \epsilon_t$, which forms a sequence of independent and identically distributed random variables with mean zero and variance σ^2. In addition, $\{\epsilon_t\}$ is independent of $\{I_t\}$. In this case, $\text{Var}(\Delta P_t) = \sigma^2 + S^2/2$, but $\text{Cov}(\Delta P_t, \Delta P_{t-j})$ remains unchanged. Therefore,

$$\rho_1(\Delta P_t) = \frac{-S^2/4}{S^2/2 + \sigma^2} \leq 0.$$

The magnitude of the lag-1 autocorrelation of ΔP_t is reduced, but the negative effect remains when $S = P_a - P_b > 0$. In finance, it might be of interest to study the components of the bid–ask spread. Interested readers are referred to Campbell, Lo, and MacKinlay (1997) and the references therein.

The effect of bid–ask spread continues to exist in portfolio returns and in multivariate financial time series. Consider the bivariate case. Denote the bivariate order-type indicator by $I_t = (I_{1t}, I_{2t})'$, where I_{1t} is for the first security and I_{2t} for the second security. If I_{1t} and I_{2t} are contemporaneously positively correlated, then the bid–ask spreads can introduce negative lag-1 cross-correlations.

5.3 EMPIRICAL CHARACTERISTICS OF TRANSACTIONS DATA

Let t_i be the calendar time, measured in seconds from midnight, at which the ith transaction of an asset takes place. Associated with the transaction are several variables such as the transaction price, the transaction volume, the prevailing bid and ask quotes, and so on. The collection of t_i and the associated measurements are referred to as the *transactions data*. These data have several important characteristics that do not exist when the observations are aggregated over time. Some of the characteristics are given next.

1. *Unequally Spaced Time Intervals*. Transactions such as stock tradings on an exchange do not occur at equally spaced time intervals. As such, the observed transaction prices of an asset do not form an equally spaced time series. The time duration between trades becomes important and might contain useful information about market microstructure (e.g., trading intensity).
2. *Discrete-Valued Prices*. The price change of an asset from one transaction to the next only occurs in multiples of tick size. On the NYSE, the tick size was one-eighth of a dollar before June 24, 1997 and was one-sixteenth of a dollar before January 29, 2001. All NYSE and AMEX stocks started to trade in decimals on January 29, 2001. Therefore, the price is a discrete-valued variable in transactions data. In some markets, price change may also be subject to limit constraints set by regulators.
3. *Existence of a Daily Periodic or Diurnal Pattern*. Under the normal trading conditions, transaction activity can exhibit a periodic pattern. For instance, on the NYSE, transactions are "heavier" at the beginning and closing of the trading hours and "thinner" during lunch hour, resulting in a U-shape transaction intensity. Consequently, time durations between transactions also exhibit a daily cyclical pattern.
4. *Multiple Transactions Within a Single Second*. It is possible that multiple transactions, even with different prices, occur at the same time. This is partly due to the fact that time is measured in seconds that may be too long a time scale in periods of heavy trading.

To demonstrate these characteristics, we consider first the IBM transactions data from November 1, 1990 to January 31, 1991. These data are from the Trades, Orders Reports, and Quotes (TORQ) dataset; see Hasbrouck (1992). There are 63 trading days and 60,328 transactions. To simplify the discussion, we ignore the price changes between trading days and focus on the transactions that occurred in the normal trading hours from 9:30 am to 4:00 pm Eastern time. It is well known that overnight stock returns differ substantially from intraday returns; see Stoll and Whaley (1990) and the references therein. Table 5.1 gives the frequencies in percentages of price change measured in the tick size of \$1/8 = \$0.125. From the table, we make the following observations:

1. About two-thirds of the intraday transactions were without price change.
2. The price changed in one tick approximately 29% of the intraday transactions.
3. Only 2.6% of the transactions were associated with two-tick price changes.
4. Only about 1.3% of the transactions resulted in price changes of three ticks or more.
5. The distribution of positive and negative price changes was approximately symmetric.

Consider next the number of transactions in a 5-minute time interval. Denote the series by x_t. That is, x_1 is the number of IBM transactions from 9:30 am to 9:35 am on November 1, 1990 Eastern time, x_2 is the number of transactions from 9:35 am to 9:40 am, and so on. The time gaps between trading days are ignored. Figure 5.1a shows the time plot of x_t, and Figure 5.1b the sample ACF of x_t for lags 1 to 260. Of particular interest is the cyclical pattern of the ACF with a periodicity of 78, which is the number of 5-minute intervals in a trading day. The number of transactions thus exhibits a daily pattern. To further illustrate the daily trading pattern, Figure 5.2 shows the average number of transactions within 5-minute time intervals over the 63 days. There are 78 such averages. The plot exhibits a "smiling" or U shape, indicating heavier trading at the opening and closing of the market and thinner trading during the lunch hours.

Table 5.1. Frequencies of Price Change in Multiples of Tick Size for IBM Stock from November 1, 1990 to January 31, 1991

Number (tick)	≤ -3	-2	-1	0	1	2	≥ 3
Percentage	0.66	1.33	14.53	67.06	14.53	1.27	0.63

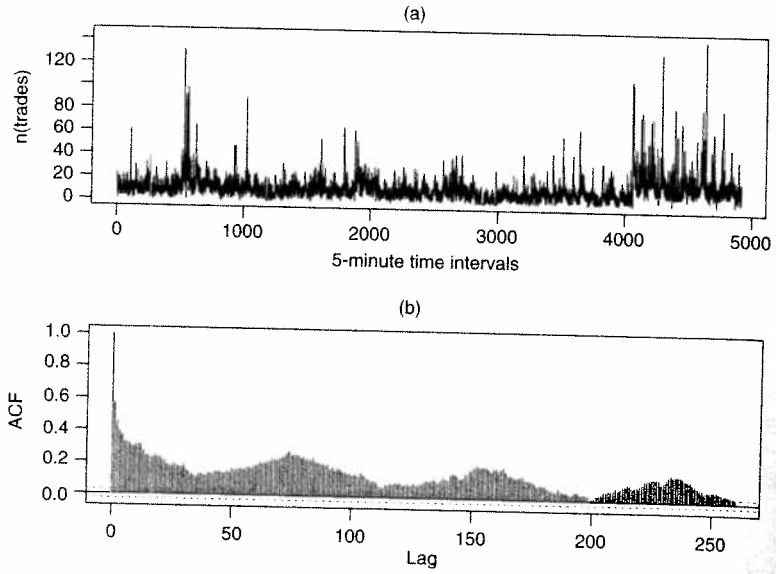

Figure 5.1. IBM intraday transactions data from 11/01/90 to 1/31/91: (a) the number of transactions in 5-minute time intervals and (b) the sample ACF of the series in part(a).

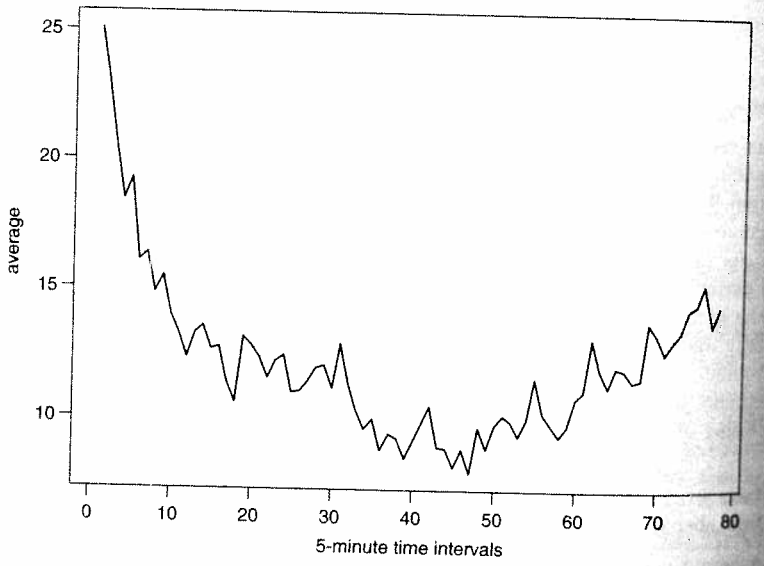

Figure 5.2. Time plot of the average number of transactions in 5-minute time intervals. There are 78 observations, averaging over the 63 trading days from 11/01/90 to 1/31/91 for IBM stock.

Since we focus on transactions that occurred during normal trading hours of a trading day, there are 59,838 time intervals in the data. These intervals are called the intraday *durations* between trades. For IBM stock, there were 6531 zero time intervals. That is, during the normal trading hours of the 63 trading days from November 1, 1990 to January 31, 1991, multiple transactions in a second occurred 6531 times, which is about 10.91%. Among these multiple transactions, 1002 of them had different prices, which is about 1.67% of the total number of intraday transactions. Therefore, multiple transactions (i.e., zero durations) may become an issue in statistical modeling of the time durations between trades.

Table 5.2 provides a two-way classification of price movements. Here price movements are classified into "up," "unchanged," and "down." We denote them by "+," "0," and "−," respectively. The table shows the price movements between two consecutive trades (i.e., from the $(i-1)$th to the ith transaction) in the sample. From the table, trade-by-trade data show that:

1. Consecutive price increases or decreases are relatively rare, which are about 441/59837 = 0.74% and 410/59837 = 0.69%, respectively.
2. There is a slight edge to move from "up" to "unchanged" rather than to "down"; see row 1 of the table.
3. There is a high tendency for the price to remain "unchanged."
4. The probabilities of moving from "down" to "up" or "unchanged" are about the same; see row 3.

The first observation mentioned before is a clear demonstration of bid–ask bounce, showing *price reversals* in intraday transactions data. To confirm this phenomenon, we consider a directional series D_i for price movements, where D_i assumes the value $+1$, 0, and -1 for up, unchanged, and down price movement, respectively, for the ith transaction. The ACF of $\{D_i\}$ has a single spike at lag 1 with value -0.389, which is highly significant for a sample size of 59,837 and confirms the price reversal in consecutive trades.

As a second illustration, we consider the transactions data of IBM stock in December 1999 obtained from the TAQ database. The normal trading hours are

Table 5.2. Two-Way Classification of Price Movements in Consecutive Intraday Trades for IBM Stock[a]

	ith Trade			
$(i-1)$th Trade	+	0	−	Margin
+	441	5498	3948	9887
0	4867	29779	5473	40119
−	4580	4841	410	9831
Margin	9888	40118	9831	59837

[a]The price movements are classified into "up," "unchanged," and "down." The data span is from 11/01/90 to 1/31/91.

from 9:30 am to 4:00 pm Eastern time, except for December 31 when the market closed at 1:00 pm. Comparing with the 1990–1991 data, two important changes have occurred. First, the number of intraday tradings has increased sixfold. There were 134,120 intraday tradings in December 1999 alone. The increased trading intensity also increased the chance of multiple transactions within a second. The percentage of trades with zero time duration doubled to 22.98%. At the extreme, there were 42 transactions within a given second that happened twice on December 3, 1999. Second, the tick size of price movement was \$1/16 = \$0.0625 instead of \$1/8. The change in tick size should reduce the bid–ask spread. Figure 5.3 shows the daily number of transactions in the new sample. Figure 5.4a shows the time plot of time durations between trades, measured in seconds, and Figure 5.4b is the time plot of price changes in consecutive intraday trades, measured in multiples of the tick size of \$1/16. As expected, Figures 5.3 and 5.4a show clearly the inverse relationship between the daily number of transactions and the time interval between trades. Figure 5.4b shows two unusual price movements for IBM stock on December 3, 1999. They were a drop of 63 ticks followed by an immediate jump of 64 ticks and a drop of 68 ticks followed immediately by a jump of 68 ticks. Unusual price movements like these occurred infrequently in intraday transactions.

Focusing on trades recorded within regular trading hours, we have 61,149 trades out of 133,475 with no price change. This is about 45.8% and substantially lower than that between November 1990 and January 1991. It seems that reducing the tick size increased the chance of a price change. Table 5.3 gives the percentages of trades associated with a price change. The price movements remain approximately symmetric with respect to zero. Large price movements in intraday tradings are still relatively rare.

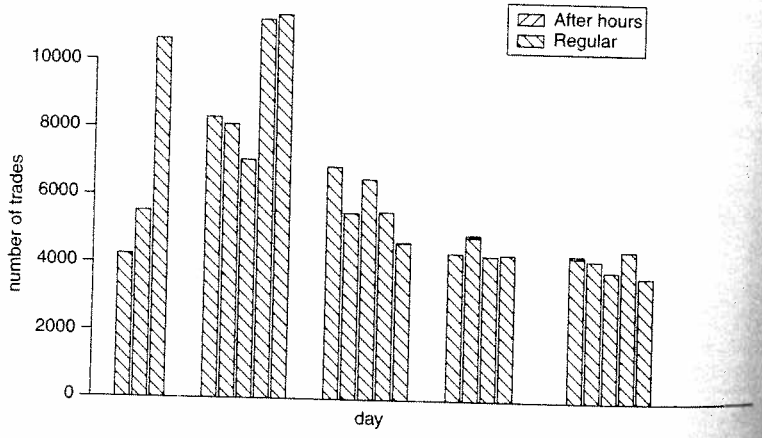

Figure 5.3. IBM transactions data for December 1999. The plot shows the number of transactions in each trading day with the after-hours portion denoting the number of trades with time stamp after 4:00 pm.

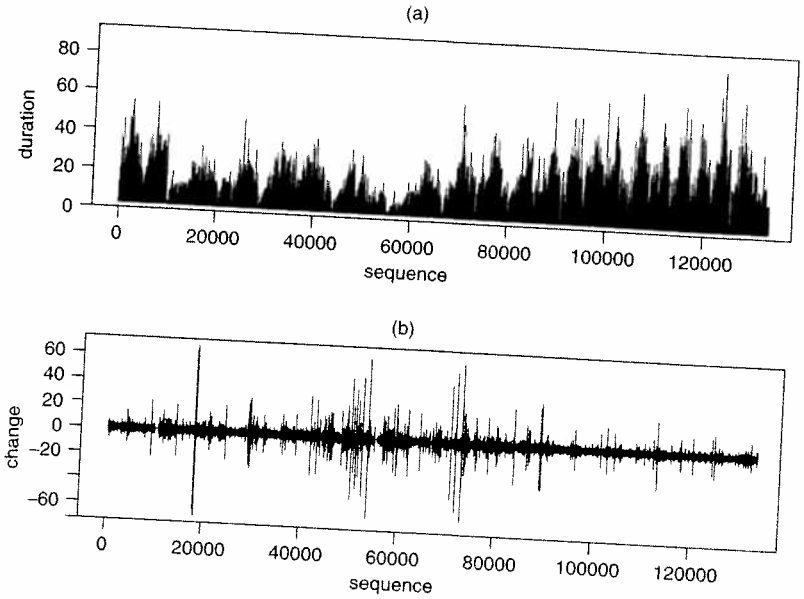

Figure 5.4. IBM transactions data for December 1999. (a) The time plot of time durations between trades. (b) The time plot of price changes in consecutive trades measured in multiples of the tick size of \$1/16. Only data during normal trading hours are included.

Table 5.3. Percentages of Intraday Transactions Associated with a Price Change for IBM Stock Traded in December 1999[a]

Size	1	2	3	4	5	6	7	>7
				Upward Movements				
Percentage	18.03	5.80	1.79	0.66	0.25	0.15	0.09	0.32
				Downward Movements				
Percentage	18.24	5.57	1.79	0.71	0.24	0.17	0.10	0.31

[a]The percentage of transactions without price change is 45.8% and the total number of transactions recorded within regular trading hours is 133,475. The size is measured in multiples of tick size \$1/16.

Remark. The recordkeeping of high-frequency data is often not as good as that of observations taken at lower frequencies. Data cleaning becomes a necessity in high-frequency data analysis. For transactions data, missing observations may happen in many ways, and the accuracy of the exact transaction time might be questionable for some trades. For example, recorded trading times may be beyond 4:00 pm Eastern time even before the opening of after-hours tradings. How to handle

these observations deserves a careful study. A proper method of data cleaning requires a deep understanding of the way in which the market operates. As such, it is important to specify clearly and precisely the methods used in data cleaning. These methods must be taken into consideration in making inference. □

Again, let t_i be the calendar time, measured in seconds from midnight, when the ith transaction took place. Let P_{t_i} be the transaction price. The price change from the $(i-1)$th to the ith trade is $y_i \equiv \Delta P_{t_i} = P_{t_i} - P_{t_{i-1}}$ and the time duration is $\Delta t_i = t_i - t_{i-1}$. Here it is understood that the subscript i in Δt_i and y_i denotes the time sequence of transactions, not the calendar time. In what follows, we consider models for y_i and Δt_i both individually and jointly.

5.4 MODELS FOR PRICE CHANGES

The discreteness and concentration on "no change" make it difficult to model the intraday price changes. Campbell, Lo, and MacKinlay (1997) discuss several econometric models that have been proposed in the literature. Here we mention two models that have the advantage of employing explanatory variables to study the intraday price movements. The first model is the ordered probit model used by Hauseman, Lo, and MacKinlay (1992) to study the price movements in transactions data. The second model has been considered recently by McCulloch and Tsay (2000) and is a simplified version of the model proposed by Rydberg and Shephard (2003); see also Ghysels (2000).

5.4.1 Ordered Probit Model

Let y_i^* be the unobservable price change of the asset under study (i.e., $y_i^* = P_{t_i}^* - P_{t_{i-1}}^*$), where P_t^* is the *virtual* price of the asset at time t. The ordered probit model assumes that y_i^* is a continuous random variable and follows the model

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i, \tag{5.15}$$

where $\boldsymbol{x}_i$ is a p-dimensional row vector of explanatory variables available at time t_{i-1}, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector, $E(\epsilon_i|\boldsymbol{x}_i) = 0$, $\text{Var}(\epsilon_i|\boldsymbol{x}_i) = \sigma_i^2$, and $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. The conditional variance σ_i^2 is assumed to be a positive function of the explanatory variable $\boldsymbol{w}_i$—that is,

$$\sigma_i^2 = g(\boldsymbol{w}_i), \tag{5.16}$$

where $g(.)$ is a positive function. For financial transactions data, $\boldsymbol{w}_i$ may contain the time interval $t_i - t_{i-1}$ and some conditional heteroscedastic variables. Typically, one also assumes that the conditional distribution of ϵ_i given $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ is Gaussian.

Suppose that the observed price change y_i may assume k possible values. In theory, k can be infinity, but countable. In practice, k is finite and may involve combining several categories into a single value. For example, we have $k = 7$ in Table 5.1, where the first value "−3 ticks" means that the price change is −3 ticks

or lower. We denote the k possible values as $\{s_1, \ldots, s_k\}$. The ordered probit model postulates the relationship between y_i and y_i^* as

$$y_i = s_j \quad \text{if} \quad \alpha_{j-1} < y_i^* \leq \alpha_j, \quad j = 1, \ldots, k, \tag{5.17}$$

where α_j are real numbers satisfying $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{k-1} < \alpha_k = \infty$. Under the assumption of conditional Gaussian distribution, we have

$$\begin{aligned} P(y_i = s_j|\boldsymbol{x}_i, \boldsymbol{w}_i) &= P(\alpha_{j-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i \leq \alpha_j|\boldsymbol{x}_i, \boldsymbol{w}_i) \\ &= \begin{cases} P(\boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i \leq \alpha_1|\boldsymbol{x}_i, \boldsymbol{w}_i) & \text{if } j = 1, \\ P(\alpha_{j-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i \leq \alpha_j|\boldsymbol{x}_i, \boldsymbol{w}_i) & \text{if } j = 2, \ldots, k-1, \\ P(\alpha_{k-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i|\boldsymbol{x}_i, \boldsymbol{w}_i) & \text{if } j = k, \end{cases} \\ &= \begin{cases} \Phi\left[\dfrac{\alpha_1 - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] & \text{if } j = 1, \\ \Phi\left[\dfrac{\alpha_j - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] - \Phi\left[\dfrac{\alpha_{j-1} - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] & \text{if } j = 2, \ldots, k-1, \\ 1 - \Phi\left[\dfrac{\alpha_{k-1} - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] & \text{if } j = k, \end{cases} \end{aligned} \tag{5.18}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal random variable evaluated at x, and we write $\sigma_i(\boldsymbol{w}_i)$ to denote that σ_i^2 is a positive function of $\boldsymbol{w}_i$. From the definition, an ordered probit model is driven by an unobservable continuous random variable. The observed values, which have a natural ordering, can be regarded as categories representing the underlying process.

The ordered probit model contains parameters $\boldsymbol{\beta}$, $\alpha_i (i = 1, \ldots, k-1)$, and those in the conditional variance function $\sigma_i(\boldsymbol{w}_i)$ in Eq. (5.16). These parameters can be estimated by the maximum likelihood or Markov chain Monte Carlo methods.

Example 5.1. Hauseman, Lo, and MacKinlay (1992) apply the ordered probit model to the 1988 transactions data of more than 100 stocks. Here we only report their result for IBM. There are 206,794 trades. The sample mean (standard deviation) of price change y_i, time duration Δt_i, and bid–ask spread are −0.0010(0.753), 27.21(34.13), and 1.9470(1.4625), respectively. The bid–ask spread is measured in ticks. The model used has nine categories for price movement, and the functional specifications are

$$\begin{aligned} \boldsymbol{x}_i\boldsymbol{\beta} = \beta_1 \Delta t_i^* &+ \sum_{v=1}^{3} \beta_{v+1} y_{i-v} + \sum_{v=1}^{3} \beta_{v+4} \text{SP5}_{i-v} + \sum_{v=1}^{3} \beta_{v+7} \text{IBS}_{i-v} \\ &+ \sum_{v=1}^{3} \beta_{v+10} [T_\lambda(V_{i-v}) \times \text{IBS}_{i-v}], \end{aligned} \tag{5.19}$$

$$\sigma_i^2(\boldsymbol{w}_i) = 1.0 + \gamma_1^2 \Delta t_i^* + \gamma_2^2 \text{AB}_{i-1}, \tag{5.20}$$

where $T_\lambda(V) = (V^\lambda - 1)/\lambda$ is the Box–Cox (1964) transformation of V with $\lambda \in [0, 1]$ and the explanatory variables are defined by the following:

- $\Delta t_i^* = (t_i - t_{i-1})/100$ is a rescaled time duration between the $(i-1)$th and ith trades with time measured in seconds.
- AB_{i-1} is the bid–ask spread prevailing at time t_{i-1} in ticks.
- y_{i-v} $(v = 1, 2, 3)$ is the lagged value of price change at t_{i-v} in ticks. With $k = 9$, the possible values of price changes are $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$ in ticks.
- V_{i-v} $(v = 1, 2, 3)$ is the lagged value of dollar volume at the $(i-v)$th transaction, defined as the price of the $(i-v)$th transaction in dollars times the number of shares traded (denominated in hundreds of shares). That is, the dollar volume is in hundreds of dollars.
- SP5_{i-v} $(v = 1, 2, 3)$ is the 5-minute continuously compounded returns of the Standard and Poor's 500 index futures price for the contract maturing in the closest month beyond the month in which transaction $(i-v)$ occurred, where the return is computed with the futures price recorded 1 minute before the nearest round minute *prior* to t_{i-v} and the price recorded 5 minutes before this.
- IBS_{i-v} $(v = 1, 2, 3)$ is an indicator variable defined by

$$\text{IBS}_{i-v} = \begin{cases} 1 & \text{if } P_{i-v} > (P_{i-v}^a + P_{i-v}^b)/2, \\ 0 & \text{if } P_{i-v} = (P_{i-v}^a + P_{i-v}^b)/2, \\ -1 & \text{if } P_{i-v} < (P_{i-v}^a + P_{i-v}^b)/2, \end{cases}$$

where P_j^a and P_j^b are the ask and bid price at time t_j.

The parameter estimates and their t-ratios are given in Table 5.4. All the t-ratios are large except one, indicating that the estimates are highly significant. Such high t-ratios are not surprising as the sample size is large. For the heavily traded IBM stock, the estimation results suggest the following conclusions:

1. The boundary partitions are not equally spaced, but are almost symmetric with respect to zero.
2. The transaction duration Δt_i affects both the conditional mean and conditional variance of y_i in Eqs. (5.19) and (5.20).
3. The coefficients of lagged price changes are negative and highly significant, indicating *price reversals*.
4. As expected, the bid–ask spread at time t_{i-1} significantly affects the conditional variance.

Table 5.4. Parameter Estimates of the Ordered Probit Model in Eqs. (5.19) and (5.20) for the 1988 Transaction Data of IBM, Where t Denotes the t-Ratio[a]

Boundary Partitions of the Probit Model								
Parameter	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
Estimate	−4.67	−4.16	−3.11	−1.34	1.33	3.13	4.21	4.73
t	−145.7	−157.8	−171.6	−155.5	154.9	167.8	152.2	138.9

Equation Parameters of the Probit Model								
Parameter	γ_1	γ_2	β_1: Δt_i^*	β_2: y_{-1}	β_3	β_4	β_5	β_6
Estimate	0.40	0.52	−0.12	−1.01	−0.53	−0.21	1.12	−0.26
t	15.6	71.1	−11.4	−135.6	−85.0	−47.2	54.2	−12.1
Parameter	β_7	β_8	β_9:	β_{10}	β_{11}	β_{12}	β_{13}	
Estimate	0.01	−1.14	−0.37	−0.17	0.12	0.05	0.02	
t	0.26	−63.6	−21.6	−10.3	47.4	18.6	7.7	

[a]Reprinted with permission from Elsevier.

5.4.2 A Decomposition Model

An alternative approach to modeling price change is to decompose it into three components and use conditional specifications for the components; see Rydberg and Shephard (2003). The three components are an indicator for price change, the direction of price movement if there is a change, and the size of price change if a change occurs. Specifically, the price change at the ith transaction can be written as

$$y_i \equiv P_{t_i} - P_{t_{i-1}} = A_i D_i S_i, \tag{5.21}$$

where A_i is a binary variable defined as

$$A_i = \begin{cases} 1 & \text{if there is a price change at the } i\text{th trade,} \\ 0 & \text{if price remains the same at the } i\text{th trade,} \end{cases} \tag{5.22}$$

D_i is also a discrete variable signifying the *direction* of the price change if a change occurs—that is,

$$D_i|(A_i = 1) = \begin{cases} 1 & \text{if price increases at the } i\text{th trade,} \\ -1 & \text{if price drops at the } i\text{th trade,} \end{cases} \tag{5.23}$$

where $D_i|(A_i = 1)$ means that D_i is defined under the condition of $A_i = 1$, and S_i is the size of the price change in ticks if there is a change at the ith trade and $S_i = 0$ if there is no price change at the ith trade. When there is a price change, S_i is a positive integer-valued random variable.

Note that D_i is not needed when $A_i = 0$, and there is a natural ordering in the decomposition. D_i is well defined only when $A_i = 1$ and S_i is meaningful when $A_i = 1$ and D_i is given. Model specification under the decomposition makes use of the ordering.

Let F_i be the information set available at the ith transaction. Examples of elements in F_i are Δt_{i-j}, A_{i-j}, D_{i-j}, and S_{i-j} for $j \geq 0$. The evolution of price change under model (5.21) can then be partitioned as

$$P(y_i|F_{i-1}) = P(A_i D_i S_i|F_{i-1}) = P(S_i|D_i, A_i, F_{i-1})P(D_i|A_i, F_{i-1})P(A_i|F_{i-1}). \tag{5.24}$$

Since A_i is a binary variable, it suffices to consider the evolution of the probability $p_i = P(A_i = 1)$ over time. We assume that

$$\ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i\boldsymbol{\beta} \quad \text{or} \quad p_i = \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}, \tag{5.25}$$

where $\boldsymbol{x}_i$ is a finite-dimensional vector consisting of elements of F_{i-1} and $\boldsymbol{\beta}$ is a parameter vector. Conditioned on $A_i = 1$, D_i is also a binary variable, and we use the following model for $\delta_i = P(D_i = 1|A_i = 1)$:

$$\ln\left(\frac{\delta_i}{1-\delta_i}\right) = \boldsymbol{z}_i\boldsymbol{\gamma} \quad \text{or} \quad \delta_i = \frac{e^{\boldsymbol{z}_i\boldsymbol{\gamma}}}{1+e^{\boldsymbol{z}_i\boldsymbol{\gamma}}}, \tag{5.26}$$

where $\boldsymbol{z}_i$ is a finite-dimensional vector consisting of elements of F_{i-1} and $\boldsymbol{\gamma}$ is a parameter vector. To allow for asymmetry between positive and negative price changes, we assume that

$$S_i|(D_i, A_i = 1) \sim 1 + \begin{cases} g(\lambda_{u,i}) & \text{if } D_i = 1, A_i = 1, \\ g(\lambda_{d,i}) & \text{if } D_i = -1, A_i = 1, \end{cases} \tag{5.27}$$

where $g(\lambda)$ is a geometric distribution with parameter λ and the parameters $\lambda_{j,i}$ evolve over time as

$$\ln\left(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}\right) = \boldsymbol{w}_i\boldsymbol{\theta}_j \quad \text{or} \quad \lambda_{j,i} = \frac{e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}{1+e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}, \quad j = u, d, \tag{5.28}$$

where $\boldsymbol{w}_i$ is again a finite-dimensional explanatory variable in F_{i-1} and $\boldsymbol{\theta}_j$ is a parameter vector.

In Eq. (5.27), the probability mass function of a random variable x, which follows the geometric distribution $g(\lambda)$, is

$$p(x = m) = \lambda(1-\lambda)^m, \quad m = 0, 1, 2, \ldots.$$

We added 1 to the geometric distribution so that the price change, if it occurs, is at least 1 tick. In Eq. (5.28), we take the logistic transformation to ensure that $\lambda_{j,i} \in [0, 1]$.

The previous specification classifies the ith trade, or transaction, into one of three categories:

1. *No price change*: $A_i = 0$ and the associated probability is $(1 - p_i)$.
2. *A price increase*: $A_i = 1$, $D_i = 1$, and the associated probability is $p_i\delta_i$. The size of the price increase is governed by $1 + g(\lambda_{u,i})$.

3. *A price drop*: $A_i = 1$, $D_i = -1$, and the associated probability is $p_i(1 - \delta_i)$. The size of the price drop is governed by $1 + g(\lambda_{d,i})$.

Let $I_i(j)$ for $j = 1, 2, 3$ be the indicator variables of the prior three categories. That is, $I_i(j) = 1$ if the jth category occurs and $I_i(j) = 0$ otherwise. The log likelihood function of Eq. (5.24) becomes

$$\begin{aligned}
&\ln[P(y_i|F_{i-1})] \\
&\quad = I_i(1)\ln[(1-p_i)] + I_i(2)[\ln(p_i) + \ln(\delta_i) + \ln(\lambda_{u,i}) + (S_i - 1)\ln(1-\lambda_{u,i})] \\
&\qquad + I_i(3)[\ln(p_i) + \ln(1-\delta_i) + \ln(\lambda_{d,i}) + (S_i - 1)\ln(1-\lambda_{d,i})],
\end{aligned}$$

and the overall log likelihood function is

$$\ln[P(y_1, \ldots, y_n|F_0)] = \sum_{i=1}^{n} \ln[P(y_i|F_{i-1})], \tag{5.29}$$

which is a function of parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta}_u$, and $\boldsymbol{\theta}_d$.

Example 5.2. We illustrate the decomposition model by analyzing the intraday transactions of IBM stock from November 1, 1990 to January 31, 1991. There were 63 trading days and 59,838 intraday transactions in the normal trading hours. The explanatory variables used are:

1. A_{i-1}: the action indicator of the previous trade (i.e., the $(i-1)$th trade within a trading day).
2. D_{i-1}: the direction indicator of the previous trade.
3. S_{i-1}: the size of the previous trade.
4. V_{i-1}: the volume of the previous trade, divided by 1000.
5. Δt_{i-1}: time duration from the $(i-2)$th to $(i-1)$th trade.
6. BA_i: The bid–ask spread prevailing at the time of transaction.

Because we use lag-1 explanatory variables, the actual sample size is 59,775. It turns out that V_{i-1}, Δt_{i-1}, and BA_i are not statistically significant for the model entertained. Thus, only the first three explanatory variables are used. The model employed is

$$\begin{aligned}
\ln\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 A_{i-1}, \\
\ln\left(\frac{\delta_i}{1-\delta_i}\right) &= \gamma_0 + \gamma_1 D_{i-1}, \\
\ln\left(\frac{\lambda_{u,i}}{1-\lambda_{u,i}}\right) &= \theta_{u,0} + \theta_{u,1} S_{i-1}, \\
\ln\left(\frac{\lambda_{d,i}}{1-\lambda_{d,i}}\right) &= \theta_{d,0} + \theta_{d,1} S_{i-1}.
\end{aligned} \tag{5.30}$$

Table 5.5. Parameter Estimates of the ADS Model in Eq. (5.30) for IBM Intraday Transactions from 11/01/90 to 1/31/91

Parameter	β_0	β_1	γ_0	γ_1
Estimate	−1.057	0.962	−0.067	−2.307
Standard error	0.104	0.044	0.023	0.056
Parameter	$\theta_{u,0}$	$\theta_{u,1}$	$\theta_{d,0}$	$\theta_{d,1}$
Estimate	2.235	−0.670	2.085	−0.509
Standard error	0.029	0.050	0.187	0.139

The parameter estimates, using the log likelihood function in Eq. (5.29), are given in Table 5.5. The estimated simple model shows some dynamic dependence in the price change. In particular, the trade-by-trade price changes of IBM stock exhibit some appealing features:

1. The probability of a price change depends on the previous price change. Specifically, we have

$$P(A_i = 1|A_{i-1} = 0) = 0.258, \quad P(A_i = 1|A_{i-1} = 1) = 0.476.$$

The result indicates that a price change may occur in clusters and, as expected, most transactions are without price change. When no price change occurred at the $(i-1)$th trade, then only about one out of four trades in the subsequent transaction has a price change. When there is a price change at the $(i-1)$th transaction, the probability of a price change in the ith trade increases to about 0.5.

2. The direction of price change is governed by

$$P(D_i = 1|F_{i-1}, A_i) = \begin{cases} 0.483 & \text{if } D_{i-1} = 0 \text{ (i.e., } A_{i-1} = 0), \\ 0.085 & \text{if } D_{i-1} = 1,\ A_i = 1, \\ 0.904 & \text{if } D_{i-1} = -1,\ A_i = 1. \end{cases}$$

This result says that (a) if no price change occurred at the $(i-1)$th trade, then the chances for a price increase or decrease at the ith trade are about even; and (b) the probabilities of consecutive price increases or decreases are very low. The probability of a price increase at the ith trade given that a price change occurs at the ith trade and there was a price increase at the $(i-1)$th trade is only 8.6%. However, the probability of a price increase is about 90% given that a price change occurs at the ith trade and there was a price decrease at the $(i-1)$th trade. Consequently, this result shows the effect of bid–ask bounce and supports price reversals in high-frequency trading.

3. There is weak evidence suggesting that big price changes have a higher probability to be followed by another big price change. Consider the size of a price increase. We have

$$S_i|(D_i = 1) \sim 1 + g(\lambda_{u,i}), \quad \lambda_{u,i} = 2.235 - 0.670 S_{i-1}.$$

Using the probability mass function of a geometric distribution, we obtain that the probability of a price increase by one tick is 0.827 at the ith trade if the transaction results in a price increase and $S_{i-1} = 1$. The probability reduces to 0.709 if $S_{i-1} = 2$ and to 0.556 if $S_{i-1} = 3$. Consequently, the probability of a large S_i is proportional to S_{i-1} given that there is a price increase at the ith trade.

A difference between the ADS and ordered probit models is that the former does not require any truncation or grouping in the size of a price change.

5.5 DURATION MODELS

Duration models are concerned with time intervals between trades. Longer durations indicate lack of trading activities, which in turn signify a period of no new information. The dynamic behavior of durations thus contains useful information about intraday market activities. Using concepts similar to the ARCH models for volatility, Engle and Russell (1998) propose an autoregressive conditional duration (ACD) model to describe the evolution of time durations for (heavily traded) stocks. Zhang, Russell, and Tsay (2001a) extend the ACD model to account for nonlinearity and structural breaks in the data. In this section, we introduce some simple duration models. As mentioned before, intraday transactions exhibit some diurnal pattern. Therefore, we focus on the adjusted time duration

$$\Delta t_i^* = \Delta t_i / f(t_i), \tag{5.31}$$

where $f(t_i)$ is a deterministic function consisting of the cyclical component of Δt_i. Obviously, $f(t_i)$ depends on the underlying asset and the systematic behavior of the market. In practice, there are many ways to estimate $f(t_i)$, but no single method dominates the others in terms of statistical properties. A common approach is to use smoothing spline. Here we use simple quadratic functions and indicator variables to take care of the deterministic component of daily trading activities.

For the IBM data employed in the illustration of ADS models, we assume

$$f(t_i) = \exp[d(t_i)], \quad d(t_i) = \beta_0 + \sum_{j=1}^{7} \beta_j f_j(t_i), \tag{5.32}$$

where

$$f_1(t_i) = -\left(\frac{t_i - 43200}{14400}\right)^2, \quad f_3(t_i) = \begin{cases} -\left(\dfrac{t_i - 38700}{7500}\right)^2 & \text{if } t_i < 43200 \\ 0 & \text{otherwise,} \end{cases}$$

$$f_2(t_i) = -\left(\frac{t_i - 48300}{9300}\right)^2, \quad f_4(t_i) = \begin{cases} -\left(\dfrac{t_i - 48600}{9000}\right)^2 & \text{if } t_i \geq 43200 \\ 0 & \text{otherwise,} \end{cases}$$

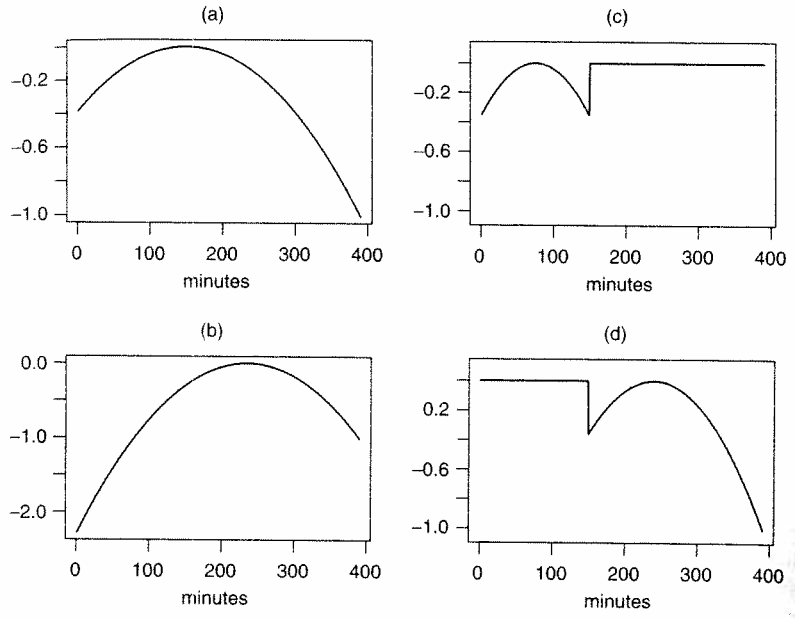

Figure 5.5. Quadratic functions used to remove the deterministic component of IBM intraday trading durations: (a)–(d) are the functions $f_1(.)$ to $f_4(.)$ of Eq. (5.32), respectively.

$f_5(t_i)$ and $f_6(t_i)$ are indicator variables for the first and second 5 minutes of market opening (i.e., $f_5(.) = 1$ if and only if t_i is between 9:30 am and 9:35 am Eastern time), and $f_7(t_i)$ is the indicator for the last 30 minutes of daily trading (i.e., $f_7(t_i) = 1$ if and only if the trade occurred between 3:30 pm and 4:00 pm Eastern time). Figure 5.5 shows the plot of $f_i(.)$ for $i = 1, \ldots, 4$, where the time scale on the x-axis is in minutes. Note that $f_3(43200) = f_4(43200)$, where 43,200 corresponds to 12:00 noon.

The coefficients β_j of Eq. (5.32) are obtained by the least squares method of the linear regression

$$\ln(\Delta t_i) = \beta_0 + \sum_{j=1}^{7} \beta_j f_j(t_i) + \epsilon_i.$$

The fitted model is

$$\ln(\widehat{\Delta t_i}) = 2.555 + 0.159 f_1(t_i) + 0.270 f_2(t_i) + 0.384 f_3(t_i) + 0.061 f_4(t_i) - 0.611 f_5(t_i) - 0.157 f_6(t_i) + 0.073 f_7(t_i).$$

Figure 5.6 shows the time plot of average durations in 5-minute time intervals over the 63 trading days before and after adjusting for the deterministic component. Figure 5.6a shows the average durations of Δt_i and, as expected, exhibits a diurnal

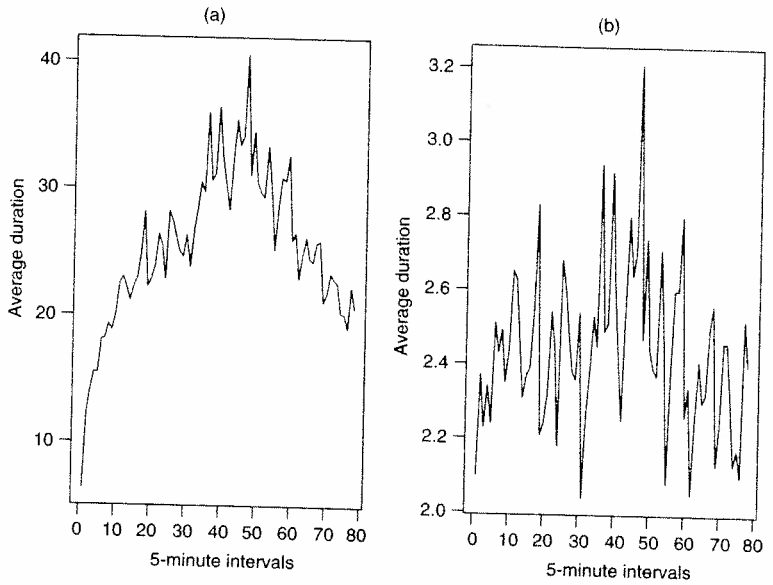

Figure 5.6. IBM transactions data from 11/01/90 to 1/31/91: (a) the average durations in 5-minute time intervals and (b) the average durations in 5-minute time intervals after adjusting for the deterministic component.

pattern. Figure 5.6b shows the average durations of Δt_i^* (i.e., after the adjustment), and the diurnal pattern is largely removed.

5.5.1 The ACD Model

The autoregressive conditional duration (ACD) model uses the idea of GARCH models to study the dynamic structure of the adjusted duration Δt_i^* of Eq. (5.31). For ease in notation, we define $x_i = \Delta t_i^*$.

Let $\psi_i = E(x_i|F_{i-1})$ be the conditional expectation of the adjusted duration between the $(i-1)$th and ith trades, where F_{i-1} is the information set available at the $(i-1)$th trade. In other words, ψ_i is the expected adjusted duration given F_{i-1}. The basic ACD model is defined as

$$x_i = \psi_i \epsilon_i, \tag{5.33}$$

where $\{\epsilon_i\}$ is a sequence of independent and identically distributed non-negative random variables such that $E(\epsilon_i) = 1$. In Engle and Russell (1998), ϵ_i follows a standard exponential or a standardized Weibull distribution, and ψ_i assumes the form

$$\psi_i = \omega + \sum_{j=1}^{r} \gamma_j x_{i-j} + \sum_{j=1}^{s} \omega_j \psi_{i-j}. \tag{5.34}$$

Such a model is referred to as an ACD(r, s) model. When the distribution of ϵ_i is exponential, the resulting model is called an EACD(r, s) model. Similarly, if ϵ_i follows a Weibull distribution, the model is a WACD(r, s) model. If necessary, readers are referred to Appendix A for a quick review of exponential and Weibull distributions.

Similar to GARCH models, the process $\eta_i = x_i - \psi_i$ is a martingale difference sequence (i.e., $E(\eta_i|F_{i-1}) = 0$), and the ACD(r, s) model can be written as

$$x_i = \omega + \sum_{j=1}^{\max(r,s)} (\gamma_j + \omega_j)x_{i-j} - \sum_{j=1}^{s} \omega_j \eta_{i-j} + \eta_j, \tag{5.35}$$

which is in the form of an ARMA process with non-Gaussian innovations. It is understood here that $\gamma_j = 0$ for $j > r$ and $\omega_j = 0$ for $j > s$. Such a representation can be used to obtain the basic conditions for weak stationarity of the ACD model. For instance, taking expectation on both sides of Eq. (5.35) and assuming weak stationarity, we have

$$E(x_i) = \frac{\omega}{1 - \sum_{j=1}^{\max(r,s)} (\gamma_j + \omega_j)}.$$

Therefore, we assume $\omega > 0$ and $1 > \sum_j (\gamma_j + \omega_j)$ because the expected duration is positive. As another application of Eq. (5.35), we study properties of the EACD(1,1) model.

EACD(1,1) Model

An EACD(1,1) model can be written as

$$x_i = \psi_i \epsilon_i, \quad \psi_i = \omega + \gamma_1 x_{i-1} + \omega_1 \psi_{i-1}, \tag{5.36}$$

where ϵ_i follows the standard exponential distribution. Using the moments of a standard exponential distribution in Appendix A, we have $E(\epsilon_i) = 1$, $\text{Var}(\epsilon_i) = 1$, and $E(\epsilon_i^2) = \text{Var}(x_i) + [E(x_i)]^2 = 2$. Assuming that x_i is weakly stationary (i.e., the first two moments of x_i are time-invariant), we derive the variance of x_i. First, taking the expectation of Eq. (5.36), we have

$$E(x_i) = E[E(\psi_i \epsilon_i|F_{i-1})] = E(\psi_i), \quad E(\psi_i) = \omega + \gamma_1 E(x_{i-1}) + \omega_1 E(\psi_{i-1}). \tag{5.37}$$

Under weak stationarity, $E(\psi_i) = E(\psi_{i-1})$ so that Eq. (5.37) gives

$$\mu_x \equiv E(x_i) = E(\psi_i) = \frac{\omega}{1 - \gamma_1 - \omega_1}. \tag{5.38}$$

Next, because $E(\epsilon_i^2) = 2$, we have $E(x_i^2) = E[E(\psi_i^2 \epsilon_i^2|F_{i-1})] = 2E(\psi_i^2)$.

Taking the square of ψ_i in Eq. (5.36) and the expectation and using weak stationarity of ψ_i and x_i, we have, after some algebra, that

$$E(\psi_i^2) = \mu_x^2 \times \frac{1 - (\gamma_1 + \omega_1)^2}{1 - 2\gamma_1^2 - \omega_1^2 - 2\gamma_1\omega_1}. \tag{5.39}$$

Finally, using $\text{Var}(x_i) = E(x_i^2) - [E(x_i)]^2$ and $E(x_i^2) = 2E(\psi_i^2)$, we have

$$\text{Var}(x_i) = 2E(\psi_i^2) - \mu_x^2 = \mu_x^2 \times \frac{1 - \omega_1^2 - 2\gamma_1\omega_1}{1 - \omega_1^2 - 2\gamma_1\omega_1 - 2\gamma_1^2},$$

where μ_x is defined in Eq. (5.38). This result shows that, to have time-invariant unconditional variance, the EACD(1,1) model in Eq. (5.36) must satisfy $1 > 2\gamma_1^2 + \omega_1^2 + 2\gamma_1\omega_1$. The variance of a WACD(1,1) model can be obtained by using the same techniques and the first two moments of a standardized Weibull distribution.

ACD Models with a Generalized Gamma Distribution

In the statistical literature, intensity function is often expressed in terms of hazard function. As shown in Appendix B, the hazard function of an EACD model is constant over time and that of a WACD model is a monotonous function. These hazard functions are rather restrictive in application as the intensity function of stock transactions might not be constant or monotone over time. To increase the flexibility of the associated hazard function, Zhang, Russell, and Tsay (2001a) employ a (standardized) generalized gamma distribution for ϵ_i. See Appendix A for some basic properties of a generalized gamma distribution. The resulting hazard function may assume various patterns, including U shape or inverted U shape. We refer to an ACD model with innovations that follow a generalized gamma distribution as a GACD(r, s) model.

5.5.2 Simulation

To illustrate ACD processes, we generated 500 observations from the ACD(1,1) model

$$x_i = \psi_i \epsilon_i, \quad \psi_i = 0.3 + 0.2x_{i-1} + 0.7\psi_{i-1} \tag{5.40}$$

using two different innovational distributions for ϵ_i. In case 1, ϵ_i is assumed to follow a standardized Weibull distribution with parameter $\alpha = 1.5$. In case 2, ϵ_i follows a (standardized) generalized gamma distribution with parameters $\kappa = 1.5$ and $\alpha = 0.5$.

Figure 5.7a shows the time plot of the WACD(1,1) series, whereas Figure 5.8a is the GACD(1,1) series. Figure 5.9 plots the histograms of both simulated series. The difference between the two models is evident. Finally, the sample ACFs of the two simulated series are shown in Figure 5.10a and Figure 5.11b, respectively. The serial dependence of the data is clearly seen.

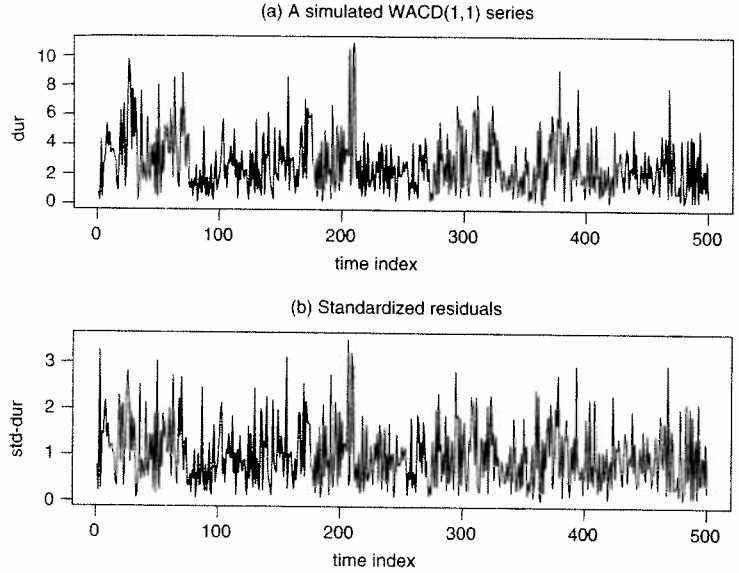

Figure 5.7. A simulated WACD(1,1) series in Eq. (5.40): (a) the original series and (b) the standardized series after estimation. There are 500 observations.

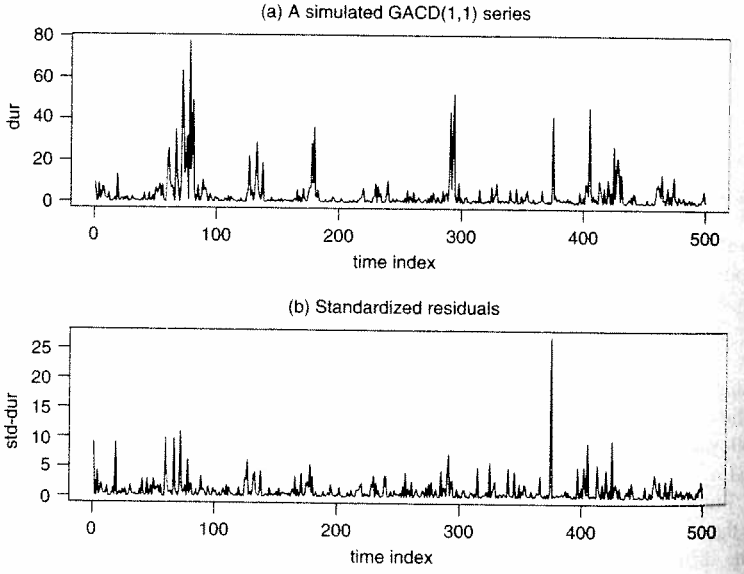

Figure 5.8. A simulated GACD(1,1) series in Eq. (5.40): (a) the original series and (b) the standardized series after estimation. There are 500 observations.

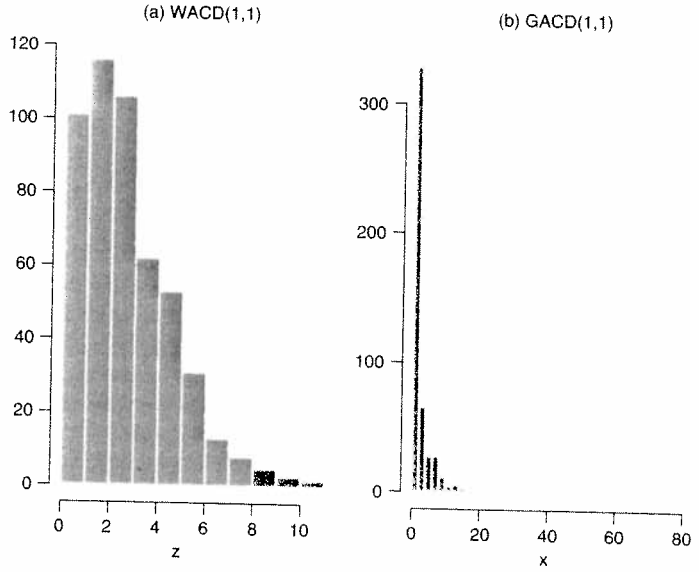

Figure 5.9. Histograms of simulated duration processes with 500 observations: (a) WACD(1,1) model and (b) GACD(1,1) model.

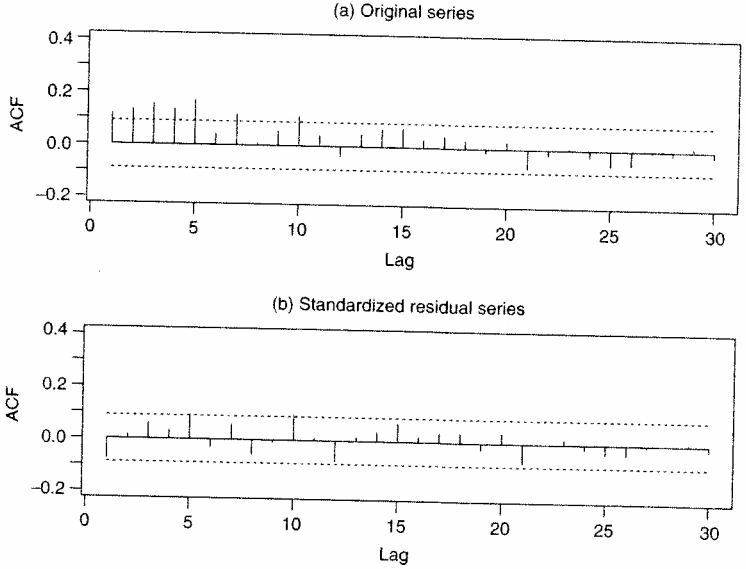

Figure 5.10. The sample autocorrelation function of a simulated WACD(1,1) series with 500 observations: (a) the original series and (b) the standardized residual series.

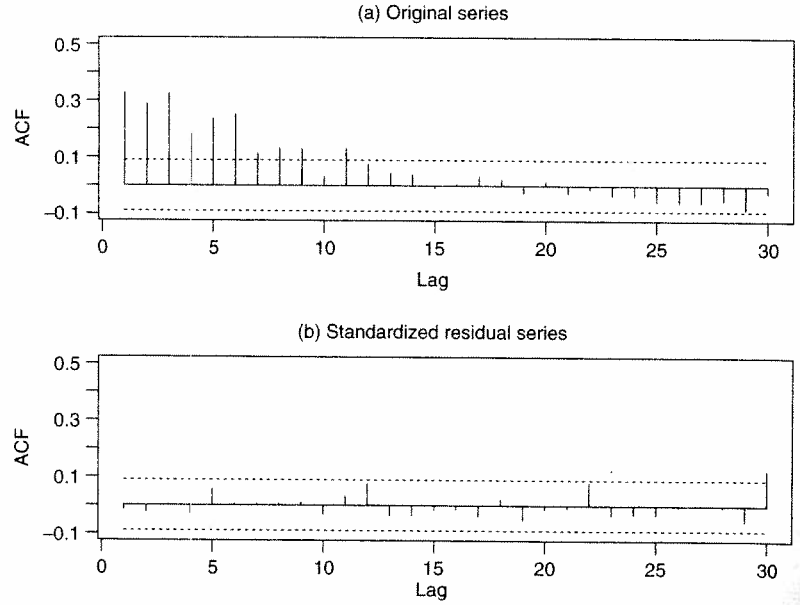

Figure 5.11. The sample autocorrelation function of a simulated GACD(1,1) series with 500 observations: (a) the original series and (b) the standardized residual series.

5.5.3 Estimation

For an ACD(r, s) model, let $i_o = \max(r, s)$ and $\boldsymbol{x}_t = (x_1, \ldots, x_t)'$. The likelihood function of the durations $x_1, \ldots, x_T$ is

$$f(\boldsymbol{x}_T|\boldsymbol{\theta}) = \left[\prod_{i=i_o+1}^{T} f(x_i|F_{i-1}, \boldsymbol{\theta})\right] \times f(\boldsymbol{x}_{i_o}|\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes the vector of model parameters, and T is the sample size. The marginal probability density function $f(\boldsymbol{x}_{i_o}|\boldsymbol{\theta})$ of the previous equation is rather complicated for a general ACD model. Because its impact on the likelihood function is diminishing as the sample size T increases, this marginal density is often ignored, resulting in use of the conditional likelihood method. For a WACD model, we use the probability density function (pdf) of Eq. (5.55) and obtain the conditional log likelihood function

$$\ell(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{x}_{i_o}) = \sum_{i=i_0+1}^{T} \alpha \ln\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right] + \ln\left(\frac{\alpha}{x_i}\right) + \alpha \ln\left(\frac{x_i}{\psi_i}\right) - \left(\frac{\Gamma(1+1/\alpha)x_i}{\psi_i}\right)^{\alpha}, \tag{5.41}$$

where $\psi_i = \omega + \sum_{j=1}^{r} \gamma_j x_{i-j} + \sum_{j=1}^{s} \omega_j \psi_{i-j}$, $\boldsymbol{\theta} = (\omega, \gamma_1, \ldots, \gamma_r, \omega_1, \ldots, \omega_s, \alpha)'$ and $\boldsymbol{x} = (x_{i_o+1}, \ldots, x_T)'$. When $\alpha = 1$, the (conditional) log likelihood function reduces to that of an EACD(r, s) model.

For a GACD(r, s) model, the conditional log likelihood function is

$$\ell(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{x}_{i_o}) = \sum_{i=i_o+1}^{T} \ln\left(\frac{\alpha}{\Gamma(\kappa)}\right) + (\kappa\alpha - 1)\ln(x_i) - \kappa\alpha \ln(\lambda\psi_i) - \left(\frac{x_i}{\lambda\psi_i}\right)^{\alpha}, \tag{5.42}$$

where $\lambda = \Gamma(\kappa)/\Gamma(\kappa + 1/\alpha)$ and the parameter vector $\boldsymbol{\theta}$ now also includes κ. As expected, when $\kappa = 1$, $\lambda = 1/\Gamma(1 + 1/\alpha)$ and the log likelihood function in Eq. (5.42) reduces to that of a WACD(r, s) model in Eq. (5.41). This log likelihood function can be rewritten in many ways to simplify the estimation.

Under some regularity conditions, the conditional maximum likelihood estimates are asymptotically normal; see Engle and Russell (1998) and the references therein. In practice, simulation can be used to obtain finite-sample reference distributions for the problem of interest once a duration model is specified.

Example 5.3. (Simulated ACD(1,1) series continued). Consider the simulated WACD(1,1) and GACD(1,1) series of Eq. (5.40). We apply the conditional likelihood method and obtain the results in Table 5.6. The estimates appear to be reasonable. Let $\hat{\psi}_i$ be the 1-step ahead prediction of ψ_i and $\hat{\epsilon}_i = x_i/\hat{\psi}_i$ be the standardized series, which can be regarded as standardized residuals of the series. If the model is adequately specified, $\{\hat{\epsilon}_i\}$ should behave as a sequence of independent and identically distributed random variables. Figure 5.7b and Figure 5.8b show the time plot of $\hat{\epsilon}_i$ for both models. The sample ACF of $\hat{\epsilon}_i$ for both fitted models are shown in Figure 5.10b and Figure 5.11b, respectively. It is evident that no significant serial correlations are found in the $\hat{\epsilon}_i$ series.

Table 5.6. Estimation Results for Simulated ACD(1,1) Series with 500 Observations for WACD(1,1) Series and GACD(1,1) Series

WACD(1,1) Model				
Parameter	ω	γ_1	ω_1	α
True	0.3	0.2	0.7	1.5
Estimate	0.364	0.100	0.767	1.477
Standard error	(0.139)	(0.025)	(0.060)	(0.052)

GACD(1,1) Model					
Parameter	ω	γ_1	ω_1	α	κ
True	0.3	0.2	0.7	0.5	1.5
Estimate	0.401	0.343	0.561	0.436	2.077
Standard error	(0.117)	(0.074)	(0.065)	(0.078)	(0.653)

Example 5.4. As an illustration of duration models, we consider the transaction durations of IBM stock on five consecutive trading days from November 1 to November 7, 1990. Focusing on positive transaction durations, we have 3534 observations. In addition, the data have been adjusted by removing the deterministic component in Eq. (5.32). That is, we employ 3534 positive adjusted durations as defined in Eq. (5.31).

Figure 5.12a shows the time plot of the adjusted (positive) durations for the first five trading days of November 1990, and Figure 5.13a gives the sample ACF of the series. There exist some serial correlations in the adjusted durations. We fit a WACD(1,1) model to the data and obtain the model

$$x_i = \psi_i \epsilon_i, \quad \psi_i = 0.169 + 0.064 x_{i-1} + 0.885 \psi_{i-1}, \tag{5.43}$$

where $\{\epsilon_i\}$ is a sequence of independent and identically distributed random variates that follow the standardized Weibull distribution with parameter $\hat{\alpha} = 0.879(0.012)$, where 0.012 is the estimated standard error. Standard errors of the estimates in Eq. (5.43) are 0.039, 0.010, and 0.018, respectively. All t-ratios of the estimates are greater than 4.2, indicating that the estimates are significant at the 1% level. Figure 5.12b shows the time plot of $\hat{\epsilon}_i = x_i/\hat{\psi}_i$, and Figure 5.13b provides the sample ACF of $\hat{\epsilon}_i$. The Ljung–Box statistics show $Q(10) = 4.96$ and $Q(20) =$

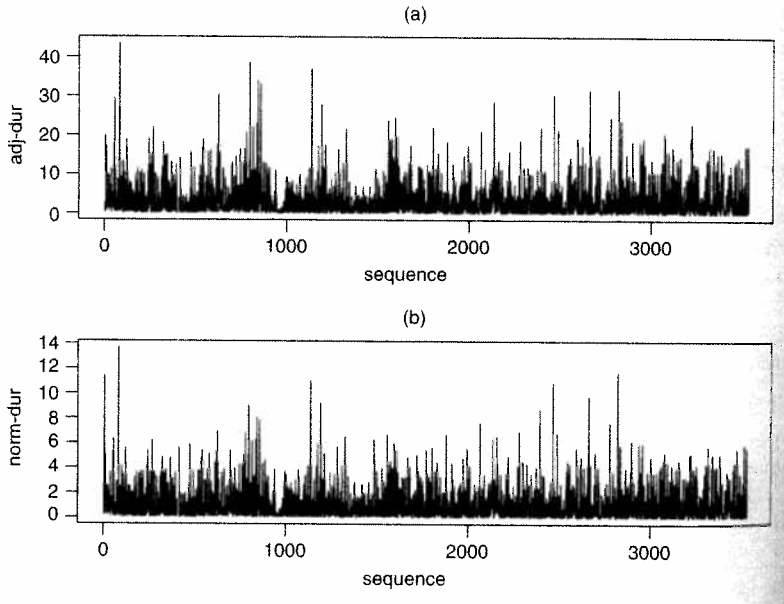

Figure 5.12. Time plots of durations for IBM stock traded in the first five trading days of November 1990: (a) the adjusted series and (b) the normalized innovations of an WACD(1,1) model. There are 3534 nonzero durations.

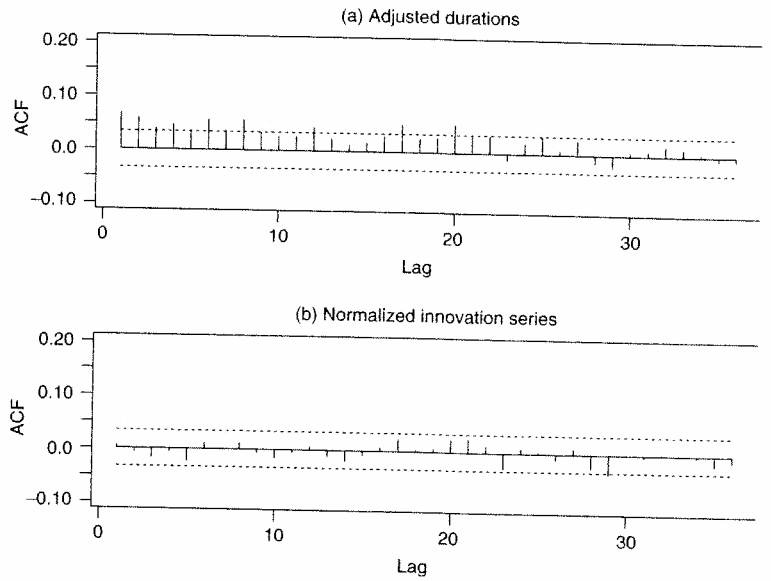

Figure 5.13. The sample autocorrelation function of adjusted durations for IBM stock traded in the first five trading days of November 1990: (a) the adjusted series and (b) the normalized innovations for a WACD(1,1) model.

10.75 for the $\hat{\epsilon}_i$ series. Clearly, the standardized innovations have no significant serial correlations. In fact, the sample autocorrelations of the squared series $\{\hat{\epsilon}_i^2\}$ are also small with $Q(10) = 6.20$ and $Q(20) = 11.16$, further confirming lack of serial dependence in the normalized innovations. In addition, the mean and standard deviation of a standardized Weibull distribution with $\alpha = 0.879$ are 1.00 and 1.14, respectively. These numbers are close to the sample mean and standard deviation of $\{\hat{\epsilon}_i\}$, which are 1.01 and 1.22, respectively. The fitted model seems adequate.

In model (5.43), the estimated coefficients show $\hat{\gamma}_1 + \hat{\omega}_1 \approx 0.949$, indicating certain persistence in the adjusted durations. The expected adjusted duration is $0.169/(1 - 0.064 - 0.885) = 3.31$ seconds, which is close to the sample mean 3.29 of the adjusted durations. The estimated α of the standardized Weibull distribution is 0.879, which is less than but close to 1. Thus, the conditional hazard function is monotonously decreasing at a slow rate.

If a generalized gamma distribution function is used for the innovations, then the fitted GACD(1,1) model is

$$x_i = \psi_i \epsilon_i, \quad \psi_i = 0.141 + 0.063 x_{i-1} + 0.897 \psi_{i-1}, \tag{5.44}$$

where $\{\epsilon_i\}$ follows a standardized, generalized gamma distribution in Eq. (5.56) with parameters $\kappa = 4.248(1.046)$ and $\alpha = 0.395(0.053)$, where the number in

parentheses denotes estimated standard error. Standard errors of the three parameters in Eq. (5.44) are 0.041, 0.010, and 0.019, respectively. All of the estimates are statistically significant at the 1% level. Again, the normalized innovational process $\{\hat{\epsilon}_i\}$ and its squared series have no significant serial correlation, where $\hat{\epsilon}_i = x_i/\hat{\psi}_i$ based on model (5.44). Specifically, for the $\hat{\epsilon}_i$ process, we have $Q(10) = 4.95$ and $Q(20) = 10.28$. For the $\hat{\epsilon}_i^2$ series, we have $Q(10) = 6.36$ and $Q(20) = 10.89$.

The expected duration of model (5.44) is 3.52, which is slightly greater than that of the WACD(1,1) model in Eq. (5.43). Similarly, the persistence parameter $\hat{\gamma}_1 + \hat{\omega}_1$ of model (5.44) is also slightly higher at 0.96.

Remark. Estimation of EACD models can be carried out by using programs for ARCH models with some minor modification; see Engle and Russell (1998). In this book, we use either the RATS program or some Fortran programs developed by the author to estimate the duration models. Limited experience indicates that it is harder to estimate a GACD model than an EACD or a WACD model. RATS programs used to estimate WACD and GACD models are given in Appendix C. □

5.6 NONLINEAR DURATION MODELS

Nonlinear features are also commonly found in high-frequency data. As an illustration, we apply some nonlinearity tests discussed in Chapter 4 to the normalized innovations $\hat{\epsilon}_i$ of the WACD(1,1) model for the IBM transaction durations in Example 5.4; see Eq. (5.43). Based on an AR(4) model, the test results are given in part (a) of Table 5.7. As expected from the model diagnostics of Example 5.4, the Ori-F test indicates no quadratic nonlinearity in the normalized innovations. However, the TAR-F test statistics suggest strong nonlinearity.

Based on the test results in Table 5.7, we entertain a threshold duration model with two regimes for the IBM intraday durations. The threshold variable is x_{t-1} (i.e., lag-1 adjusted duration). The estimated threshold value is 3.79. The fitted

Table 5.7. Nonlinearity Tests for IBM Transaction Durations from November 1 to November 7, 1990[a]

Type	Ori-F	TAR-$F(1)$	TAR-$F(2)$	TAR-$F(3)$	TAR-$F(4)$
(a) Normalized Innovations of a WACD(1,1) Model					
Test	0.343	3.288	3.142	3.128	0.297
p-Value	0.969	0.006	0.008	0.008	0.915
(b) Normalized Innovations of a Threshold WACD(1,1) Model					
Test	0.163	0.746	1.899	1.752	0.270
p-Value	0.998	0.589	0.091	0.119	0.929

[a]Only intraday durations are used. The number in parentheses of TAR-F tests denotes time delay.

threshold WACD(1,1) model is $x_i = \psi_i \epsilon_i$, where

$$\psi_i = \begin{cases} 0.020 + 0.257x_{i-1} + 0.847\psi_{i-1}, & \epsilon_i \sim w(0.901) \text{ if } x_{i-1} \le 3.79, \\ 1.808 + 0.027x_{i-1} + 0.501\psi_{i-1}, & \epsilon_i \sim w(0.845) \text{ if } x_{i-1} > 3.79, \end{cases} \tag{5.45}$$

where $w(\alpha)$ denotes a standardized Weibull distribution with parameter α. The number of observations in the two regimes are 2503 and 1030, respectively. In Eq. (5.45), the standard errors of the parameters for the first regime are 0.043, 0.041, 0.024, and 0.014, whereas those for the second regime are 0.526, 0.020, 0.147, and 0.020, respectively.

Consider the normalized innovations $\hat{\epsilon}_i = x_i/\hat{\psi}_i$ of the threshold WACD(1,1) model in Eq. (5.45). We obtain $Q(12) = 9.8$ and $Q(24) = 23.9$ for $\hat{\epsilon}_i$ and $Q(12) = 8.0$ and $Q(24) = 16.7$ for $\hat{\epsilon}_i^2$. Thus, there are no significant serial correlations in the $\hat{\epsilon}_i$ and $\hat{\epsilon}_i^2$ series. Furthermore, applying the same nonlinearity tests as before to this newly normalized innovational series $\hat{\epsilon}_i$, we detect no nonlinearity; see part (b) of Table 5.7. Consequently, the two-regime threshold WACD(1,1) model in Eq. (5.45) is adequate.

If we classify the two regimes as heavy and thin trading periods, then the threshold model suggests that the trading dynamics measured by intraday transaction durations are different between heavy and thin trading periods for IBM stock even after the adjustment of diurnal pattern. This is not surprising as market activities are often driven by the arrival of news and other information.

The estimated threshold WACD(1,1) model in Eq. (5.45) contains some insignificant parameters. We refine the model and obtain the result:

$$\psi_i = \begin{cases} 0.225x_{i-1} + 0.867\psi_{i-1}, & \epsilon_i \sim w(0.902) \text{ if } x_{i-1} \le 3.79, \\ 1.618 + 0.614\psi_{i-1}, & \epsilon_i \sim w(0.846) \text{ if } x_{i-1} > 3.79. \end{cases}$$

All of the estimates of the refined model are highly significant. The Ljung–Box statistics of the standardized innovations $\hat{\epsilon}_i = x_i/\hat{\psi}_i$ show $Q(10) = 5.91(0.82)$ and $Q(20) = 16.04(0.71)$ and those of $\hat{\epsilon}_i^2$ give $Q(10) = 5.35(0.87)$ and $Q(20) = 15.20(0.76)$, where the number in parentheses is the p-value. Therefore, the refined model is adequate. The RATS program used to estimate the prior model is given in Appendix C.

5.7 BIVARIATE MODELS FOR PRICE CHANGE AND DURATION

In this section, we introduce a model that considers jointly the process of price change and the associated duration. As mentioned before, many intraday transactions of a stock result in no price change. Those transactions are highly relevant to trading intensity, but they do not contain direct information on price movement. Therefore, to simplify the complexity involved in modeling price change, we focus on transactions that result in a price change and consider a price change

and duration (PCD) model to describe the multivariate dynamics of price change and the associated time duration.

We continue to use the same notation as before, but the definition is changed to transactions with a price change. Let t_i be the calendar time of the ith price change of an asset. As before, t_i is measured in seconds from midnight of a trading day. Let P_{t_i} be the transaction price when the ith price change occurred and $\Delta t_i = t_i - t_{i-1}$ be the time duration between price changes. In addition, let N_i be the number of trades in the time interval (t_{i-1}, t_i) that result in no price change. This new variable is used to represent trading intensity during a period of no price change. Finally, let D_i be the direction of the ith price change with $D_i = 1$ when price goes up and $D_i = -1$ when the price comes down, and let S_i be the size of the ith price change measured in ticks. Under the new definitions, the price of a stock evolves over time by

$$P_{t_i} = P_{t_{i-1}} + D_i S_i, \tag{5.46}$$

and the transactions data consist of $\{\Delta t_i, N_i, D_i, S_i\}$ for the ith price change. The PCD model is concerned with the joint analysis of $(\Delta t_i, N_i, D_i, S_i)$.

Remark. Focusing on transactions associated with a price change can reduce the sample size dramatically. For example, consider the intraday data of IBM stock from November 1, 1990 to January 31, 1991. There were 60,265 intraday trades, but only 19,022 of them resulted in a price change. In addition, there is no diurnal pattern in time durations between price changes. □

To illustrate the relationship among the price movements of all transactions and those of transactions associated with a price change, we consider the intraday tradings of IBM stock on November 21, 1990. There were 726 transactions on that day during normal trading hours, but only 195 trades resulted in a price change. Figure 5.14 shows the time plot of the price series for both cases. As expected, the price series are the same.

The PCD model decomposes the joint distribution of $(\Delta t_i, N_i, D_i, S_i)$ given F_{i-1} as

$$\begin{aligned} &f(\Delta t_i, N_i, D_i, S_i | F_{i-1}) \\ &\quad = f(S_i | D_i, N_i, \Delta t_i, F_{i-1}) f(D_i | N_i, \Delta t_i, F_{i-1}) f(N_i | \Delta t_i, F_{i-1}) f(\Delta t_i | F_{i-1}). \end{aligned} \tag{5.47}$$

This partition enables us to specify suitable econometric models for the conditional distributions and, hence, to simplify the modeling task. There are many ways to specify models for the conditional distributions. A proper specification might depend on the asset under study. Here we employ the specifications used by McCulloch and Tsay (2000), who use generalized linear models for the discrete-valued variables and a time series model for the continuous variable $\ln(\Delta t_i)$.

For the time duration between price changes, we use the model

$$\ln(\Delta t_i) = \beta_0 + \beta_1 \ln(\Delta t_{i-1}) + \beta_2 S_{i-1} + \sigma \epsilon_i, \tag{5.48}$$

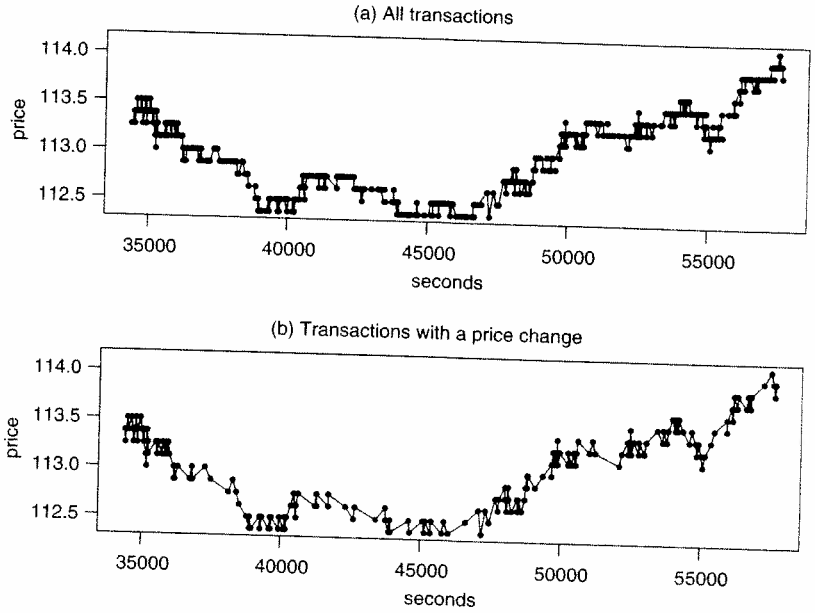

Figure 5.14. Time plots of the intraday transaction prices of IBM stock on November 21, 1990: (a) all transactions and (b) transactions that resulted in a price change.

where σ is a positive number and $\{\epsilon_i\}$ is a sequence of iid $N(0, 1)$ random variables. This is a multiple linear regression model with lagged variables. Other explanatory variables can be added if necessary. The log transformation is used to ensure the positiveness of time duration.

The conditional model for N_i is further partitioned into two parts because empirical data suggest a concentration of N_i at 0. The first part of the model for N_i is the logit model

$$p(N_i = 0 | \Delta t_i, F_{i-1}) = \text{logit}[\alpha_0 + \alpha_1 \ln(\Delta t_i)], \tag{5.49}$$

where $\text{logit}(x) = \exp(x)/[1 + \exp(x)]$, whereas the second part of the model is

$$N_i | (N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}{1 + \exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}, \tag{5.50}$$

where $\sim$ means "is distributed as," and $g(\lambda)$ denotes a geometric distribution with parameter λ, which is in the interval $(0, 1)$.

The model for direction D_i is

$$D_i | (N_i, \Delta t_i, F_{i-1}) = \text{sign}(\mu_i + \sigma_i \epsilon), \tag{5.51}$$

where ϵ is a $N(0, 1)$ random variable, and

$$\mu_i = \omega_0 + \omega_1 D_{i-1} + \omega_2 \ln(\Delta t_i),$$

$$\ln(\sigma_i) = \beta \left| \sum_{j=1}^{4} D_{i-j} \right| = \beta |D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|.$$

In other words, D_i is governed by the sign of a normal random variable with mean μ_i and variance σ_i^2. A special characteristic of the prior model is the function for $\ln(\sigma_i)$. For intraday transactions, a key feature is the *price reversal* between consecutive price changes. This feature is modeled by the dependence of D_i on D_{i-1} in the mean equation with a negative ω_1 parameter. However, there exists an occasional local trend in the price movement. The previous variance equation allows for such a local trend by increasing the uncertainty in the direction of price movement when the past data showed evidence of a local trend. For a normal distribution with a fixed mean, increasing its variance makes a random draw have the same chance to be positive and negative. This in turn increases the chance for a sequence of all positive or all negative draws. Such a sequence produces a local trend in price movement.

To allow for different dynamics between positive and negative price movements, we use different models for the size of a price change. Specifically, we have

$$S_i | (D_i = -1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{d,i}) + 1, \quad \text{with}$$

$$\ln(\lambda_{d,i}) = \eta_{d,0} + \eta_{d,1} N_i + \eta_{d,2} \ln(\Delta t_i) + \eta_{d,3} S_{i-1} \tag{5.52}$$

$$S_i | (D_i = 1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{u,i}) + 1, \quad \text{with}$$

$$\ln(\lambda_{u,i}) = \eta_{u,0} + \eta_{u,1} N_i + \eta_{u,2} \ln(\Delta t_i) + \eta_{u,3} S_{i-1}, \tag{5.53}$$

where $p(\lambda)$ denotes a Poisson distribution with parameter λ, and 1 is added to the size because the minimum size is 1 tick when there is a price change.

The specified models in Eqs. (5.48)–(5.53) can be estimated jointly by either the maximum likelihood method or the Markov chain Monte Carlo methods. Based on Eq. (5.47), the models consist of six conditional models that can be estimated separately.

Example 5.5. Consider the intraday transactions of IBM stock on November 21, 1990. There are 194 price changes within normal trading hours. Figure 5.15 shows the histograms of $\ln(\Delta t_i)$, N_i, D_i, and S_i. The data for D_i are about equally distributed between "upward" and "downward" movements. Only a few transactions resulted in a price change of more than 1 tick; as a matter of fact, there were seven changes with two ticks and one change with three ticks. Using Markov chain Monte Carlo (MCMC) methods (see Chapter 12), we obtained the following models for the data. The reported estimates and their standard deviations are the posterior means and standard deviations of MCMC draws with 9500 iterations. The model for the time duration between price changes is

$$\ln(\Delta t_i) = 4.023 + 0.032 \ln(\Delta t_{i-1}) - 0.025 S_{i-1} + 1.403 \epsilon_i,$$

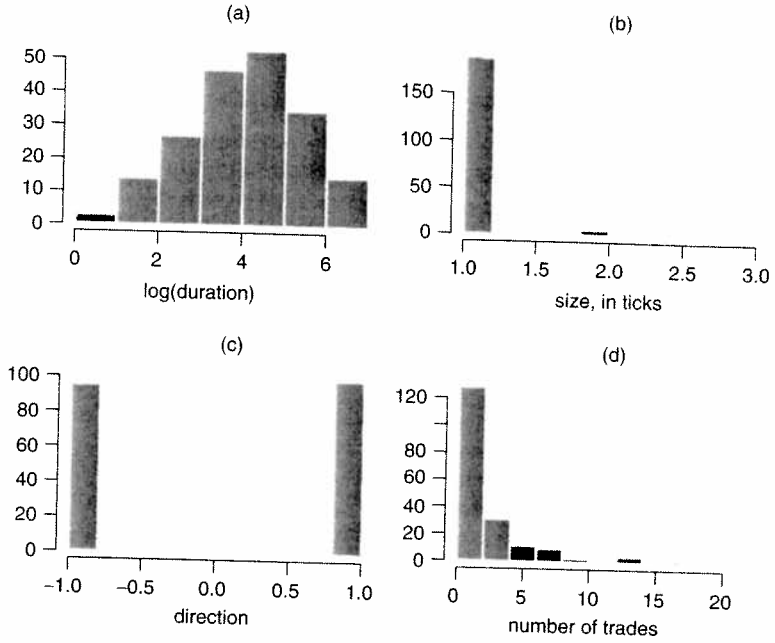

Figure 5.15. Histograms of intraday transactions data for IBM stock on November 21, 1990: (a) log durations between price changes, (b) direction of price movement, (c) size of price change measured in ticks, and (d) number of trades without a price change.

where standard deviations of the coefficients are 0.415, 0.073, 0.384, and 0.073, respectively. The fitted model indicates that there was no dynamic dependence in the time duration. For the N_i variable, we have

$$\Pr(N_i > 0 | \Delta t_i, F_{i-1}) = \text{logit}[-0.637 + 1.740 \ln(\Delta t_i)],$$

where standard deviations of the estimates are 0.238 and 0.248, respectively. Thus, as expected, the number of trades with no price change in the time interval (t_{i-1}, t_i) depends positively on the length of the interval. The magnitude of N_i when it is positive is

$$N_i | (N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[0.178 - 0.910 \ln(\Delta t_i)]}{1 + \exp[0.178 - 0.910 \ln(\Delta t_i)]},$$

where standard deviations of the estimates are 0.246 and 0.138, respectively. The negative and significant coefficient of $\ln(\Delta t_i)$ means that N_i is positively related to the length of the duration Δt_i because a large $\ln(\Delta t_i)$ implies a small λ_i, which in turn implies higher probabilities for larger N_i; see the geometric distribution in Eq. (5.27).

The fitted model for D_i is

$$\mu_i = 0.049 - 0.840 D_{i-1} - 0.004 \ln(\Delta t_i),$$
$$\ln(\sigma_i) = 0.244|D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|,$$

where standard deviations of the parameters in the mean equation are 0.129, 0.132, and 0.082, respectively, whereas the standard deviation for the parameter in the variance equation is 0.182. The price reversal is clearly shown by the highly significant negative coefficient of D_{i-1}. The marginally significant parameter in the variance equation is exactly as expected. Finally, the fitted models for the size of a price change are

$$\ln(\lambda_{d,i}) = 1.024 - 0.327 N_i + 0.412 \ln(\Delta t_i) - 4.474 S_{i-1},$$
$$\ln(\lambda_{u,i}) = -3.683 - 1.542 N_i + 0.419 \ln(\Delta t_i) + 0.921 S_{i-1},$$

where standard deviations of the parameters for the "down size" are 3.350, 0.319, 0.599, and 3.188, respectively, whereas those for the "up size" are 1.734, 0.976, 0.453, and 1.459. The interesting estimates of the prior two equations are the negative estimates of the coefficient of N_i. A large N_i means there were more transactions in the time interval (t_{i-1}, t_i) with no price change. This can be taken as evidence of no new information available in the time interval (t_{i-1}, t_i). Consequently, the size for the price change at t_i should be small. A small $\lambda_{u,i}$ or $\lambda_{d,i}$ for a Poisson distribution gives precisely that.

In summary, granted that a sample of 194 observations in a given day may not contain sufficient information about the trading dynamics of IBM stock, but the fitted models appear to provide some sensible results. McCulloch and Tsay (2000) extend the PCD model to a hierarchical framework to handle all the data of the 63 trading days between November 1, 1990 and January 31, 1991. Many of the parameter estimates become significant in this extended sample, which has more than 19,000 observations. For example, the overall estimate of the coefficient of $\ln(\Delta t_{i-1})$ in the model for time duration ranges from 0.04 to 0.1, which is small, but significant.

Finally, using transactions data to test microstructure theory often requires a careful specification of the variables used. It also requires a deep understanding of the way by which the market operates and the data are collected. However, ideas of the econometric models discussed in this chapter are useful and widely applicable in analysis of high-frequency data.

APPENDIX A: REVIEW OF SOME PROBABILITY DISTRIBUTIONS

Exponential Distribution

A random variable X has an exponential distribution with parameter $\beta > 0$ if its probability density function (pdf) is given by

$$f(x|\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Denoting such a distribution by $X \sim \exp(\beta)$, we have $E(X) = \beta$ and $\text{Var}(X) = \beta^2$. The cumulative distribution function (CDF) of X is

$$F(x|\beta) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-x/\beta} & \text{if } x \geq 0. \end{cases}$$

When $\beta = 1$, X is said to have a standard exponential distribution.

Gamma Function

For $\kappa > 0$, the gamma function $\Gamma(\kappa)$ is defined by

$$\Gamma(\kappa) = \int_0^\infty x^{\kappa-1} e^{-x} dx.$$

The most important properties of the gamma function are:

1. For any $\kappa > 1$, $\Gamma(\kappa) = (\kappa - 1)\Gamma(\kappa - 1)$.
2. For any positive integer m, $\Gamma(m) = (m - 1)!$.
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

The integration

$$\Gamma(y|\kappa) = \int_0^y x^{\kappa-1} e^{-x} dx, \quad y > 0$$

is an *incomplete* gamma function. Its values have been tabulated in the literature. Computer programs are now available to evaluate the incomplete gamma function.

Gamma Distribution

A random variable X has a gamma distribution with parameter κ and β ($\kappa > 0$, $\beta > 0$) if its pdf is given by

$$f(x|\kappa, \beta) = \begin{cases} \frac{1}{\beta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\beta} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

By changing variable $y = x/\beta$, one can easily obtain the moments of X:

$$E(X^m) = \int_0^\infty x^m f(x|\kappa, \beta) dx = \frac{1}{\beta^\kappa \Gamma(\kappa)} \int_0^\infty x^{\kappa+m-1} e^{-x/\beta} dx$$
$$= \frac{\beta^m}{\Gamma(\kappa)} \int_0^\infty y^{\kappa+m-1} e^{-y} dy = \frac{\beta^m \Gamma(\kappa + m)}{\Gamma(\kappa)}.$$

In particular, the mean and variance of X are $E(X) = \kappa\beta$ and $\text{Var}(X) = \kappa\beta^2$. When $\beta = 1$, the distribution is called a standard gamma distribution with parameter κ.

We use the notation $G \sim \text{gamma}(\kappa)$ to denote that G follows a standard gamma distribution with parameter κ. The moments of G are

$$E(G^m) = \frac{\Gamma(\kappa + m)}{\Gamma(\kappa)}, \quad m > 0. \tag{5.54}$$

Weibull Distribution

A random variable X has a Weibull distribution with parameters α and β ($\alpha > 0$, $\beta > 0$) if its pdf is given by

$$f(x|\alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

where β and α are the scale and shape parameters of the distribution. The mean and variance of X are

$$E(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right), \quad \text{Var}(X) = \beta^2 \left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\},$$

and the CDF of X is

$$F(x|\alpha, \beta) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-(x/\beta)^\alpha} & \text{if } x \geq 0. \end{cases}$$

When $\alpha = 1$, the Weibull distribution reduces to an exponential distribution.

Define $Y = X/[\beta\Gamma(1 + 1/\alpha)]$. We have $E(Y) = 1$ and the pdf of Y is

$$f(y|\alpha) = \begin{cases} \alpha\left[\Gamma\left(1 + \dfrac{1}{\alpha}\right)\right]^\alpha y^{\alpha-1} \exp\left\{-\left[\Gamma\left(1 + \dfrac{1}{\alpha}\right) y\right]^\alpha\right\} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5.55}$$

where the scale parameter β disappears due to standardization. The CDF of the standardized Weibull distribution is

$$F(y|\alpha) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - \exp\left\{-\left[\Gamma\left(1 + \dfrac{1}{\alpha}\right) y\right]^\alpha\right\} & \text{if } y > 0, \end{cases}$$

and we have $E(Y) = 1$ and $\text{Var}(Y) = \Gamma(1 + 2/\alpha)/[\Gamma(1 + 1/\alpha)]^2 - 1$. For a duration model with Weibull innovations, the pdf in Eq. (5.55) is used in the maximum likelihood estimation.

Generalized Gamma Distribution

A random variable X has a generalized gamma distribution with parameter α, β, κ ($\alpha > 0$, $\beta > 0$, and $\kappa > 0$) if its pdf is given by

$$f(x|\alpha, \beta, \kappa) = \begin{cases} \dfrac{\alpha x^{\kappa\alpha-1}}{\beta^{\kappa\alpha}\Gamma(\kappa)} \exp\left[-\left(\dfrac{x}{\beta}\right)^\alpha\right] & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where β is a scale parameter, and α and κ are shape parameters. This distribution can be written as

$$G = \left(\frac{X}{\beta}\right)^\alpha,$$

where G is a standard gamma random variable with parameter κ. The pdf of X can be obtained from that of G by the technique of changing variables. Similarly, the moments of X can be obtained from that of G in Eq. (5.54) by

$$E(X^m) = E[(\beta G^{1/\alpha})^m] = \beta^m E(G^{m/\alpha}) = \beta^m \frac{\Gamma(\kappa + m/\alpha)}{\Gamma(\kappa)} = \frac{\beta^m \Gamma(\kappa + m/\alpha)}{\Gamma(\kappa)}.$$

When $\kappa = 1$, the generalized gamma distribution reduces to that of a Weibull distribution. Thus, the exponential and Weibull distributions are special cases of the generalized gamma distribution.

The expectation of a generalized gamma distribution is $E(X) = \beta\Gamma(\kappa + 1/\alpha)/\Gamma(\kappa)$. In duration models, we need a distribution with unit expectation. Therefore, defining a random variable $Y = \lambda X/\beta$, where $\lambda = \Gamma(\kappa)/\Gamma(\kappa + 1/\alpha)$, we have $E(Y) = 1$ and the pdf of Y is

$$f(y|\alpha, \kappa) = \begin{cases} \dfrac{\alpha y^{\kappa\alpha-1}}{\lambda^{\kappa\alpha}\Gamma(\kappa)} \exp\left[-\left(\dfrac{y}{\lambda}\right)^\alpha\right] & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5.56}$$

where again the scale parameter β disappears and $\lambda = \Gamma(\kappa)/\Gamma(\kappa + 1/\alpha)$.

APPENDIX B: HAZARD FUNCTION

A useful concept in modeling duration is the *hazard function* implied by a distribution function. For a random variable X, the *survival function* is defined as

$$S(x) \equiv P(X > x) = 1 - P(X \leq x) = 1 - \text{CDF}(x), \quad x > 0,$$

which gives the probability that a subject, which follows the distribution of X, survives at the time x. The hazard function (or intensity function) of X is then defined by

$$h(x) = \frac{f(x)}{S(x)}, \tag{5.57}$$

where $f(.)$ and $S(.)$ are the pdf and survival function of X, respectively.

Example 5.6. For the Weibull distribution with parameters α and β, the survival function and hazard function are

$$S(x|\alpha,\beta) = \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \quad h(x|\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}, \quad x > 0.$$

In particular, when $\alpha = 1$, we have $h(x|\beta) = 1/\beta$. Therefore, for an exponential distribution, the hazard function is constant. For a Weibull distribution, the hazard is a monotone function. If $\alpha > 1$, then the hazard function is monotonously increasing. If $\alpha < 1$, the hazard function is monotonously decreasing. For the generalized gamma distribution, the survival function and, hence, the hazard function involve the incomplete gamma function. Yet the hazard function may exhibit various patterns, including U shape or inverted U shape. Thus, the generalized gamma distribution provides a flexible approach to modeling the duration of stock transactions.

For the standardized Weibull distribution, the survival and hazard functions are

$$S(y|\alpha) = \exp\left\{-\left[\Gamma\left(1+\frac{1}{\alpha}\right)y\right]^{\alpha}\right\}, \quad h(y|\alpha) = \alpha\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^{\alpha} y^{\alpha-1}, \quad y > 0.$$

APPENDIX C: SOME RATS PROGRAMS FOR DURATION MODELS

The data used are adjusted time durations of intraday transactions of IBM stock from November 1 to November 9, 1990. The file name is `ibm1to5.txt` and it has 3534 observations.

Program for Estimating a WACD(1,1) Model

```
all 0  3534:1
open data ibm1to5.txt
data(org=obs) / x r1
set psi = 1.0
nonlin a0 a1 b1 a1
frml gvar = a0+a1*x(t-1)+b1*psi(t-1)
frml gma = %LNGAMMA(1.0+1.0/a1)
frml gln =a1*gma(t)+log(a1)-log(x(t)) $
  +a1*log(x(t)/(psi(t)=gvar(t)))-(exp(gma(t))*x(t)/psi(t))**a1
smpl 2 3534
compute a0 = 0.2, a1 = 0.1, b1 = 0.1, a1 = 0.8
maximize(method=bhhh,recursive,iterations=150) gln
set fv = gvar(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```


Program for Estimating a GACD(1,1) Model

```
all 0  3534:1
open data ibm1to5.txt
data(org=obs) / x r1
set psi = 1.0
nonlin a0 a1 b1 a1 ka
frml cv = a0+a1*x(t-1)+b1*psi(t-1)
frml gma = %LNGAMMA(ka)
frml lam = exp(gma(t))/exp(%LNGAMMA(ka+(1.0/a1)))
frml xlam = x(t)/(lam(t)*(psi(t)=cv(t)))
frml gln =-gma(t)+log(a1/x(t))+ka*a1*log(xlam(t))-(xlam(t))**a1
smpl 2 3534
compute a0 = 0.238, a1 = 0.075, b1 = 0.857, a1 = 0.5, ka = 4.0
nlpar(criterion=value,cvcrit=0.00001)
maximize(method=bhhh,recursive,iterations=150) gln
set fv = cv(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

Program for Estimating a TAR-WACD(1,1) Model

The threshold 3.79 is prespecified.

```
all 0 3534:1
open data ibm1to5.txt
data(org=obs) / x rt
set psi = 1.0
nonlin a1 a2 a1 b0 b2 b1
frml u = ((x(t-1)-3.79)/abs(x(t-1)-3.79)+1.0)/2.0
frml cp1 = a1*x(t-1)+a2*psi(t-1)
frml gma1 = %LNGAMMA(1.0+1.0/a1)
frml cp2 = b0+b2*psi(t-1)
frml gma2 = %LNGAMMA(1.0+1.0/b1)
frml cp = cp1(t)*(1-u(t))+cp2(t)*u(t)
frml gln1 =a1*gma1(t)+log(a1)-log(x(t)) $
 +a1*log(x(t)/(psi(t)=cp(t)))-(exp(gma1(t))*x(t)/psi(t))**a1
frml gln2 =b1*gma2(t)+log(b1)-log(x(t)) $
 +b1*log(x(t)/(psi(t)=cp(t)))-(exp(gma2(t))*x(t)/psi(t))**b1
frml gln = gln1(t)*(1-u(t))+gln2(t)*u(t)
smpl 2 3534
compute a1 = 0.2, a2 = 0.85, a1 = 0.9
compute b0 = 1.8, b2 = 0.5, b1 = 0.8
maximize(method=bhhh,recursive,iterations=150) gln
set fv = cp(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

EXERCISES

5.1. Let r_t be the log return of an asset at time t. Assume that $\{r_t\}$ is a Gaussian white noise series with mean 0.05 and variance 1.5. Suppose that the probability of a trade at each time point is 40% and is independent of r_t. Denote the observed return by r_t^o. Is r_t^o serially correlated? If yes, calculate the first three lags of autocorrelations of r_t^o.

5.2. Let P_t be the observed market price of an asset, which is related to the fundamental value of the asset P_t^* via Eq. (5.9). Assume that $\Delta P_t^* = P_t^* - P_{t-1}^*$ forms a Gaussian white noise series with mean zero and variance 1.0. Suppose that the bid–ask spread is two ticks. What is the lag-1 autocorrelation of the price change series $\Delta P_t = P_t - P_{t-1}$ when the tick size is \$1/8? What is the lag-1 autocorrelation of the price change when the tick size is \$1/16?

5.3. The file `ibm-d2-dur.txt` contains the adjusted durations between trades of IBM stock on November 2, 1990. The file has three columns consisting of day, time of trade measured in seconds from midnight, and adjusted durations.

(a) Build an EACD model for the adjusted duration and check the fitted model.

(b) Build a WACD model for the adjusted duration and check the fitted model.

(c) Build a GACD model for the adjusted duration and check the fitted model.

(d) Compare the prior three duration models.

5.4. The file `mmm9912-dtp.txt` contains the transactions data of the stock of 3M Company in December 1999. There are three columns: day of the month, time of transaction in seconds from midnight, and transaction price. Transactions that occurred after 4:00 pm Eastern time are excluded.

(a) Is there a diurnal pattern in 3M stock trading? You may construct a time series n_t, which denotes the number of trades in a 5-minute time interval to answer this question.

(b) Use the price series to confirm the existence of a bid–ask bounce in intraday trading of 3M stock.

(c) Tabulate the frequencies of price change in multiples of tick size \$1/16. You may combine changes with 5 ticks or more into a category and those with -5 ticks or beyond into another category.

5.5. Consider again the transactions data of 3M stock in December 1999.

(a) Use the data to construct an intraday 5-minute log return series. Use the simple average of all transaction prices within a 5-minute interval as the stock price for the interval. Is the series serially correlated? You may use Ljung–Box statistics to test the hypothesis with the first 10 lags of the sample autocorrelation function.

(b) There are seventy-seven 5-minute returns in a normal trading day. Some researchers suggest that the sum of squares of the intraday 5-minute returns can be used as a measure of daily volatility. Apply this approach and calculate the daily volatility of the log return of 3M stock in December 1999. Discuss the validity of such a procedure to estimate daily volatility.

5.6. The file `mmm9912-adur.txt` contains an adjusted intraday trading duration of 3M stock in December 1999. There are thirty-nine 10-minute time intervals in a trading day. Let d_i be the average of all log durations for the ith 10-minute interval across all trading days in December 1999. Define an adjusted duration as $t_j / \exp(d_i)$, where j is in the ith 10-minute interval. Note that more sophisticated methods can be used to adjust the diurnal pattern of trading duration. Here we simply use a local average.

(a) Is there a diurnal pattern in the adjusted duration series? Why?

(b) Build a duration model for the adjusted series using exponential innovations. Check the fitted model.

(c) Build a duration model for the adjusted series using Weibull innovations. Check the fitted model.

(d) Build a duration model for the adjusted series using generalized gamma innovations. Check the fitted model.

(e) Compare and comment on the three duration models built before.

5.7. To gain experience in analyzing high-frequency financial data, consider the trade data of GE stock from December 1 to December 5, 2003 in the file `taq-t-ge-dec5.txt`. The file has four major columns; day, time (hour, minute, second), price, and volume. Ignore all transactions outside normal trading hours (9:30 am to 4:00 pm Eastern time). Construct a time series of the number of trades in an intraday 5-minute time interval. Is there any diurnal pattern in the constructed series? You can simply compute the sample ACF of the series to answer this question. The number of trades is in the file `taq-ge-dec5-nt.txt`.

5.8. Again, consider the high-frequency data of GE stock from December 1 to December 5, 2003 and ignore the transactions outside normal trading hours. Construct an intraday 5-minute return series. Note that the price of the stock in a 5-minute interval (e.g., 9:30 and 9:35 am) is the last transaction price within the time interval. For simplicity, ignore overnight returns. Are there serial correlations in the 5-minute return series? Use 10 lags of the ACF and 5% level to perform the test. See file `taq-ge-dec5-5m.txt`.

5.9. Consider the same problem as in Exercise 5.8, but use 10-minute time intervals. See file `taq-ge-dec5-10m.txt`.

5.10. Again, consider the high-frequency data of GE stock and ignore transactions outside normal trading hours. Compute the percentage of consecutive transactions without price change in the sample.

REFERENCES

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B* **26**: 211–243.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

Cho, D., Russell, J. R., Tiao, G. C., and Tsay, R. S. (2003). The magnet effect of price limits: Evidence from high frequency data on Taiwan stock exchange. *Journal of Empirical Finance* **10**: 133–168.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**: 1127–1162.

Ghysels, E. (2000). Some econometric recipes for high-frequency data cooking. *Journal of Business and Economic Statistics* **18**: 154–163.

Hasbrouck, J. (1992). *Using the TORQ Database*. Stern School of Business, New York University, New York.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *Journal of Finance* **54**: 2109–2142.

Hauseman, J., Lo, A., and MacKinlay, C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* **31**: 319–379.

Lo, A. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics* **45**: 181–212.

McCulloch, R. E. and Tsay, R. S. (2000). Nonlinearity in high frequency data and hierarchical models. *Studies in Nonlinear Dynamics and Econometrics* **5**: 1–17.

Roll, R. (1984). A simple implicit measure of the effective bid–ask spread in an efficient market. *Journal of Finance* **39**: 1127–1140.

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics* **1**: 2–25.

Stoll, H. and Whaley, R. (1990). Stock market structure and volatility. *Review of Financial Studies* **3**: 37–71.

Wood, R. A. (2000). Market microstructure research databases: History and projections. *Journal of Business & Economic Statistics* **18**: 140–145.

Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001a). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* **104**: 179–207.

Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001b). Determinants of bid and ask quotes and implications for the cost of trading. Working paper, Graduate School of Business, University of Chicago.

CHAPTER 6

Continuous-Time Models and Their Applications

The price of a financial asset evolves over time and forms a *stochastic process*, which is a statistical term used to describe the evolution of a random variable over time. The observed prices are a realization of the underlying stochastic process. The theory of stochastic process is the basis on which the observed prices are analyzed and statistical inference is made.

There are two types of stochastic process for modeling the price of an asset. The first type is called the *discrete-time stochastic process*, in which the price changes at discrete time points. All the processes discussed in the previous chapters belong to this category. For example, the daily closing price of IBM stock on the New York Stock Exchange forms a discrete-time stochastic process. Here the price changes only at the closing of a trading day. Price movements within a trading day are not necessarily relevant to the observed daily price. The second type of stochastic process is the *continuous-time process*, in which the price changes continuously, even though the price is only observed at discrete time points. One can think of the price as the "true value" of the stock that always exists and is time varying.

For both types of process, the price can be continuous or discrete. A continuous price can assume any positive real number, whereas a discrete price can only assume a countable number of possible values. Assume that the price of an asset is a continuous-time stochastic process. If the price is a continuous random variable, then we have a continuous-time continuous process. If the price itself is discrete, then we have a continuous-time discrete process. Similar classifications apply to discrete-time processes. The series of price change in Chapter 5 is an example of a discrete-time discrete process.

In this chapter, we treat the price of an asset as a continuous-time continuous stochastic process. Our goal is to introduce the statistical theory and tools needed to model financial assets and to price options. We begin the chapter with some terminologies of stock options used in the chapter. In Section 6.2, we provide a brief

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81**: 158–171.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk*. McGraw-Hill, Chicago.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**: 33–50.

Koenker, R. W. and D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics* **36**: 383–393.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

Longerstaey, J. and More, L. (1995). *Introduction to RiskMetrics*™, 4th edition. Morgan Guaranty Trust Company, New York.

Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business* **69**: 383–408.

Longin, F. M. (1999a). Optimal margin level in futures markets: Extreme price movements. *The Journal of Futures Markets* **19**: 127–152.

Longin, F. M. (1999b). From value at risk to stress testing: the extreme value approach. Working paper, Centre for Economic Policy Research, London, UK.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**: 119–131.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* **4**: 367–393.

Smith, R. L. (1999). Measuring risk with extreme value theory. Working paper, Department of Statistics, University of North Carolina at Chapel Hill.

Smith, R. L. and Shively, T. S. (1995). A point process approach to modeling trends in tropospheric ozone. *Atmospheric Environment* **29**: 3489–3499.

Tsay, R. S. (1999). Extreme value analysis of financial data. Working paper, Graduate School of Business, University of Chicago.

Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer-Verlag, New York.

CHAPTER 8

Multivariate Time Series Analysis and Its Applications

Economic globalization and internet communication have accelerated the integration of world financial markets in recent years. Price movements in one market can spread easily and instantly to another market. For this reason, financial markets are more dependent on each other than ever before, and one must consider them jointly to better understand the dynamic structure of global finance. One market may lead the other market under some circumstances, yet the relationship may be reversed under other circumstances. Consequently, knowing how the markets are interrelated is of great importance in finance. Similarly, for an investor or a financial institution holding multiple assets, the dynamic relationships between returns of the assets play an important role in decision making. In this and the next two chapters, we introduce econometric models and methods useful for studying jointly multiple return series. In the statistical literature, these models and methods belong to vector or multivariate time series analysis.

A multivariate time series consists of multiple single series referred to as *components*. As such, concepts of vector and matrix are important in multivariate time series analysis. We use boldface notation to indicate vectors and matrices. If necessary, readers may consult Appendix A of this chapter for some basic operations and properties of vectors and matrices. Appendix B provides some results of multivariate normal distribution, which is widely used in multivariate statistical analysis (e.g., Johnson and Wichern, 1998).

Let $\boldsymbol{r}_t = (r_{1t}, r_{2t}, \ldots, r_{kt})'$ be the log returns of k assets at time t, where $\boldsymbol{a}'$ denotes the transpose of $\boldsymbol{a}$. For example, an investor holding stocks of IBM, Microsoft, Exxon Mobil, General Motors, and Wal-Mart Stores may consider the five-dimensional daily log returns of these companies. Here r_{1t} denotes the daily log return of IBM stock, r_{2t} is that of Microsoft, and so on. As a second example, an investor who is interested in global investment may consider the return series of the S&P 500 index of the United States, the FTSE 100 index of the United

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

Kingdom, and the Nikkei 225 index of Japan. Here the series is three-dimensional, with r_{1t} denoting the return of the S&P 500 index, r_{2t} the return of the FTSE 100 index, and r_{3t} the return of the Nikkei 225. The goal of this chapter is to study econometric models for analyzing the multivariate process $\boldsymbol{r}_t$.

Many of the models and methods discussed in previous chapters can be generalized directly to the multivariate case. But there are situations in which the generalization requires some attention. In some situations, one needs new models and methods to handle the complicated relationships between multiple series. In this chapter, we discuss these issues with emphasis on intuition and applications. For statistical theory of multivariate time series analysis, readers are referred to Lütkepohl (1991) and Reinsel (1993).

8.1 WEAK STATIONARITY AND CROSS-CORRELATION MATRICES

Consider a k-dimensional time series $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{kt})'$. The series $\boldsymbol{r}_t$ is *weakly stationary* if its first and second moments are time-invariant. In particular, the mean vector and covariance matrix of a weakly stationary series are constant over time. Unless stated explicitly to the contrary, we assume that the return series of financial assets are weakly stationary.

For a weakly stationary time series $\boldsymbol{r}_t$, we define its mean vector and covariance matrix as

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t), \quad \boldsymbol{\Gamma}_0 = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_t - \boldsymbol{\mu})'], \tag{8.1}$$

where the expectation is taken element by element over the joint distribution of $\boldsymbol{r}_t$. The mean $\boldsymbol{\mu}$ is a k-dimensional vector consisting of the unconditional expectations of the components of $\boldsymbol{r}_t$. The covariance matrix $\boldsymbol{\Gamma}_0$ is a $k \times k$ matrix. The ith diagonal element of $\boldsymbol{\Gamma}_0$ is the variance of r_{it}, whereas the (i, j)th element of $\boldsymbol{\Gamma}_0$ is the covariance between r_{it} and r_{jt}. We write $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\Gamma}_0 = [\Gamma_{ij}(0)]$ when the elements are needed.

8.1.1 Cross-Correlation Matrices

Let $\boldsymbol{D}$ be a $k \times k$ diagonal matrix consisting of the standard deviations of r_{it} for $i = 1, \ldots, k$. In other words, $\boldsymbol{D} = \text{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{kk}(0)}\}$. The concurrent, or lag-zero, cross-correlation matrix of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\rho}_0 \equiv [\rho_{ij}(0)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1}.$$

More specifically, the (i, j)th element of $\boldsymbol{\rho}_0$ is

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(r_{it}, r_{jt})}{\text{std}(r_{it})\text{std}(r_{jt})},$$

which is the correlation coefficient between r_{it} and r_{jt}. In time series analysis, such a correlation coefficient is referred to as a concurrent, or contemporaneous, correlation coefficient because it is the correlation of the two series at time t. It is easy to see that $\rho_{ij}(0) = \rho_{ji}(0)$, $-1 \le \rho_{ij}(0) \le 1$, and $\rho_{ii}(0) = 1$ for $1 \le i, j \le k$. Thus, $\boldsymbol{\rho}(0)$ is a symmetric matrix with unit diagonal elements.

An important topic in multivariate time series analysis is the lead–lag relationships between component series. To this end, the cross-correlation matrices are used to measure the strength of linear dependence between time series. The lag-ℓ cross-covariance matrix of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\Gamma}_\ell \equiv [\Gamma_{ij}(\ell)] = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_{t-\ell} - \boldsymbol{\mu})'], \tag{8.2}$$

where $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{r}_t$. Therefore, the (i, j)th element of $\boldsymbol{\Gamma}_\ell$ is the covariance between r_{it} and $r_{j,t-\ell}$. For a weakly stationary series, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ is a function of ℓ, not the time index t.

The lag-ℓ cross-correlation matrix (CCM) of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\rho}_\ell \equiv [\rho_{ij}(\ell)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell\boldsymbol{D}^{-1}, \tag{8.3}$$

where, as before, $\boldsymbol{D}$ is the diagonal matrix of standard deviations of the individual series r_{it}. From the definition,

$$\rho_{ij}(\ell) = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(r_{it}, r_{j,t-\ell})}{\text{std}(r_{it})\text{std}(r_{jt})}, \tag{8.4}$$

which is the correlation coefficient between r_{it} and $r_{j,t-\ell}$. When $\ell > 0$, this correlation coefficient measures the linear dependence of r_{it} on $r_{j,t-\ell}$, which occurred prior to time t. Consequently, if $\rho_{ij}(\ell) \neq 0$ and $\ell > 0$, we say that the series r_{jt} *leads* the series r_{it} at lag ℓ. Similarly, $\rho_{ji}(\ell)$ measures the linear dependence of r_{jt} and $r_{i,t-\ell}$, and we say that the series r_{it} *leads* the series r_{jt} at lag ℓ if $\rho_{ji}(\ell) \neq 0$ and $\ell > 0$. Equation (8.4) also shows that the diagonal element $\rho_{ii}(\ell)$ is simply the lag-ℓ autocorrelation coefficient of r_{it}.

Based on this discussion, we obtain some important properties of the cross-correlations when $\ell > 0$. First, in general, $\rho_{ij}(\ell) \neq \rho_{ji}(\ell)$ for $i \neq j$ because the two correlation coefficients measure different linear relationships between $\{r_{it}\}$ and $\{r_{jt}\}$. Therefore, $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$ are in general not symmetric. Second, using $\text{Cov}(x, y) = \text{Cov}(y, x)$ and the weak stationarity assumption, we have

$$\text{Cov}(r_{it}, r_{j,t-\ell}) = \text{Cov}(r_{j,t-\ell}, r_{it}) = \text{Cov}(r_{jt}, r_{i,t+\ell}) = \text{Cov}(r_{jt}, r_{i,t-(-\ell)}),$$

so that $\Gamma_{ij}(\ell) = \Gamma_{ji}(-\ell)$. Because $\Gamma_{ji}(-\ell)$ is the (j, i)th element of the matrix $\boldsymbol{\Gamma}_{-\ell}$ and the equality holds for $1 \le i, j \le k$, we have $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{-\ell}$ and $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$. Consequently, unlike the univariate case, $\boldsymbol{\rho}_\ell \neq \boldsymbol{\rho}_{-\ell}$ for a general vector time series when $\ell > 0$. Because $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$, it suffices in practice to consider the cross-correlation matrices $\boldsymbol{\rho}_\ell$ for $\ell \ge 0$.

8.1.2 Linear Dependence

Considered jointly, the cross-correlation matrices $\{\boldsymbol{\rho}_\ell | \ell = 0, 1, \ldots\}$ of a weakly stationary vector time series contain the following information:

1. The diagonal elements $\{\rho_{ii}(\ell) | \ell = 0, 1, \ldots\}$ are the autocorrelation function of r_{it}.

2. The off-diagonal element $\rho_{ij}(0)$ measures the concurrent linear relationship between r_{it} and r_{jt}.
3. For $\ell > 0$, the off-diagonal element $\rho_{ij}(\ell)$ measures the linear dependence of r_{it} on the past value $r_{j,t-\ell}$.

Therefore, if $\rho_{ij}(\ell) = 0$ for all $\ell > 0$, then r_{it} does not depend linearly on any past value $r_{j,t-\ell}$ of the r_{jt} series.

In general, the linear relationship between two time series $\{r_{it}\}$ and $\{r_{jt}\}$ can be summarized as follows:

1. r_{it} and r_{jt} have no linear relationship if $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ for all $\ell \geq 0$.
2. r_{it} and r_{jt} are concurrently correlated if $\rho_{ij}(0) \neq 0$.
3. r_{it} and r_{jt} have no lead–lag relationship if $\rho_{ij}(\ell) = 0$ and $\rho_{ji}(\ell) = 0$ for all $\ell > 0$. In this case, we say the two series are uncoupled.
4. There is a *unidirectional relationship* from r_{it} to r_{jt} if $\rho_{ij}(\ell) = 0$ for all $\ell > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$. In this case, r_{it} does not depend on any past value of r_{jt}, but r_{jt} depends on some past values of r_{it}.
5. There is a *feedback relationship* between r_{it} and r_{jt} if $\rho_{ij}(\ell) \neq 0$ for some $\ell > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

The conditions stated earlier are sufficient conditions. A more informative approach to study the relationship between time series is to build a multivariate model for the series because a properly specified model considers simultaneously the serial and cross-correlations among the series.

8.1.3 Sample Cross-Correlation Matrices

Given the data $\{\boldsymbol{r}_t | t = 1, \ldots, T\}$, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ can be estimated by

$$\widehat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (\boldsymbol{r}_t - \overline{\boldsymbol{r}})(\boldsymbol{r}_{t-\ell} - \overline{\boldsymbol{r}})', \quad \ell \geq 0, \tag{8.5}$$

where $\overline{\boldsymbol{r}} = \left(\sum_{t=1}^{T} \boldsymbol{r}_t\right)/T$ is the vector of sample means. The cross-correlation matrix $\boldsymbol{\rho}_\ell$ is estimated by

$$\widehat{\boldsymbol{\rho}}_\ell = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{D}}^{-1}, \quad \ell \geq 0, \tag{8.6}$$

where $\widehat{\boldsymbol{D}}$ is the $k \times k$ diagonal matrix of the sample standard deviations of the component series.

Similar to the univariate case, asymptotic properties of the sample cross-correlation matrix $\widehat{\boldsymbol{\rho}}_\ell$ have been investigated under various assumptions; see, for instance, Fuller (1976, Chapter 6). The estimate is consistent but is biased in a finite sample. For asset return series, the finite sample distribution of $\widehat{\boldsymbol{\rho}}_\ell$ is

rather complicated partly because of the presence of conditional heteroscedasticity and high kurtosis. If the finite-sample distribution of cross-correlations is needed, we recommend that proper bootstrap resampling methods be used to obtain an approximate estimate of the distribution. For many applications, a crude approximation of the variance of $\hat{\rho}_{ij}(\ell)$ is sufficient.

Example 8.1. Consider the monthly log returns of IBM stock and the S&P 500 index from January 1926 to December 1999 with 888 observations. The returns include dividend payments and are in percentages. Denote the returns of IBM stock and the S&P 500 index by r_{1t} and r_{2t}, respectively. These two returns form a bivariate time series $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$. Figure 8.1 shows the time plots of $\boldsymbol{r}_t$ using the same scale. Figure 8.2 shows some scatterplots of the two series. The plots show that the two return series are concurrently correlated. Indeed, the sample concurrent correlation coefficient between the two returns is 0.64, which is statistically significant at the 5% level. However, the cross-correlations at lag 1 are weak if any.

Table 8.1 provides some summary statistics and cross-correlation matrices of the two series. For a bivariate series, each CCM is a 2×2 matrix with four correlations. Empirical experience indicates that it is rather hard to absorb simultaneously many cross-correlation matrices, especially when the dimension k is greater than 3. To overcome this difficulty, we use the simplifying notation of Tiao and Box (1981)

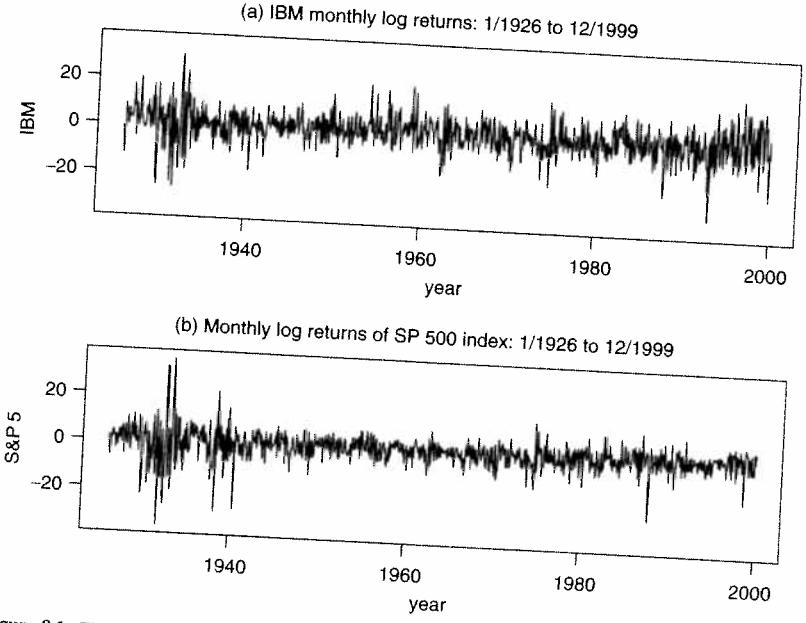

Figure 8.1. Time plots of (a) monthly log returns in percentages for IBM stock and (b) the S&P 500 index from January 1926 to December 1999.

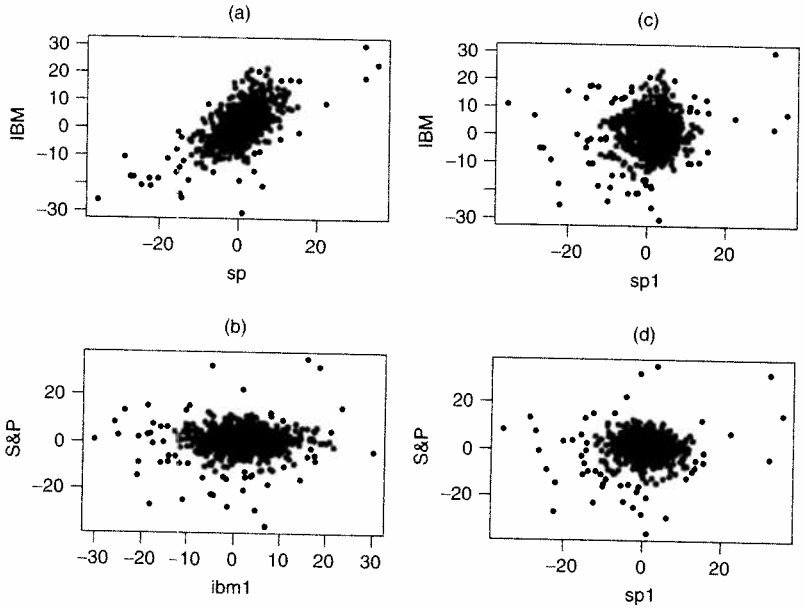

Figure 8.2. Some scatterplots for monthly log returns of IBM stock and the S&P 500 index: (a) concurrent plot of IBM versus S&P 500 (b) S&P 500 versus lag-1 IBM, (c) IBM versus lag-1 S&P 500, and (d) S&P 500 versus lag-1 S&P 500.

and define a simplified cross-correlation matrix consisting of three symbols "+," "−," and "." where

1. "+" means that the corresponding correlation coefficient is greater than or equal to $2/\sqrt{T}$,
2. "−" means that the corresponding correlation coefficient is less than or equal to $-2/\sqrt{T}$, and
3. "." means that the corresponding correlation coefficient is between $-2/\sqrt{T}$ and $2/\sqrt{T}$,

where $1/\sqrt{T}$ is the asymptotic 5% critical value of the sample correlation under the assumption that $\boldsymbol{r}_t$ is a white noise series.

Table 8.1c shows the simplified CCM for the monthly log returns of IBM stock and the S&P 500 index. It is easily seen that significant cross-correlations at the approximate 5% level appear mainly at lags 1 and 3. An examination of the sample CCMs at these two lags indicates that (a) S&P 500 index returns have some marginal autocorrelations at lags 1 and 3, and (b) IBM stock returns depend weakly on the previous returns of the S&P 500 index. The latter observation is based on the significance of cross-correlations at the (1, 2)th element of lag-1 and lag-3 CCMs.

Table 8.1. Summary Statistics and Cross-Correlation Matrices of Monthly Log Returns of IBM Stock and the S&P 500 Index: January 1926 to December 1999

(a) Summary Statistics

Ticker	Mean	Standard Error	Skewness	Excess Kurtosis	Minimum	Maximum
IBM	1.240	6.729	−0.237	1.917	−30.37	30.10
S&P 500	0.537	5.645	−0.521	8.117	−35.58	35.22

(b) Cross-Correlation Matrices

Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
0.08	0.10	0.02	−0.06	−0.02	−0.07	−0.02	−0.03	0.00	0.07
0.04	0.08	0.02	−0.02	−0.07	−0.11	0.04	0.02	0.00	0.08

(c) Simplified Notation

$$\begin{bmatrix} + & + \\ \cdot & + \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & - \\ - & - \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & + \end{bmatrix}$$

Figure 8.3 shows the sample autocorrelations and cross-correlations of the two series. Since the ACF is symmetric with respect to lag 0, only those of positive lags are shown. Because lagged values of the S&P 500 index return are used to compute the cross-correlations, the plot associated with positive lags in Figure 8.3c shows the dependence of IBM stock return on the past S&P 500 index returns, and the plot associated with negative lags shows the linear dependence of the index return on the past IBM stock returns. The horizontal lines in the plots are the asymptotic two standard-error limits of the sample auto- and cross-correlation coefficients. From the plots, the dynamic relationship is weak between the two return series, but their contemporaneous correlation is statistically significant.

Example 8.2. Consider the simple returns of monthly indexes of U.S. government bonds with maturities in 30 years, 20 years, 10 years, 5 years, and 1 year. The data obtained from the CRSP database have 696 observations starting from January 1942 to December 1999. Let $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{5t})'$ be the return series with decreasing time to maturity. Figure 8.4 shows the time plots of $\boldsymbol{r}_t$ on the same scale. The variability of the 1-year bond returns is much smaller than that of returns with longer maturities. The sample means and standard deviations of the data are $\widehat{\boldsymbol{\mu}} = 10^{-2}(0.43, 0.45, 0.45, 0.46, 0.44)'$ and $\widehat{\boldsymbol{\sigma}} = 10^{-2}(2.53, 2.43, 1.97, 1.39, 0.53)'$. The concurrent correlation matrix of the series is

$$\widehat{\boldsymbol{\rho}}_0 = \begin{bmatrix} 1.00 & 0.98 & 0.92 & 0.85 & 0.63 \\ 0.98 & 1.00 & 0.91 & 0.86 & 0.64 \\ 0.92 & 0.91 & 1.00 & 0.90 & 0.68 \\ 0.85 & 0.86 & 0.90 & 1.00 & 0.82 \\ 0.63 & 0.64 & 0.68 & 0.82 & 1.00 \end{bmatrix}.$$

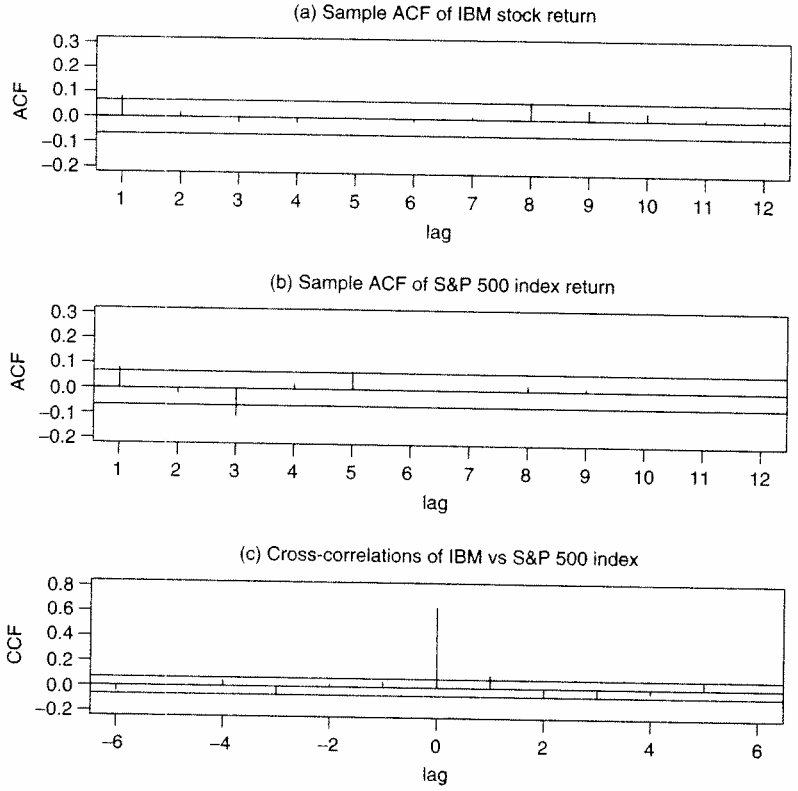

Figure 8.3. Sample auto- and cross-correlation functions of two monthly log returns: (a) sample ACF of IBM stock returns, (b) sample ACF of S&P 500 index returns, and (c) cross-correlations between IBM stock return and lagged S&P 500 index returns.

It is not surprising that (a) the series have high concurrent correlations, and (b) the correlations between long-term bonds are higher than those between short-term bonds.

Table 8.2 gives the lag-1 and lag-2 cross-correlation matrices of $\boldsymbol{r}_t$ and the corresponding simplified matrices. Most of the significant cross-correlations are at lag 1, and the five return series appear to be intercorrelated. In addition, lag-1 and lag-2 sample ACFs of the 1-year bond returns are substantially higher than those of other series with longer maturities.

8.1.4 Multivariate Portmanteau Tests

The univariate Ljung–Box statistic $Q(m)$ has been generalized to the multivariate case by Hosking (1980, 1981) and Li and McLeod (1981). For a multivariate series, the null hypothesis of the test statistic is $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}$, and the

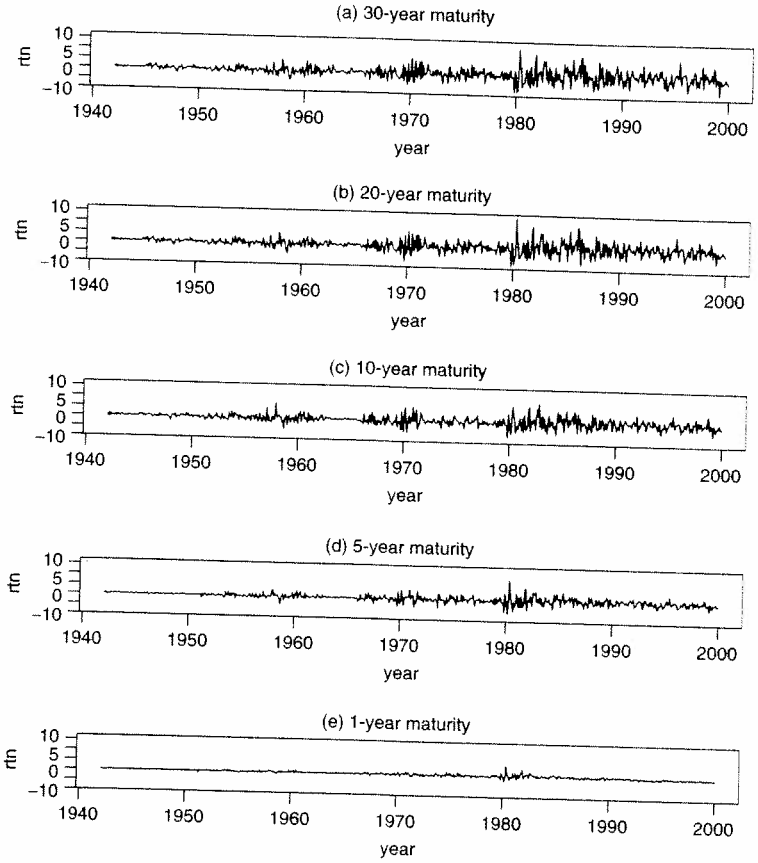

Figure 8.4. Time plots of monthly simple returns of five indexes of U.S. government bonds with maturities in (a) 30 years, (b) 20 years, (c) 10 years, (d) 5 years, and (e) 1 year. The sample period is from January 1942 to December 1999.

alternative hypothesis $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, \ldots, m\}$. Thus, the statistic is used to test that there are no auto- and cross-correlations in the vector series $\boldsymbol{r}_t$. The test statistic assumes the form

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr(\widehat{\boldsymbol{\Gamma}}_\ell' \widehat{\boldsymbol{\Gamma}}_0^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{\Gamma}}_0^{-1}), \tag{8.7}$$

where T is the sample size, k is the dimension of $\boldsymbol{r}_t$, and $tr(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$, which is the sum of the diagonal elements of $\boldsymbol{A}$. Under the null hypothesis and some regularity conditions, $Q_k(m)$ follows asymptotically a chi-squared distribution with $k^2 m$ degrees of freedom.

Table 8.2. Sample Cross-Correlation Matrices of Monthly Simple Returns of Five Indexes of U.S. Government Bonds: January 1942 to December 1999

Lag 1					Lag 2				
Cross-Correlations									
0.10	0.08	0.11	0.12	0.16	−0.01	0.00	0.00	−0.03	0.03
0.10	0.08	0.12	0.14	0.17	−0.01	0.00	0.00	−0.04	0.02
0.09	0.08	0.09	0.13	0.18	0.01	0.01	0.01	−0.02	0.07
0.14	0.12	0.15	0.14	0.22	−0.02	−0.01	0.00	−0.04	0.07
0.17	0.15	0.21	0.22	0.40	−0.02	0.00	0.02	0.02	0.22
Simplified Cross-Correlation Matrices									
$\begin{bmatrix} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{bmatrix}$					$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + \end{bmatrix}$				

Remark. The $Q_k(m)$ statistics can be rewritten in terms of the sample cross-correlation matrixes $\widehat{\boldsymbol{\rho}}_\ell$. Using the Kronecker product $\otimes$ and vectorization of matrices discussed in Appendix A of this chapter, we have

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \boldsymbol{b}_\ell'(\widehat{\boldsymbol{\rho}}_0^{-1} \otimes \widehat{\boldsymbol{\rho}}_0^{-1})\boldsymbol{b}_\ell,$$

where $\boldsymbol{b}_\ell = \text{vec}(\widehat{\boldsymbol{\rho}}_\ell')$. The test statistic proposed by Li and McLeod (1981) is

$$Q_k^*(m) = T \sum_{\ell=1}^{m} \boldsymbol{b}_\ell'(\widehat{\boldsymbol{\rho}}_0^{-1} \otimes \widehat{\boldsymbol{\rho}}_0^{-1})\boldsymbol{b}_\ell + \frac{k^2 m(m+1)}{2T},$$

which is asymptotically equivalent to $Q_k(m)$. □

Applying the $Q_k(m)$ statistics to the bivariate monthly log returns of IBM stock and the S&P 500 index of Example 8.1, we have $Q_2(1) = 9.81$, $Q_2(5) = 47.06$, and $Q_2(10) = 71.65$. Based on asymptotic chi-squared distributions with degrees of freedom 4, 20, and 40, the p-values of these $Q_2(m)$ statistics are 0.044, 0.001, and 0.002, respectively. The portmanteau tests thus confirm the existence of serial dependence in the bivariate return series at the 5% significance level. For the five-dimensional monthly simple returns of bond indexes in Example 8.2, we have $Q_5(5) = 1065.63$, which is highly significant compared with a chi-squared distribution with 125 degrees of freedom.

The $Q_k(m)$ statistic is a joint test for checking the first m cross-correlation matrices of $\boldsymbol{r}_t$. If it rejects the null hypothesis, then we build a multivariate model for the series to study the lead–lag relationships between the component series. In what follows, we discuss some simple vector models useful for modeling the linear dynamic structure of a multivariate financial time series.

8.2 VECTOR AUTOREGRESSIVE MODELS

A simple vector model useful in modeling asset returns is the vector autoregressive (VAR) model. A multivariate time series $\boldsymbol{r}_t$ is a VAR process of order 1, or VAR(1) for short, if it follows the model

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t, \tag{8.8}$$

where $\boldsymbol{\phi}_0$ is a k-dimensional vector, $\boldsymbol{\Phi}$ is a $k \times k$ matrix, and $\{\boldsymbol{a}_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. In application, the covariance matrix $\boldsymbol{\Sigma}$ is required to be positive definite; otherwise, the dimension of $\boldsymbol{r}_t$ can be reduced. In the literature, it is often assumed that $\boldsymbol{a}_t$ is multivariate normal.

Consider the bivariate case (i.e., $k = 2$, $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$, and $\boldsymbol{a}_t = (a_{1t}, a_{2t})'$). The VAR(1) model consists of the following two equations:

$$r_{1t} = \phi_{10} + \Phi_{11} r_{1,t-1} + \Phi_{12} r_{2,t-1} + a_{1t},$$
$$r_{2t} = \phi_{20} + \Phi_{21} r_{1,t-1} + \Phi_{22} r_{2,t-1} + a_{2t},$$

where Φ_{ij} is the (i, j)th element of $\boldsymbol{\Phi}$ and ϕ_{i0} is the ith element of $\boldsymbol{\phi}_0$. Based on the first equation, Φ_{12} denotes the linear dependence of r_{1t} on $r_{2,t-1}$ in the presence of $r_{1,t-1}$. Therefore, Φ_{12} is the conditional effect of $r_{2,t-1}$ on r_{1t} given $r_{1,t-1}$. If $\Phi_{12} = 0$, then r_{1t} does not depend on $r_{2,t-1}$, and the model shows that r_{1t} only depends on its own past. Similarly, if $\Phi_{21} = 0$, then the second equation shows that r_{2t} does not depend on $r_{1,t-1}$ when $r_{2,t-1}$ is given.

Consider the two equations jointly. If $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$, then there is a unidirectional relationship from r_{1t} to r_{2t}. If $\Phi_{12} = \Phi_{21} = 0$, then r_{1t} and r_{2t} are uncoupled. If $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$, then there is a feedback relationship between the two series.

8.2.1 Reduced and Structural Forms

In general, the coefficient matrix $\boldsymbol{\Phi}$ of Eq. (8.8) measures the dynamic dependence of $\boldsymbol{r}_t$. The concurrent relationship between r_{1t} and r_{2t} is shown by the off-diagonal element σ_{12} of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$. If $\sigma_{12} = 0$, then there is no concurrent linear relationship between the two component series. In the econometric literature, the VAR(1) model in Eq. (8.8) is called a *reduced-form* model because it

does not show explicitly the concurrent dependence between the component series. If necessary, an explicit expression involving the concurrent relationship can be deduced from the reduced-form model by a simple linear transformation. Because $\boldsymbol{\Sigma}$ is positive definite, there exists a lower triangular matrix $\boldsymbol{L}$ with unit diagonal elements and a diagonal matrix $\boldsymbol{G}$ such that $\boldsymbol{\Sigma} = \boldsymbol{LGL}'$; see Appendix A on Cholesky decomposition. Therefore, $\boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{G}$.

Define $\boldsymbol{b}_t = (b_{1t}, \ldots, b_{kt})' = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then

$$E(\boldsymbol{b}_t) = \boldsymbol{L}^{-1}E(\boldsymbol{a}_t) = \boldsymbol{0}, \quad \text{Cov}(\boldsymbol{b}_t) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^{-1})' = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{G}.$$

Since $\boldsymbol{G}$ is a diagonal matrix, the components of $\boldsymbol{b}_t$ are uncorrelated. Multiplying $\boldsymbol{L}^{-1}$ from the left to model (8.8), we obtain

$$\boldsymbol{L}^{-1}\boldsymbol{r}_t = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0 + \boldsymbol{L}^{-1}\boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{L}^{-1}\boldsymbol{a}_t = \boldsymbol{\phi}_0^* + \boldsymbol{\Phi}^*\boldsymbol{r}_{t-1} + \boldsymbol{b}_t, \tag{8.9}$$

where $\boldsymbol{\phi}_0^* = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0$ is a k-dimensional vector and $\boldsymbol{\Phi}^* = \boldsymbol{L}^{-1}\boldsymbol{\Phi}$ is a $k \times k$ matrix. Because of the special matrix structure, the kth row of $\boldsymbol{L}^{-1}$ is in the form $(w_{k1}, w_{k2}, \ldots, w_{k,k-1}, 1)$. Consequently, the kth equation of model (8.9) is

$$r_{kt} + \sum_{i=1}^{k-1} w_{ki}r_{it} = \phi_{k,0}^* + \sum_{i=1}^{k} \Phi_{ki}^* r_{i,t-1} + b_{kt}, \tag{8.10}$$

where $\phi_{k,0}^*$ is the kth element of $\boldsymbol{\phi}_0^*$ and Φ_{ki}^* is the (k, i)th element of $\boldsymbol{\Phi}^*$. Because b_{kt} is uncorrelated with b_{it} for $1 \le i < k$, Eq. (8.10) shows explicitly the concurrent linear dependence of r_{kt} on r_{it}, where $1 \le i \le k-1$. This equation is referred to as a *structural equation* for r_{kt} in the econometric literature.

For any other component r_{it} of $\boldsymbol{r}_t$, we can rearrange the VAR(1) model so that r_{it} becomes the last component of $\boldsymbol{r}_t$. The prior transformation method can then be applied to obtain a structural equation for r_{it}. Therefore, the reduced-form model (8.8) is equivalent to the structural form used in the econometric literature. In time series analysis, the reduced-form model is commonly used for two reasons. The first reason is ease in estimation. The second and main reason is that the concurrent correlations cannot be used in forecasting.

Example 8.3. To illustrate the transformation from a reduced-form model to structural equations, consider the bivariate AR(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

For this particular covariance matrix $\boldsymbol{\Sigma}$, the lower triangular matrix

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix}$$

provides a Cholesky decomposition (i.e., $\boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1}$ is a diagonal matrix). Premultiplying $\boldsymbol{L}^{-1}$ to the previous bivariate AR(1) model, we obtain

$$\begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.7 & 0.95 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix},$$

$$\boldsymbol{G} = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix},$$

where $\boldsymbol{G} = \text{Cov}(\boldsymbol{b}_t)$. The second equation of this transformed model gives

$$r_{2t} = 0.3 + 0.5r_{1t} - 0.7r_{1,t-1} + 0.95r_{2,t-1} + b_{2t},$$

which shows explicitly the linear dependence of r_{2t} on r_{1t}.

Rearranging the order of elements in $\boldsymbol{r}_t$, the bivariate AR(1) model becomes

$$\begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} a_{2t} \\ a_{1t} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

The lower triangular matrix needed in the Cholesky decomposition of $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix}.$$

Premultiplying $\boldsymbol{L}^{-1}$ to the earlier rearranged VAR(1) model, we obtain

$$\begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix} \begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ -0.8 & 0.8 \end{bmatrix} \begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix},$$

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where $\boldsymbol{G} = \text{Cov}(\boldsymbol{c}_t)$. The second equation now gives

$$r_{1t} = -0.2 + 1.0r_{2t} - 0.8r_{2,t-1} + 0.8r_{1,t-1} + c_{2t}.$$

Again this equation shows explicitly the concurrent linear dependence of r_{1t} on r_{2t}.

8.2.2 Stationarity Condition and Moments of a VAR(1) Model

Assume that the VAR(1) model in Eq. (8.8) is weakly stationary. Taking expectation of the model and using $E(\boldsymbol{a}_t) = \boldsymbol{0}$, we obtain

$$E(\boldsymbol{r}_t) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}E(\boldsymbol{r}_{t-1}).$$

Since $E(\boldsymbol{r}_t)$ is time-invariant, we have

$$\boldsymbol{\mu} \equiv E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_0$$

provided that the matrix $\boldsymbol{I} - \boldsymbol{\Phi}$ is nonsingular, where $\boldsymbol{I}$ is the $k \times k$ identity matrix.

Using $\boldsymbol{\phi}_0 = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}$, the VAR(1) model in Eq. (8.8) can be written as

$$(\boldsymbol{r}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t.$$

Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\mu}$ be the mean-corrected time series. Then the VAR(1) model becomes

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t. \tag{8.11}$$

This model can be used to derive properties of a VAR(1) model. By repeated substitutions, we can rewrite Eq. (8.11) as

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \boldsymbol{\Phi}\boldsymbol{a}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{a}_{t-2} + \boldsymbol{\Phi}^3\boldsymbol{a}_{t-3} + \cdots.$$

This expression shows several characteristics of a VAR(1) process. First, since $\boldsymbol{a}_t$ is serially uncorrelated, it follows that $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-1}) = \boldsymbol{0}$. In fact, $\boldsymbol{a}_t$ is not correlated with $\boldsymbol{r}_{t-\ell}$ for all $\ell > 0$. For this reason, $\boldsymbol{a}_t$ is referred to as the *shock* or *innovation* of the series at time t. It turns out that, similar to the univariate case, $\boldsymbol{a}_t$ is uncorrelated with the past value $\boldsymbol{r}_{t-j}$ $(j > 0)$ for all time series models. Second, postmultiplying the expression by $\boldsymbol{a}_t'$, taking expectation, and using the fact of no serial correlations in the $\boldsymbol{a}_t$ process, we obtain $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$. Third, for a VAR(1) model, $\boldsymbol{r}_t$ depends on the past innovation $\boldsymbol{a}_{t-j}$ with coefficient matrix $\boldsymbol{\Phi}^j$. For such dependence to be meaningful, $\boldsymbol{\Phi}^j$ must converge to zero as $j \to \infty$. This means that the k eigenvalues of $\boldsymbol{\Phi}$ must be less than 1 in modulus; otherwise, $\boldsymbol{\Phi}^j$ will either explode or converge to a nonzero matrix as $j \to \infty$. As a matter of fact, the requirement that all eigenvalues of $\boldsymbol{\Phi}$ are less than 1 in modulus is the necessary and sufficient condition for weak stationarity of $\boldsymbol{r}_t$ provided that the covariance matrix of $\boldsymbol{a}_t$ exists. Notice that this stationarity condition reduces to that of the univariate AR(1) case in which the condition is $|\phi| < 1$. Furthermore, because

$$|\boldsymbol{\Phi} - \lambda\boldsymbol{I}| = |\lambda\boldsymbol{I} - \boldsymbol{\Phi}| = \lambda^k \left|\boldsymbol{I} - \boldsymbol{\Phi}\frac{1}{\lambda}\right|,$$

the eigenvalues of $\boldsymbol{\Phi}$ are the inverses of the zeros of the determinant $|\boldsymbol{I} - \boldsymbol{\Phi}B|$. Thus, an equivalent sufficient and necessary condition for stationarity of $\boldsymbol{r}_t$ is that all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ are greater than one in modulus; that is, all zeros are outside the unit circle in the complex plane. Fourth, using the expression, we have

$$\text{Cov}(\boldsymbol{r}_t) = \boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}' + \boldsymbol{\Phi}^2\boldsymbol{\Sigma}(\boldsymbol{\Phi}^2)' + \cdots = \sum_{i=0}^{\infty} \boldsymbol{\Phi}^i\boldsymbol{\Sigma}(\boldsymbol{\Phi}^i)',$$

where it is understood that $\boldsymbol{\Phi}^0 = \boldsymbol{I}$, the $k \times k$ identity matrix.

Postmultiplying $\tilde{\boldsymbol{r}}_{t-\ell}'$ to Eq. (8.11), taking expectation, and using the result $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-j}) = E(\boldsymbol{a}_t\tilde{\boldsymbol{r}}_{t-j}') = \boldsymbol{0}$ for $j > 0$, we obtain

$$E(\tilde{\boldsymbol{r}}_t\tilde{\boldsymbol{r}}_{t-\ell}') = \boldsymbol{\Phi}E(\tilde{\boldsymbol{r}}_{t-1}\tilde{\boldsymbol{r}}_{t-\ell})', \quad \ell > 0.$$

Therefore,

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0, \tag{8.12}$$

where $\boldsymbol{\Gamma}_j$ is the lag-j cross-covariance matrix of $\boldsymbol{r}_t$. Again this result is a generalization of that of a univariate AR(1) process. By repeated substitutions, Eq. (8.12) shows that

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}^\ell\boldsymbol{\Gamma}_0, \quad \text{for} \quad \ell > 0.$$

Pre- and postmultiplying Eq. (8.12) by $\boldsymbol{D}^{-1/2}$, we obtain

$$\boldsymbol{\rho}_\ell = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{D}^{1/2}\boldsymbol{D}^{-1/2}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1/2} = \boldsymbol{\Upsilon}\boldsymbol{\rho}_{\ell-1},$$

where $\boldsymbol{\Upsilon} = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{D}^{1/2}$. Consequently, the CCM of a VAR(1) model satisfies

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}^\ell\boldsymbol{\rho}_0, \quad \text{for} \quad \ell > 0.$$

8.2.3 Vector AR(p) Models

The generalization of VAR(1) to VAR(p) models is straightforward. The time series $\boldsymbol{r}_t$ follows a VAR(p) model if it satisfies

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \quad p > 0, \tag{8.13}$$

where $\boldsymbol{\phi}_0$ and $\boldsymbol{a}_t$ are defined as before, and $\boldsymbol{\Phi}_j$ are $k \times k$ matrices. Using the back-shift operator B, the VAR(p) model can be written as

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{I}$ is the $k \times k$ identity matrix. This representation can be written in a compact form as

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ is a matrix polynomial. If $\boldsymbol{r}_t$ is weakly stationary, then we have

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$$

provided that the inverse exists. Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\mu}$. The VAR($p$) model becomes

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}_1\tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Phi}_p\tilde{\boldsymbol{r}}_{t-p} + \boldsymbol{a}_t. \tag{8.14}$$

Using this equation and the same techniques as those for VAR(1) models, we obtain that:

- $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{a}_t$;
- $\text{Cov}(\boldsymbol{r}_{t-\ell}, \boldsymbol{a}_t) = \boldsymbol{0}$ for $\ell > 0$;
- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_1\boldsymbol{\Gamma}_{\ell-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}_{\ell-p}$ for $\ell > 0$.

The last property is called the moment equations of a VAR(p) model. It is a multivariate version of the Yule–Walker equation of a univariate AR(p) model. In terms of CCM, the moment equations become

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}_1 \boldsymbol{\rho}_{\ell-1} + \cdots + \boldsymbol{\Upsilon}_p \boldsymbol{\rho}_{\ell-p} \quad \text{for} \quad \ell > 0,$$

where $\boldsymbol{\Upsilon}_i = \boldsymbol{D}^{-1/2} \boldsymbol{\Phi}_i \boldsymbol{D}^{1/2}$.

A simple approach to understanding properties of the VAR(p) model in Eq. (8.13) is to make use of the results of the VAR(1) model in Eq. (8.8). This can be achieved by transforming the VAR(p) model of $\boldsymbol{r}_t$ into a kp-dimensional VAR(1) model. Specifically, let $\boldsymbol{x}_t = (\tilde{\boldsymbol{r}}'_{t-p+1}, \tilde{\boldsymbol{r}}'_{t-p+2}, \ldots, \tilde{\boldsymbol{r}}'_t)'$ and $\boldsymbol{b}_t = (0, \ldots, 0, \boldsymbol{a}'_t)'$ be two kp-dimensional processes. The mean of $\boldsymbol{b}_t$ is zero and the covariance matrix of $\boldsymbol{b}_t$ is a $kp \times kp$ matrix with zero everywhere except for the lower right corner, which is $\boldsymbol{\Sigma}$. The VAR(p) model for $\boldsymbol{r}_t$ can then be written in the form

$$\boldsymbol{x}_t = \boldsymbol{\Phi}^* \boldsymbol{x}_{t-1} + \boldsymbol{b}_t, \tag{8.15}$$

where $\boldsymbol{\Phi}^*$ is a $kp \times kp$ matrix given by

$$\boldsymbol{\Phi}^* = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} \\ \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_{p-2} & \boldsymbol{\Phi}_{p-3} & \cdots & \boldsymbol{\Phi}_1 \end{bmatrix},$$

where $\boldsymbol{0}$ and $\boldsymbol{I}$ are the $k \times k$ zero matrix and identity matrix, respectively. In the literature, $\boldsymbol{\Phi}^*$ is called the *companion* matrix of the matrix polynomial $\boldsymbol{\Phi}(B)$.

Equation (8.15) is a VAR(1) model for $\boldsymbol{x}_t$, which contains $\boldsymbol{r}_t$ as its last k components. The results of a VAR(1) model shown in the previous subsection can now be used to derive properties of the VAR(p) model via Eq. (8.15). For example, from the definition, $\boldsymbol{x}_t$ is weakly stationary if and only if $\boldsymbol{r}_t$ is weakly stationary. Therefore, the necessary and sufficient condition of weak stationarity for the VAR(p) model in Eq. (8.13) is that all eigenvalues of $\boldsymbol{\Phi}^*$ in Eq. (8.15) are less than 1 in modulus. Similar to the VAR(1) case, it can be shown that the condition is equivalent to all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ being outside the unit circle.

Of particular relevance to financial time series analysis is the structure of the coefficient matrices $\boldsymbol{\Phi}_\ell$ of a VAR(p) model. For instance, if the (i, j)th element $\Phi_{ij}(\ell)$ of $\boldsymbol{\Phi}_\ell$ is zero for all ℓ, then r_{it} does not depend on the past values of r_{jt}. The structure of the coefficient matrices $\boldsymbol{\Phi}_\ell$ thus provides information on the lead–lag relationship between the components of $\boldsymbol{r}_t$.

8.2.4 Building a VAR(p) Model

We continue to use the iterative procedure of order specification, estimation, and model checking to build a vector AR model for a given time series. The concept of

partial autocorrelation function of a univariate series can be generalized to specify the order p of a vector series. Consider the following consecutive VAR models:

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{a}_t \\ \boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{r}_{t-2} + \boldsymbol{a}_t \\ &\vdots = \vdots \\ \boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_i \boldsymbol{r}_{t-i} + \boldsymbol{a}_t \\ &\vdots = \vdots \end{aligned} \tag{8.16}$$

Parameters of these models can be estimated by the ordinary least squares (OLS) method. This is called the multivariate linear regression estimation in multivariate statistical analysis; see Johnson and Wichern (1998).

For the ith equation in Eq. (8.16), let $\widehat{\boldsymbol{\Phi}}_j^{(i)}$ be the OLS estimate of $\boldsymbol{\Phi}_j$ and let $\widehat{\boldsymbol{\phi}}_0^{(i)}$ be the estimate of $\boldsymbol{\phi}_0$, where the superscript (i) is used to denote that the estimates are for a VAR(i) model. Then the residual is

$$\widehat{\boldsymbol{a}}_t^{(i)} = \boldsymbol{r}_t - \widehat{\boldsymbol{\phi}}_0^{(i)} - \widehat{\boldsymbol{\Phi}}_1^{(i)} \boldsymbol{r}_{t-1} - \cdots - \widehat{\boldsymbol{\Phi}}_i^{(i)} \boldsymbol{r}_{t-i}.$$

For $i = 0$, the residual is defined as $\widehat{\boldsymbol{r}}_t^{(0)} = \boldsymbol{r}_t - \bar{\boldsymbol{r}}$, where $\bar{\boldsymbol{r}}$ is the sample mean of $\boldsymbol{r}_t$. The residual covariance matrix is defined as

$$\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{T - 2i - 1} \sum_{t=i+1}^{T} \widehat{\boldsymbol{a}}_t^{(i)} (\widehat{\boldsymbol{a}}_t^{(i)})', \quad i \geq 0. \tag{8.17}$$

To specify the order p, one can test the hypothesis $H_o : \boldsymbol{\Phi}_\ell = \boldsymbol{0}$ versus the alternative hypothesis $H_a : \boldsymbol{\Phi}_\ell \neq \boldsymbol{0}$ sequentially for $\ell = 1, 2, \ldots$. For example, using the first equation in Eq. (8.16), we can test the hypothesis $H_o : \boldsymbol{\Phi}_1 = \boldsymbol{0}$ versus the alternative hypothesis $H_a : \boldsymbol{\Phi}_1 \neq \boldsymbol{0}$. The test statistic is

$$M(1) = -(T - k - \tfrac{5}{2}) \ln \left(\frac{|\widehat{\boldsymbol{\Sigma}}_1|}{|\widehat{\boldsymbol{\Sigma}}_0|} \right),$$

where $\widehat{\boldsymbol{\Sigma}}_i$ is defined in Eq. (8.17) and $|\boldsymbol{A}|$ denotes the determinant of the matrix $\boldsymbol{A}$. Under some regularity conditions, the test statistic $M(1)$ is asymptotically a chi-squared distribution with k^2 degrees of freedom; see Tiao and Box (1981).

In general, we use the ith and $(i-1)$th equations in Eq. (8.16) to test $H_o : \boldsymbol{\Phi}_i = \boldsymbol{0}$ versus $H_a : \boldsymbol{\Phi}_i \neq \boldsymbol{0}$; that is, testing a VAR(i) model versus a VAR($i-1$) model. The test statistic is

$$M(i) = -(T - k - i - \tfrac{3}{2}) \ln \left(\frac{|\widehat{\boldsymbol{\Sigma}}_i|}{|\widehat{\boldsymbol{\Sigma}}_{i-1}|} \right). \tag{8.18}$$

Asymptotically, $M(i)$ is distributed as a chi-squared distribution with k^2 degrees of freedom.

Alternatively, one can use the Akaike information criterion (AIC) or its variants to select the order p. Assume that $\boldsymbol{a}_t$ is multivariate normal and consider the ith equation in Eq. (8.16). One can estimate the model by the maximum likelihood (ML) method. For AR models, the OLS estimates $\hat{\boldsymbol{\phi}}_0$ and $\hat{\boldsymbol{\Phi}}_j$ are equivalent to the (conditional) ML estimates. However, there are differences between the estimates of $\boldsymbol{\Sigma}$. The ML estimate of $\boldsymbol{\Sigma}$ is

$$\tilde{\boldsymbol{\Sigma}}_i = \frac{1}{T}\sum_{t=i+1}^{T} \hat{\boldsymbol{a}}_t^{(i)}[\hat{\boldsymbol{a}}_t^{(i)}]'. \tag{8.19}$$

The AIC of a VAR(i) model under the normality assumption is defined as

$$\text{AIC}(i) = \ln(|\tilde{\boldsymbol{\Sigma}}_i|) + \frac{2k^2 i}{T}.$$

For a given vector time series, one selects the AR order p such that $\text{AIC}(p) = \min_{0\le i\le p_0} \text{AIC}(i)$, where p_0 is a prespecified positive integer.

Other information criteria available for VAR(i) models are

$$\text{BIC}(i) = \ln(|\tilde{\boldsymbol{\Sigma}}_i|) + \frac{k^2 i \ln(T)}{T},$$

$$\text{HQ}(i) = \ln(|\tilde{\boldsymbol{\Sigma}}_i|) + \frac{2k^2 i \ln(\ln(T))}{T}.$$

The HQ criterion is proposed by Hannan and Quinn (1979).

Example 8.4. Assuming that the bivariate series of monthly log returns of IBM stock and the S&P 500 index discussed in Example 8.1 follows a VAR model, we apply the $M(i)$ statistics and AIC to the data. Table 8.3 shows the results of these statistics. Both statistics indicate that a VAR(3) model might be adequate for the data. The $M(i)$ statistics are marginally significant at lags 1, 3, and 5 at the 5% level. The minimum of AIC occurs at order 3. For this particular instance, the $M(i)$ statistics are nonsignificant at the 1% level, confirming the previous observation that the dynamic linear dependence between the two return series is weak.

Table 8.3. Order-Specification Statistics[a] for the Monthly Log Returns of IBM Stock and the S&P 500 Index from January 1926 to December 1999

Order	1	2	3	4	5	6
$M(i)$	9.81	8.93	12.57	6.08	9.56	2.80
AIC	6.757	6.756	6.750	6.753	6.751	6.756

[a]The 5% and 1% critical values of a chi-squared distribution with 4 degrees of freedom are 9.5 and 13.3.

Estimation and Model Checking

For a specified VAR model, one can estimate the parameters using either the ordinary least squares method or the maximum likelihood method. The two methods are asymptotically equivalent. Under some regularity conditions, the estimates are asymptotically normal; see Reinsel (1993). A fitted model should then be checked carefully for any possible inadequacy. The $Q_k(m)$ statistic can be applied to the residual series to check the assumption that there are no serial or cross-correlations in the residuals. For a fitted VAR(p) model, the $Q_k(m)$ statistic of the residuals is asymptotically a chi-squared distribution with $k^2m - g$ degrees of freedom, where g is the number of estimated parameters in the AR coefficient matrices.

Example 8.4 (Continued). Table 8.4a shows the estimation results of a VAR(3) model for the bivariate series of monthly log returns of IBM stock and the S&P 500 index. The specified model is in the form

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_3 \boldsymbol{r}_{t-3} + \boldsymbol{a}_t, \tag{8.20}$$

where the first component of $\boldsymbol{r}_t$ denotes IBM stock returns. For this particular instance, we only use AR coefficient matrices at lags 1 and 3 because of the weak serial dependence of the data. In general, when the $M(i)$ statistics and the AIC criterion specify a VAR(3) model, all three AR lags should be used. Table 8.4b shows the estimation results after some statistically insignificant parameters are set to zero. The $Q_k(m)$ statistics of the residual series for the fitted model in Table 8.4b give $Q_2(4) = 18.17$ and $Q_2(8) = 41.26$. Since the fitted VAR(3) model has four parameters in the AR coefficient matrices, these two $Q_k(m)$ statistics are distributed asymptotically as a chi-squared distribution with degrees of freedom 12 and 28, respectively. The p-values of the test statistics are 0.111 and 0.051, and hence the fitted model is adequate at the 5% significance level. As shown by the univariate analysis, the return series are likely to have conditional heteroscedasticity. We discuss multivariate volatility in Chapter 10.

From the fitted model in Table 8.4b, we make the following observations. (a) The concurrent correlation coefficient between the two innovational series is $23.51/\sqrt{44.48 \times 31.29} = 0.63$, which, as expected, is close to the sample correlation coefficient between r_{1t} and r_{2t}. (b) The two log return series have positive and significant means, implying that the log prices of the two series had an upward trend over the data span. (c) The model shows that

$$\text{IBM}_t = 1.24 + 0.117\text{SP5}_{t-1} - 0.083\text{SP5}_{t-3} + a_{1t},$$

$$\text{SP5}_t = 0.57 + 0.073\text{SP5}_{t-1} - 0.109\text{SP5}_{t-3} + a_{2t}.$$

Consequently, at the 5% significance level, there is a unidirectional dynamic relationship from the monthly S&P 500 index return to the IBM return. If the S&P 500 index represents the U.S. stock market, then IBM return is affected by the past movements of the market. However, past movements of IBM stock returns do not

Table 8.4. Estimation Results of a VAR(3) Model for the Monthly Log Returns, in Percentages, of IBM Stock and the S&P 500 Index from January 1926 to December 1999

Parameter	ϕ_0	Φ_1		Φ_3		Σ	
(a) Full Model							
Estimate	1.20	0.011	0.108	0.039	−0.112	44.44	23.51
	0.58	−0.013	0.084	−0.007	−0.105	23.51	31.29
Standard error	0.23	0.043	0.051	0.044	0.052		
	0.19	0.036	0.043	0.037	0.044		
(b) Simplified Model							
Estimate	1.24	0	0.117	0	−0.083	44.48	23.51
	0.57	0	0.073	0	−0.109	23.51	31.29
Standard error	0.23	—	0.040	—	0.040		
	0.19	—	0.033	—	0.033		

significantly affect the U.S. market, even though the two returns have substantial concurrent correlation. Finally, the fitted model can be written as

$$\begin{bmatrix} \text{IBM}_t \\ \text{SP5}_t \end{bmatrix} = \begin{bmatrix} 1.24 \\ 0.57 \end{bmatrix} + \begin{bmatrix} 0.117 \\ 0.073 \end{bmatrix} \text{SP5}_{t-1} - \begin{bmatrix} 0.083 \\ 0.109 \end{bmatrix} \text{SP5}_{t-3} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

indicating that SP5_t is the driving factor of the bivariate series.

Forecasting

Treating a properly built model as the true model, one can apply the same techniques as those in the univariate analysis to produce forecasts and standard deviations of the associated forecast errors. For a VAR(p) model, the 1-step ahead forecast at the time origin h is $\boldsymbol{r}_h(1) = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+1-i}$, and the associated forecast error is $\boldsymbol{e}_h(1) = \boldsymbol{a}_{h+1}$. The covariance matrix of the forecast error is $\boldsymbol{\Sigma}$. For 2-step ahead forecasts, we substitute $\boldsymbol{r}_{h+1}$ by its forecast to obtain

$$\boldsymbol{r}_h(2) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_h(1) + \sum_{i=2}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+2-i},$$

and the associated forecast error is

$$\boldsymbol{e}_h(2) = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1[\boldsymbol{r}_t - \boldsymbol{r}_h(1)] = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1 \boldsymbol{a}_{h+1}.$$

The covariance matrix of the forecast error is $\boldsymbol{\Sigma} + \boldsymbol{\Phi}_1 \boldsymbol{\Sigma} \boldsymbol{\Phi}_1'$. If $\boldsymbol{r}_t$ is weakly stationary, then the ℓ-step ahead forecast $\boldsymbol{r}_h(\ell)$ converges to its mean vector $\boldsymbol{\mu}$ as

Table 8.5. Forecasts of a VAR(3) Model for the Monthly Log Returns, in Percentages, of IBM Stock and the S&P 500 Index: Forecast Origin December 1999

Step	1	2	3	4	5	6
IBM forecast	1.40	1.12	0.82	1.21	1.27	1.31
Standard error	6.67	6.70	6.70	6.72	6.72	6.72
S&P forecast	0.32	0.38	−0.02	0.53	0.56	0.61
Standard error	5.59	5.61	5.61	5.64	5.64	5.64

the forecast horizon ℓ increases and the covariance matrix of its forecast error converges to the covariance matrix of $\boldsymbol{r}_t$.

Table 8.5 provides 1-step to 6-step ahead forecasts of the monthly log returns, in percentages, of IBM stock and the S&P 500 index at the forecast origin $h = 888$. These forecasts are obtained by the refined VAR(3) model in Table 8.4.

In summary, building a VAR model involves three steps: (a) use the test statistic $M(i)$ or some information criterion to identify the order, (b) estimate the specified model by using the least squares method and, if necessary, reestimate the model by removing statistically insignificant parameters, and (c) use the $Q_k(m)$ statistic of the residuals to check the adequacy of a fitted model. Other characteristics of the residual series, such as conditional heteroscedasticity and outliers, can also be checked. If the fitted model is adequate, then it can be used to obtain forecasts and make inference concerning the dynamic relationship between the variables.

We used SCA to perform the analysis in this subsection. The commands used include `miden`, `mtsm`, `mest`, and `mfore`, where the prefix `m` stands for multivariate. Details of the commands and output are shown below.

SCA Demonstration

Output edited and % denotes explanation.

```
input ibm,sp5. file 'm-ibmspln.txt'
 -- % Order selection
miden ibm,sp5. no ccm. arfits 1 to 6.

 TIME PERIOD ANALYZED . . . . . .  1  TO   888

 SERIES   NAME          MEAN       STD. ERROR
    1     IBM           1.2402      6.7249
    2     SP5           0.5372      5.6415
=========  STEPWISE AUTOREGRESSION SUMMARY  ======
 ---------------------------------------------------
     I RESIDUAL I EIGENVAL.I CHI-SQ  I          I SIGN.
 LAG I VARIANCESI OF SIGMA I  TEST   I   AIC    I
 ----+----------+---------+--------+----------+------
   1 I .447E+02 I .135E+02 I    9.81 I    6.757 I . +
```

```
     I .318E+02 I .629E+02 I          I          I . .
 ----+----------+----------+---------+----------+------
   2 I .443E+02 I .135E+02 I    8.93 I    6.756 I + -
     I .317E+02 I .625E+02 I         I          I . .
 ----+----------+----------+---------+----------+------
   3 I .441E+02 I .134E+02 I   12.57 I    6.750 I . -
     I .313E+02 I .619E+02 I         I          I . -
 ----+----------+----------+---------+----------+------
   4 I .441E+02 I .133E+02 I    6.08 I    6.753 I . .
     I .312E+02 I .619E+02 I         I          I . .
 ----+----------+----------+---------+----------+------
   5 I .437E+02 I .133E+02 I    9.56 I    6.751 I . +
     I .309E+02 I .613E+02 I         I          I - +
 ----+----------+----------+---------+----------+------
   6 I .437E+02 I .133E+02 I    2.80 I    6.756 I . .
     I .308E+02 I .613E+02 I         I          I . .
 -----------------------------------------------------
CHI-SQUARED CRITICAL VALUES WITH 4 DEGREES OF FREEDOM ARE
          5 PERCENT:   9.5     1 PERCENT:  13.3
  -- % Specify a VAR(3) model with lags 1 & 3 only.
mtsm fit1. series ibm, sp5. @
model (i-p1*b-p3*b**3)series=c+noise.

 SUMMARY FOR MULTIVARIATE ARMA MODEL --   FIT1

  PARAMETER      FACTOR      ORDER     CONSTRAINT
  1        C   CONSTANT        0         CC
  2       P1    REG AR         1         CP1
  3       P3    REG AR         3         CP3
 -- % Perform multivariate estimation
mestim fit1. hold resi(r1,r2)
 ----- CONSTANT VECTOR (STD ERROR) -----
    1.201  (    0.232  )
    0.583  (    0.194  )
  ----- PHI MATRICES -----
 ESTIMATES OF  PHI(1) MATRIX AND SIGNIFICANCE
     .011     .108      . +
    -.013     .084      . .
  STANDARD ERRORS
     .043     .051
     .036     .043
 ESTIMATES OF  PHI(3) MATRIX AND SIGNIFICANCE
     .039    -.112      . -
    -.007    -.105      . -
  STANDARD ERRORS
     .044     .052
     .037     .044
----------------------
ERROR COVARIANCE MATRIX
```



```
----------------------
                1           2
  1    44.438125
  2    23.518578   31.287280
 -- % Set parameter to 0
p1(1,1)=0
 --
p1(2,1)=0
 -- % Set constraint to fix the parameter
cp1(1,1)=1
 --
cp1(2,1)=1
 --
p3(1,1)=0
 --
p3(2,1)=0
 --
cp3(1,1)=1
 --
cp3(2,1)=1
 --
mestim fit1. hold resi(r1,r2).
 ----- CONSTANT VECTOR (STD ERROR) -----
    1.243  (    0.226  )
    0.566  (    0.190  )
 ----- PHI MATRICES -----
 ESTIMATES OF  PHI(1) MATRIX AND SIGNIFICANCE
      .000     .117      . +
      .000     .073      . +
 STANDARD ERRORS
         --    .040
         --    .033
 ESTIMATES OF  PHI(3) MATRIX AND SIGNIFICANCE
     .000    -.083      . -
     .000    -.109      . -
 STANDARD ERRORS
         --    .040
         --    .033
----------------------
ERROR COVARIANCE MATRIX
----------------------
                1           2
  1    44.482888
  2    23.506951   31.293592
 -- % Compute residual CCM
miden r1,r2. maxl 12.
 -- % Produce 1 to 6-step ahead forecasts
mfore fit1. nofs 6.
```

8.2.5 Impulse Response Function

Similar to the univariate case, a VAR(p) model can be written as a linear function of the past innovations, that is,

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{a}_{t-2} + \cdots, \tag{8.21}$$

where $\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$ provided that the inverse exists, and the coefficient matrices $\boldsymbol{\psi}_i$ can be obtained by equating the coefficients of B^i in the equation

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)(\boldsymbol{I} + \boldsymbol{\Psi}_1 B + \boldsymbol{\Psi}_2 B^2 + \cdots) = \boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity matrix. This is a moving-average representation of $\boldsymbol{r}_t$ with the coefficient matrix $\boldsymbol{\Psi}_i$ being the impact of the past innovation $\boldsymbol{a}_{t-i}$ on $\boldsymbol{r}_t$. Equivalently, $\boldsymbol{\Psi}_i$ is the effect of $\boldsymbol{a}_t$ on the future observation $\boldsymbol{r}_{t+i}$. Therefore, $\boldsymbol{\Psi}_i$ is often referred to as the *impulse response function* of $\boldsymbol{r}_t$. However, since the components of $\boldsymbol{a}_t$ are often correlated, the interpretation of elements in $\boldsymbol{\Psi}_i$ of Eq. (8.21) is not straightforward. To aid interpretation, one can use the Cholesky decomposition mentioned earlier to transform the innovations so that the resulting components are uncorrelated. Specifically, there exists a lower triangular matrix $\boldsymbol{L}$ such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$, where $\boldsymbol{G}$ is a diagonal matrix and the diagonal elements of $\boldsymbol{L}$ are unity. See Eq. (8.9). Let $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then, $\text{Cov}(\boldsymbol{b}_t) = \boldsymbol{G}$ so that the elements b_{jt} are uncorrelated. Rewrite Eq. (8.21) as

$$\begin{aligned}\boldsymbol{r}_t &= \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{a}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \boldsymbol{\Psi}_0^* \boldsymbol{b}_t + \boldsymbol{\Psi}_1^* \boldsymbol{b}_{t-1} + \boldsymbol{\Psi}_2^* \boldsymbol{b}_{t-2} + \cdots,\end{aligned} \tag{8.22}$$

where $\boldsymbol{\Psi}_0^* = \boldsymbol{L}$ and $\boldsymbol{\Psi}_i^* = \boldsymbol{\Psi}_i \boldsymbol{L}$. The coefficient matrices $\boldsymbol{\Psi}_i^*$ are called the *impulse response function* of $\boldsymbol{r}_t$ with the orthogonal innovations $\boldsymbol{b}_t$. Specifically, the (i, j)th element of $\boldsymbol{\Psi}_\ell^*$, that is, $\Psi_{ij}^*(\ell)$, is the impact of $b_{j,t}$ on the future observation $r_{i,t+\ell}$. In practice, one can further normalize the orthogonal innovation $\boldsymbol{b}_t$ such that the variance of b_{it} is one. A weakness of the above orthogonalization is that the result depends on the ordering of the components of $\boldsymbol{r}_t$. In particular, $b_{1t} = a_{1t}$ so that a_{1t} is not transformed. Different orderings of the components of $\boldsymbol{r}_t$ may lead to different impulse response functions.

Both SCA and S-Plus enable one to obtain the impulse response function of a fitted VAR model. To demonstrate analysis of VAR models in S-Plus, we again use the monthly log return series of IBM stock and the S&P 500 index of Example 8.1. For details of S-Plus commands, see Zivot and Wang (2003).

S-Plus Demonstration

Output edited.

```
> x=matrix(scan(file='m-ibmspln.txt'),2) % Load data
> ibm=x[1,]
> sp5=x[2,]
```



```
> y=cbind(ibm,sp5)    % Create a vector series
> y1=data.frame(y)    % create a data frame
> ord.choice=VAR(y1,max.ar=6) % order selection
> ord.choice$info
       ar(1)     ar(2)     ar(3)     ar(4)     ar(5)     ar(6)
BIC 10998.47 11016.61 11031.07 11052.05 11069.49 11093.78
> ord.choice=VAR(y1,max.ar=6,criterion='AIC')
> ord.choice$info
       ar(1)     ar(2)     ar(3)     ar(4)     ar(5)     ar(6)
AIC 10969.78 10968.79 10964.11 10965.97 10964.28 10969.44
```

The AIC selects a VAR(3) model as before, but BIC selects a VAR(1) model. For simplicity, we shall use VAR(1) specification in the demonstration. Note that different normalizations are used between the two packages so that the values of information criteria appear to be different; see the AIC in Table 8.3. This is not important because normalization does not affect order selection. Turn to estimation.

```
> var1.fit=VAR(y~ar(1))   % Estimation
> summary(var1.fit)
Call:
VAR(formula = y ~ ar(1))
Coefficients:
                ibm      sp5
(Intercept)  1.1627   0.4993
  (std.err)  0.2290   0.1925
   (t.stat)  5.0777   2.5935

   ibm.lag1  0.0192 -0.0054
  (std.err)  0.0433  0.0364
   (t.stat)  0.4429 -0.1487

   sp5.lag1  0.1062  0.0802
  (std.err)  0.0517  0.0435
   (t.stat)  2.0544  1.8454
Regression Diagnostics:
                ibm    sp5
     R-squared 0.0105 0.0058
Adj. R-squared 0.0082 0.0036
  Resid. Scale 6.7043 5.6376

> plot(var1.fit)
Make a plot selection (or 0 to exit):
1: plot: All
2: plot: Response and Fitted Values
3: plot: Residuals
...
8: plot: PACF of Squared Residuals
Selection: 3
```

The fitted model is

$$\text{IBM}_t = 1.16 + 0.02\text{IBM}_{t-1} + 0.11\text{SP5}_{t-1} + a_{1t},$$
$$\text{SP5}_t = 0.50 - 0.01\text{IBM}_{t-1} + 0.08\text{SP5}_{t-1} + a_{2t}.$$

Based on t-statistics of the estimates in the output, only the lagged variable SP5_{t-1} is informative in both equations. Figure 8.5 shows the time plots of the two residual series, where the two horizontal lines indicate the two standard-error limits. As expected, there exist clusters of outlying observations.

Next, we compute 1-step to 6-step ahead forecasts and the impulse response function of the fitted VAR(1) model when the IBM stock return is the first component of $\boldsymbol{r}_t$. Compared with those of a VAR(3) model in Table 8.5, the forecasts of the VAR(1) model converge faster to the sample mean of the series.

```
> var1.pred=predict(var1.fit,n.predict=6) % Compute prediction
> summary(var1.pred)
Predicted Values with Standard Errors:
                ibm    sp5
1-step-ahead 1.8472 0.9255
   (std.err) 6.7043 5.6376
2-step-ahead 1.2964 0.5636
   (std.err) 6.7394 5.6539
3-step-ahead 1.2474 0.5375
   (std.err) 6.7397 5.6540
...
6-step-ahead 1.2434 0.5356
   (std.err) 6.7397 5.6540
```

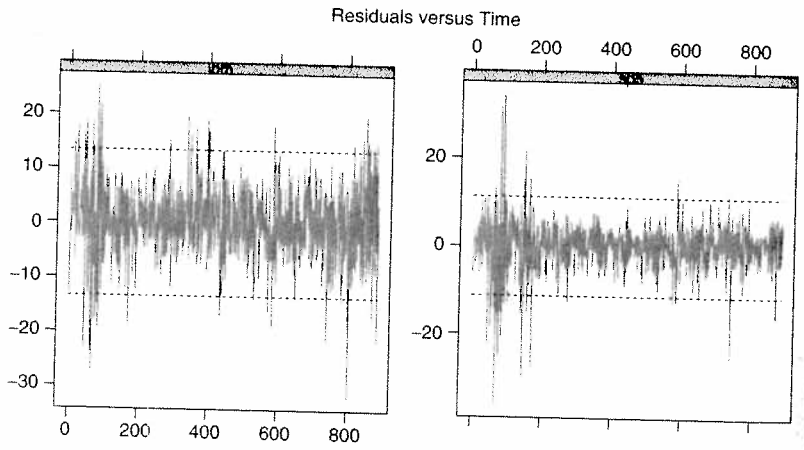

Figure 8.5. Residual plots of fitting a VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.


```
> plot(var1.pred,y, n.old=12)  % Plot forecasts
> var1.irf=impRes(var1.fit,period=6,std.err='asymptotic')
> summary(var1.irf)
Impulse Response Function:

(with responses in rows, and innovations in columns)
, , lag.0
              ibm    sp5
      ibm 6.6929 0.0000
(std.err) 0.1589 0.0000
      sp5 3.5645 4.3553
(std.err) 0.1690 0.1034
, , lag.1
              ibm    sp5
      ibm 0.5069 0.4624
(std.err) 0.2244 0.2249
      sp5 0.2496 0.3492
(std.err) 0.1885 0.1891
....
> plot(var1.irf)
```

Figure 8.6 shows the forecasts and their pointwise 95% confidence intervals along with the last 12 data points of the series. Figure 8.7 shows the impulse response functions of the fitted VAR(1) model where the IBM stock return is the first component of $\boldsymbol{r}_t$. Since the dynamic dependence of the returns is weak, the impulse response functions exhibit simple patterns and decay quickly.

8.3 VECTOR MOVING-AVERAGE MODELS

A vector moving-average model of order q, or VMA(q), is in the form

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1\boldsymbol{a}_{t-1} - \cdots - \boldsymbol{\Theta}_q\boldsymbol{a}_{t-q} \quad \text{or} \quad \boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{8.23}$$

where $\boldsymbol{\theta}_0$ is a k-dimensional vector, $\boldsymbol{\Theta}_i$ are $k \times k$ matrices, and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ is the MA matrix polynomial in the back-shift operator B. Similar to the univariate case, VMA(q) processes are weakly stationary provided that the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$ exists. Taking expectation of Eq. (8.23), we obtain that $\boldsymbol{\mu} = E(\boldsymbol{r}_t) = \boldsymbol{\theta}_0$. Thus, the constant vector $\boldsymbol{\theta}_0$ is the mean vector of $\boldsymbol{r}_t$ for a VMA model.

Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$ be the mean-corrected VAR(q) process. Then using Eq. (8.23) and the fact that $\{\boldsymbol{a}_t\}$ has no serial correlations, we have

1. $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$,
2. $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1\boldsymbol{\Sigma}\boldsymbol{\Theta}_1' + \cdots + \boldsymbol{\Theta}_q\boldsymbol{\Sigma}\boldsymbol{\Theta}_q'$,
3. $\boldsymbol{\Gamma}_\ell = \boldsymbol{0}$ if $\ell > q$, and
4. $\boldsymbol{\Gamma}_\ell = \sum_{j=\ell}^{q} \boldsymbol{\Theta}_j\boldsymbol{\Sigma}\boldsymbol{\Theta}_{j-\ell}'$ if $1 \le \ell \le q$, where $\boldsymbol{\Theta}_0 = -\boldsymbol{I}$.

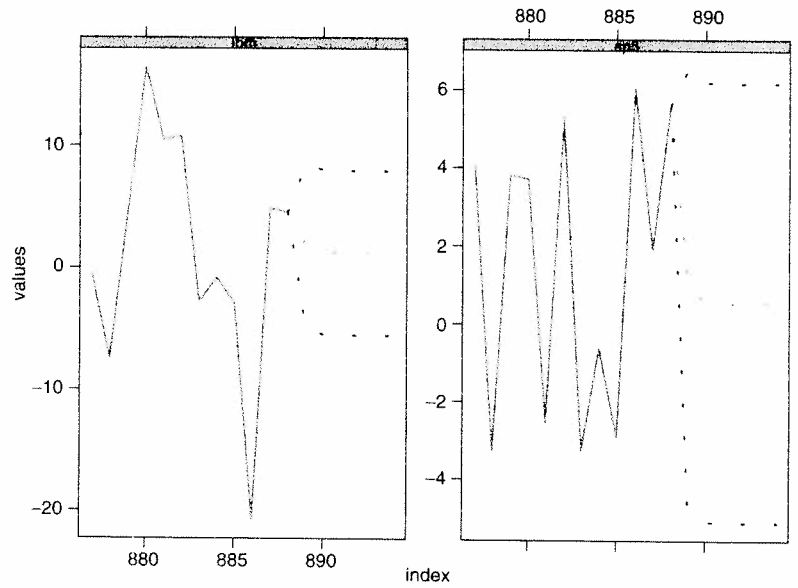

Figure 8.6. Forecasting plots of a fitted VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.

Since $\mathbf{\Gamma}_\ell = \mathbf{0}$ for $\ell > q$, the cross-correlation matrices (CCMs) of a VMA(q) process $\boldsymbol{r}_t$ satisfy

$$\boldsymbol{\rho}_\ell = \mathbf{0}, \quad \ell > q. \tag{8.24}$$

Therefore, similar to the univariate case, the sample CCMs can be used to identify the order of a VMA process.

To better understand the VMA processes, let us consider the bivariate MA(1) model

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \mathbf{\Theta}\boldsymbol{a}_{t-1} = \boldsymbol{\mu} + \boldsymbol{a}_t - \mathbf{\Theta}\boldsymbol{a}_{t-1}, \tag{8.25}$$

where, for simplicity, the subscript of $\mathbf{\Theta}_1$ is removed. This model can be written explicitly as

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \tag{8.26}$$

It says that the current return series $\boldsymbol{r}_t$ only depends on the current and past shocks. Therefore, the model is a finite-memory model.

Consider the equation for r_{1t} in Eq. (8.26). The parameter Θ_{12} denotes the linear dependence of r_{1t} on $a_{2,t-1}$ in the presence of $a_{1,t-1}$. If $\Theta_{12} = 0$, then r_{1t} does not depend on the lagged values of a_{2t} and, hence, the lagged values of r_{2t}. Similarly,

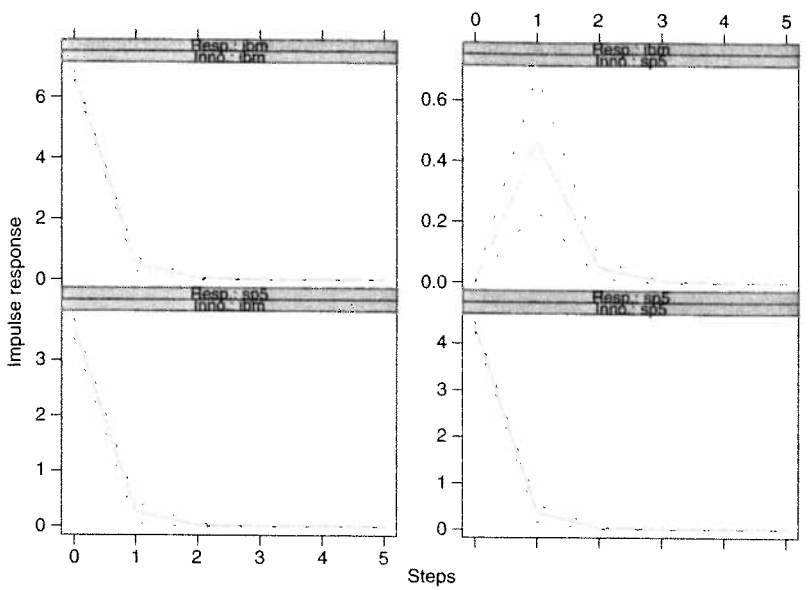

Figure 8.7. Plots of impulse response functions of orthogonal innovations for a fitted VAR(1) model to the monthly log returns, in percentages, of IBM stock and the S&P 500 index. The sample period is from January 1926 to December 1999.

if $\Theta_{21} = 0$, then r_{2t} does not depend on the past values of r_{1t}. The off-diagonal elements of $\mathbf{\Theta}$ thus show the dynamic dependence between the component series. For this simple VMA(1) model, we can classify the relationships between r_{1t} and r_{2t} as follows:

1. They are uncoupled series if $\Theta_{12} = \Theta_{21} = 0$.
2. There is a unidirectional dynamic relationship from r_{1t} to r_{2t} if $\Theta_{12} = 0$, but $\Theta_{21} \neq 0$. The opposite unidirectional relationship holds if $\Theta_{21} = 0$, but $\Theta_{12} \neq 0$.
3. There is a feedback relationship between r_{1t} and r_{2t} if $\Theta_{12} \neq 0$ and $\Theta_{21} \neq 0$.

Finally, the concurrent correlation between r_{it} is the same as that between a_{it}. The previous classification can be generalized to a VMA(q) model.

Estimation

Unlike the VAR models, estimation of VMA models is much more involved; see Hillmer and Tiao (1979), Lütkepohl (1991), and the references therein. For the likelihood approach, there are two methods available. The first is the conditional likelihood method that assumes that $\boldsymbol{a}_t = \mathbf{0}$ for $t \leq 0$. The second is the exact likelihood method that treats $\boldsymbol{a}_t$ with $t \leq 0$ as additional parameters of

the model. To gain some insight into the problem of estimation, we consider the VMA(1) model in Eq. (8.25). Suppose that the data are $\{\boldsymbol{r}_t | t = 1, \ldots, T\}$ and $\boldsymbol{a}_t$ is multivariate normal. For a VMA(1) model, the data depend on $\boldsymbol{a}_0$.

Conditional MLE

The conditional likelihood method assumes that $\boldsymbol{a}_0 = \boldsymbol{0}$. Under such an assumption and rewriting the model as $\boldsymbol{a}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}$, we can compute the shock $\boldsymbol{a}_t$ recursively as

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0, \quad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}_1 \boldsymbol{a}_1, \quad \cdots.$$

Consequently, the likelihood function of the data becomes

$$f(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T | \boldsymbol{\theta}_0, \boldsymbol{\Theta}_1, \boldsymbol{\Sigma}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}\boldsymbol{a}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_t\right),$$

which can be evaluated to obtain the parameter estimates.

Exact MLE

For the exact likelihood method, $\boldsymbol{a}_0$ is an unknown vector that must be estimated from the data to evaluate the likelihood function. For simplicity, let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$ be the mean-corrected series. Using $\tilde{\boldsymbol{r}}_t$ and Eq. (8.25), we have

$$\boldsymbol{a}_t = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}. \tag{8.27}$$

By repeated substitutions, $\boldsymbol{a}_0$ is related to all $\tilde{\boldsymbol{r}}_t$ as

$$\begin{aligned}
\boldsymbol{a}_1 &= \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\boldsymbol{a}_0, \\
\boldsymbol{a}_2 &= \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^2\boldsymbol{a}_0, \\
&\vdots = \vdots \\
\boldsymbol{a}_T &= \tilde{\boldsymbol{r}}_T + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{T-1} + \cdots + \boldsymbol{\Theta}^{T-1}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^T\boldsymbol{a}_0.
\end{aligned} \tag{8.28}$$

Thus, $\boldsymbol{a}_0$ is a linear function of the data if $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}$ are given. This result enables us to estimate $\boldsymbol{a}_0$ using the data and initial estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}$. More specifically, given $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and the data, we can define

$$\boldsymbol{r}_t^* = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Theta}^{t-1}\tilde{\boldsymbol{r}}_1, \quad \text{for} \quad t = 1, 2, \ldots, T.$$

Equation (8.28) can then be rewritten as

$$\begin{aligned}
\boldsymbol{r}_1^* &= -\boldsymbol{\Theta}\boldsymbol{a}_0 + \boldsymbol{a}_1, \\
\boldsymbol{r}_2^* &= -\boldsymbol{\Theta}^2\boldsymbol{a}_0 + \boldsymbol{a}_2, \\
&\vdots = \vdots \\
\boldsymbol{r}_T^* &= -\boldsymbol{\Theta}^T\boldsymbol{a}_0 + \boldsymbol{a}_T.
\end{aligned}$$

This is in the form of a multiple linear regression with parameter vector $\boldsymbol{a}_0$, even though the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$ may not be a diagonal matrix. If initial estimate of $\boldsymbol{\Sigma}$ is also available, one can premultiply each equation of the prior system by $\boldsymbol{\Sigma}^{-1/2}$, which is the square-root matrix of $\boldsymbol{\Sigma}$. The resulting system is indeed a multiple linear regression, and the ordinary least squares method can be used to obtain an estimate of $\boldsymbol{a}_0$. Denote the estimate by $\widehat{\boldsymbol{a}}_0$.

Using the estimate $\widehat{\boldsymbol{a}}_0$, we can compute the shocks $\boldsymbol{a}_t$ recursively as

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\widehat{\boldsymbol{a}}_0, \quad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_1, \quad \ldots.$$

This recursion is a linear transformation from $(\boldsymbol{a}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ to $(\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_T)$, from which we can (a) obtain the joint distribution of $\boldsymbol{a}_0$ and the data, and (2) integrate out $\boldsymbol{a}_0$ to derive the exact likelihood function of the data. The resulting likelihood function can then be evaluated to obtain the exact ML estimates. For details, see Hillmer and Tiao (1979).

In summary, the exact likelihood method works as follows. Given initial estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$, one uses Eq. (8.28) to derive an estimate of $\boldsymbol{a}_0$. This estimate is in turn used to compute $\boldsymbol{a}_t$ recursively using Eq. (8.27) and starting with $\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\widehat{\boldsymbol{a}}_0$. The resulting $\{\boldsymbol{a}_t\}_{t=1}^T$ are then used to evaluate the exact likelihood function of the data to update the estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$. The whole process is then repeated until the estimates converge. This iterative method to evaluate the exact likelihood function applies to the general VMA(q) models.

From the previous discussion, the exact likelihood method requires more intensive computation than the conditional likelihood approach does. But it provides more accurate parameter estimates, especially when some eigenvalues of $\boldsymbol{\Theta}$ are close to 1 in modulus. Hillmer and Tiao (1979) provide some comparison between the conditional and exact likelihood estimations of VMA models. In multivariate time series analysis, the exact maximum likelihood method becomes important if one suspects that the data might have been overdifferenced. Overdifferencing may occur in many situations (e.g., differencing individual components of a cointegrated system; see discussion later on cointegration).

In summary, building a VMA model involves three steps: (a) use the sample cross-correlation matrices to specify the order q—for a VMA(q) model, $\boldsymbol{\rho}_\ell = \boldsymbol{0}$ for $\ell > q$; (b) estimate the specified model by using either the conditional or exact likelihood method—the exact method is preferred when the sample size is not large; and (c) the fitted model should be checked for adequacy (e.g., applying the $Q_k(m)$ statistics to the residual series). Finally, forecasts of a VMA model can be obtained by using the same procedure as a univariate MA model.

Example 8.5. Consider again the bivariate series of monthly log returns in percentages of IBM stock and the S&P 500 index from January 1926 to December 1999. Since significant cross-correlations occur mainly at lags 1 and 3, we employ

Table 8.6. Estimation Results for Monthly Log Returns of IBM Stock and the S&P 500 Index Using the Vector Moving-Average Model in Eq. (8.29): January 1926 to December 1999

Parameter	θ_0	Θ_1		Θ_3		Σ	
(a) Full Model with Conditional Likelihood Method							
Estimate	1.24	−0.013	−0.121	−0.038	0.108	44.48	23.52
	0.54	0.020	−0.101	0.014	0.105	23.52	31.20
Standard error	0.24	0.043	0.051	0.044	0.052		
	0.18	0.036	0.043	0.036	0.043		
(b) Full Model with Exact Likelihood Method							
Estimate	1.24	−0.013	−0.121	−0.038	0.108	44.48	23.52
	0.54	0.020	−0.101	0.013	0.105	23.52	31.20
Standard error	0.24	0.043	0.051	0.044	0.052		
	0.18	0.036	0.043	0.036	0.043		
(c) Simplified Model with Exact Likelihood Method							
Estimate	1.24	0.000	−0.126	0.000	0.082	44.54	23.51
	0.54	0.000	−0.084	0.000	0.114	23.51	31.21
Standard error	0.23	—	0.040	—	0.040		
	0.18	—	0.033	—	0.033		

the VMA(3) model

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1 \boldsymbol{a}_{t-1} - \boldsymbol{\Theta}_3 \boldsymbol{a}_{t-3} \tag{8.29}$$

for the data. Table 8.6 shows the estimation results of the model. The $Q_k(m)$ statistics for the residuals of the simplified model give $Q_2(4) = 17.25$ and $Q_2(8) = 39.30$. Compared with chi-squared distributions with 12 and 28 degrees of freedom, the p-values of these statistics are 0.1404 and 0.0762, respectively. Thus, the model is adequate at the 5% significance level.

From Table 8.6, we make the following observations:

1. The difference between conditional and exact likelihood estimates is small for this particular example. This is not surprising because the sample size is not small and, more important, the dynamic structure of the data is weak.
2. The VMA(3) model provides essentially the same dynamic relationship for the series as that of the VAR(3) model in Example 8.4. The monthly log return of IBM stock depends on the previous returns of the S&P 500 index. The market return, in contrast, does not depend on lagged returns of IBM stock. In other words, the dynamic structure of the data is driven by the market return, not by IBM return. The concurrent correlation between the two returns remains strong, however.

8.4 VECTOR ARMA MODELS

Univariate ARMA models can also be generalized to handle vector time series. The resulting models are called VARMA models. The generalization, however, encounters some new issues that do not occur in developing VAR and VMA models. One of the issues is the *identifiability* problem. Unlike the univariate ARMA models, VARMA models may not be uniquely defined. For example, the VMA(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

is *identical* to the VAR(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The equivalence of the two models can easily be seen by examining their component models. For the VMA(1) model, we have

$$r_{1t} = a_{1t} - 2a_{2,t-1}, \quad r_{2t} = a_{2t}.$$

For the VAR(1) model, the equations are

$$r_{1t} + 2r_{2,t-1} = a_{1t}, \quad r_{2t} = a_{2t}.$$

From the model for r_{2t}, we have $r_{2,t-1} = a_{2,t-1}$. Therefore, the models for r_{1t} are identical. This type of identifiability problem is harmless because either model can be used in a real application.

Another type of identifiability problem is more troublesome. Consider the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

This model is identical to the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix},$$

for any nonzero ω and η. In this particular instance, the equivalence occurs because we have $r_{2t} = a_{2t}$ in both models. The effects of the parameters ω and η on the system cancel out between AR and MA parts of the second model. Such an identifiability problem is serious because, without proper constraints, the likelihood function of a vector ARMA(1,1) model for the data is not uniquely defined, resulting in a situation similar to the exact multicollinearity in a regression analysis.

This type of identifiability problem can occur in a vector model even if none of the components is a white noise series.

These two simple examples highlight the new issues involved in the generalization to VARMA models. Building a VARMA model for a given data set thus requires some attention. In the time series literature, methods of *structural specification* have been proposed to overcome the identifiability problem; see Tiao and Tsay (1989), Tsay (1991), and the references therein. We do not discuss the detail of structural specification here because VAR and VMA models are sufficient in most financial applications. When VARMA models are used, only lower order models are entertained (e.g., a VARMA(1,1) or VARMA(2,1) model) especially when the time series involved are not seasonal.

A VARMA(p, q) model can be written as

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t,$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are two $k \times k$ matrix polynomials. We assume that the two matrix polynomials have no left common factors; otherwise, the model can be simplified. The necessary and sufficient condition of weak stationarity for $\boldsymbol{r}_t$ is the same as that for the VAR(p) model with matrix polynomial $\boldsymbol{\Phi}(B)$. For $v > 0$, the (i, j)th elements of the coefficient matrices $\boldsymbol{\Phi}_v$ and $\boldsymbol{\Theta}_v$ measure the linear dependence of r_{1t} on $r_{j,t-v}$ and $a_{j,t-v}$, respectively. If the (i, j)th element is zero for all AR and MA coefficient matrices, then r_{it} does not depend on the lagged values of r_{jt}. However, the converse proposition does not hold in a VARMA model. In other words, nonzero coefficients at the (i, j)th position of AR and MA matrices may exist even when r_{it} does not depend on any lagged value of r_{jt}.

To illustrate, consider the following bivariate model:

$$\begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \Theta_{11}(B) & \Theta_{12}(B) \\ \Theta_{21}(B) & \Theta_{22}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Here the necessary and sufficient conditions for the existence of a unidirectional dynamic relationship from r_{1t} to r_{2t} are

$$\Phi_{22}(B)\Theta_{12}(B) - \Phi_{12}(B)\Theta_{22}(B) = 0,$$

but

$$\Phi_{11}(B)\Theta_{21}(B) - \Phi_{21}(B)\Theta_{11}(B) \neq 0. \tag{8.30}$$

These conditions can be obtained as follows. Letting

$$\Omega(B) = |\boldsymbol{\Phi}(B)| = \Phi_{11}(B)\Phi_{22}(B) - \Phi_{12}(B)\Phi_{21}(B)$$

be the determinant of the AR matrix polynomial and premultiplying the model by the matrix

$$\begin{bmatrix} \Phi_{22}(B) & -\Phi_{12}(B) \\ -\Phi_{21}(B) & \Phi_{11}(B) \end{bmatrix},$$

we can rewrite the bivariate model as

$$\Omega(B)\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \Phi_{22}(B)\Theta_{11}(B) - \Phi_{12}(B)\Theta_{21}(B) & \Phi_{22}(B)\Theta_{12}(B) - \Phi_{12}(B)\Theta_{22}(B) \\ \Phi_{11}(B)\Theta_{21}(B) - \Phi_{21}(B)\Theta_{11}(B) & \Phi_{11}(B)\Theta_{22}(B) - \Phi_{21}(B)\Theta_{12}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Consider the equation for r_{1t}. The first condition in Eq. (8.30) shows that r_{1t} does not depend on any past value of a_{2t} or r_{2t}. From the equation for r_{2t}, the second condition in Eq. (8.30) implies that r_{2t} indeed depends on some past values of a_{1t}. Based on Eq. (8.30), $\Theta_{12}(B) = \Phi_{12}(B) = 0$ is a sufficient, but not necessary, condition for the unidirectional relationship from r_{1t} to r_{2t}.

Estimation of a VARMA model can be carried out by either the conditional or exact maximum likelihood method. The $Q_k(m)$ statistic continues to apply to the residual series of a fitted model, but the degrees of freedom of its asymptotic chi-squared distribution are $k^2m - g$, where g is the number of estimated parameters in both the AR and MA coefficient matrices.

Example 8.6. To demonstrate VARMA modeling, we consider two U.S. monthly interest-rate series. The first series is the 1-year Treasury constant maturity rate, and the second series is the 3-year Treasury constant maturity rate. The data are obtained from the Federal Reserve Bank of St. Louis, and the sampling period is from April 1953 to January 2001. There are 574 observations. To ensure the positiveness of U.S. interest rates, we analyze the log series. Figure 8.8 shows the time plots of the two log interest-rate series. The solid line denotes the 1-year maturity rate. The two series moved closely in the sampling period.

The $M(i)$ statistics and AIC criterion specify a VAR(4) model for the data. However, we employ a VARMA(2,1) model because the two models provide similar fits. Table 8.7 shows the parameter estimates of the VARMA(2,1) model obtained by the exact likelihood method. We removed the insignificant parameters and reestimated the simplified model. The residual series of the fitted model has some minor serial and cross-correlations at lags 7 and 11. Figure 8.9 shows the residual plots and indicates the existence of some outlying data points. The model can be further improved, but it seems to capture the dynamic structure of the data reasonably well.

The final VARMA(2,1) model shows some interesting characteristics of the data. First, the interest-rate series are highly contemporaneously correlated. The concurrent correlation coefficient is $2.5/\sqrt{3.58 \times 2.19} = 0.893$. Second, there is a unidirectional linear relationship from the 3-year rate to the 1-year rate because the $(2, 1)$th elements of all AR and MA matrices are zero, but some $(1, 2)$th element is not zero. As a matter of fact, the model in Table 8.7 shows that

$$\begin{aligned} r_{3t} &= 0.025 + 0.99r_{3,t-1} + a_{3t} + 0.47a_{3,t-1}, \\ r_{1t} &= 0.028 + 1.82r_{1,t-1} - 0.84r_{1,t-2} - 0.97r_{3,t-1} + 0.98r_{3,t-2} \\ &\quad + a_{1t} - 0.90a_{1,t-1} + 1.66a_{3,t-1}, \end{aligned}$$

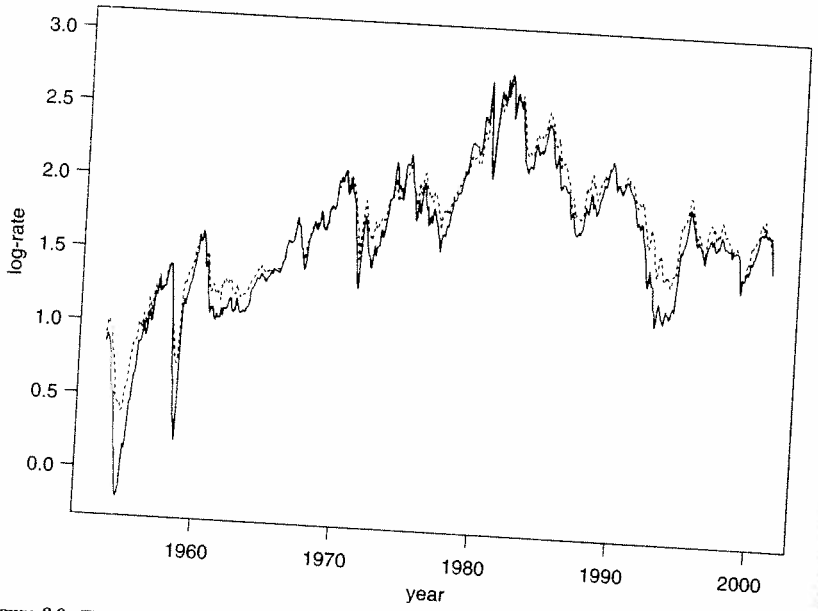

Figure 8.8. Time plots of log U.S. monthly interest rates from April 1953 to January 2001. The solid line denotes the 1-year Treasury constant maturity rate, and the dashed line denotes the 3-year rate.

Table 8.7. Parameter Estimates of a VARMA(2,1) Model for Two Monthly U.S. Interest-Rate Series Based on the Exact Likelihood Method

Parameter	Φ_1		Φ_2		ϕ_0	Θ_1		$\Sigma \times 10^3$	
Estimate	1.82	−0.97	−0.84	0.98	0.028	0.90	−1.66	3.58	2.50
		0.99			0.025		−0.47	2.50	2.19
Standard error	0.03	0.08	0.03	0.08	0.014	0.03	0.10		
		0.01			0.011		0.04		

where r_{it} is the log series of i-year interest rate and a_{it} is the corresponding shock series. Therefore, the 3-year interest rate does not depend on the past values of the 1-year rate, but the 1-year rate depends on the past values of the 3-year rate. Third, the two interest-rate series appear to be unit-root nonstationary. Using the back-shift operator B, the model can be rewritten approximately as

$$(1 - B)r_{3t} = 0.03 + (1 + 0.47B)a_{3t},$$
$$(1 - B)(1 - 0.82B)r_{1t} = 0.03 - 0.97B(1 - B)r_{3,t} + (1 - 0.9B)a_{1t} + 1.66Ba_{3,t}.$$

Finally, the SCA commands used in the analysis are given in Appendix C.

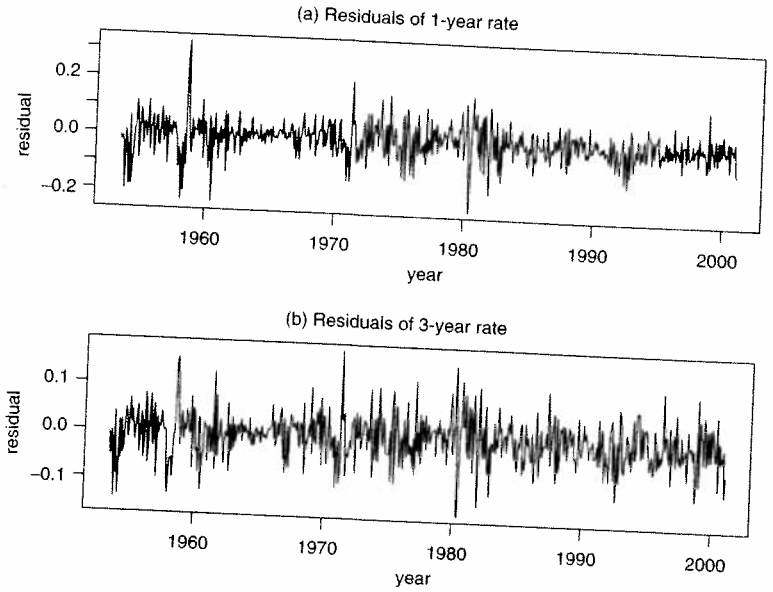

Figure 8.9. Residual plots for log U.S. monthly interest-rate series of Example 8.6. The fitted model is a VARMA(2,1).

8.4.1 Marginal Models of Components

Given a vector model for $\boldsymbol{r}_t$, the implied univariate models for the components r_{it} are the *marginal* models. For a k-dimensional ARMA(p, q) model, the marginal models are ARMA$[kp, (k-1)p + q]$. This result can be obtained in two steps. First, the marginal model of a VMA(q) model is univariate MA(q). Assume that $\boldsymbol{r}_t$ is a VMA(q) process. Because the cross-correlation matrix of $\boldsymbol{r}_t$ vanishes after lag q (i.e., $\boldsymbol{\rho}_\ell = \mathbf{0}$ for $\ell > q$), the ACF of r_{it} is zero beyond lag q. Therefore, r_{it} is an MA process and its univariate model is in the form $r_{it} = \theta_{i,0} + \sum_{j=1}^{q} \theta_{i,j} b_{i,t-j}$, where $\{b_{it}\}$ is a sequence of uncorrelated random variables with mean zero and variance σ_{ib}^2. The parameters $\theta_{i,j}$ and σ_{ib} are functions of the parameters of the VMA model for $\boldsymbol{r}_t$.

The second step to obtain the result is to diagonalize the AR matrix polynomial of a VARMA(p, q) model. For illustration, consider the bivariate AR(1) model

$$\begin{bmatrix} 1 - \Phi_{11}B & -\Phi_{12}B \\ -\Phi_{21}B & 1 - \Phi_{22}B \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Premultiplying the model by the matrix polynomial

$$\begin{bmatrix} 1 - \Phi_{22}B & \Phi_{12}B \\ \Phi_{21}B & 1 - \Phi_{11}B \end{bmatrix},$$

we obtain

$$[(1-\Phi_{11}B)(1-\Phi_{22}B)-\Phi_{12}\Phi_{22}B^2]\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 1-\Phi_{22}B & -\Phi_{12}B \\ -\Phi_{21}B & 1-\Phi_{11}B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The left-hand side of the prior equation shows that the univariate AR polynomials for r_{it} are of order 2. In contrast, the right-hand side of the equation is in a VMA(1) form. Using the result of VMA models in step 1, we show that the univariate model for r_{it} is ARMA(2,1). The technique generalizes easily to the k-dimensional VAR(1) model, and the marginal models are ARMA$(k, k-1)$. More generally, for a k-dimensional VAR(p) model, the marginal models are ARMA$[kp, (k-1)p]$. The result for VARMA models follows directly from those of VMA and VAR models.

The order $[kp, (k-1)p+q]$ is the maximum order (i.e., the upper bound) for the marginal models. The actual marginal order of r_{it} can be much lower.

8.5 UNIT-ROOT NONSTATIONARITY AND COINTEGRATION

When modeling several unit-root nonstationary time series jointly, one may encounter the case of *cointegration*. Consider the bivariate ARMA(1,1) model

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{bmatrix}\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix}\begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}, \tag{8.31}$$

where the covariance matrix $\boldsymbol{\Sigma}$ of the shock $\boldsymbol{a}_t$ is positive definite. This is not a weakly stationary model because the two eigenvalues of the AR coefficient matrix are 0 and 1. Figure 8.10 shows the time plots of a simulated series of the model with 200 data points and $\boldsymbol{\Sigma} = \boldsymbol{I}$, whereas Figure 8.11 shows the sample autocorrelations of the two component series x_{it}. It is easy to see that the two series have high autocorrelations and exhibit features of unit-root nonstationarity. The two marginal models of $\boldsymbol{x}_t$ are indeed unit-root nonstationary. Rewrite the model as

$$\begin{bmatrix} 1-0.5B & B \\ 0.25B & 1-0.5B \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.2B & 0.4B \\ 0.1B & 1-0.2B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Premultiplying the above equation by

$$\begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix},$$

we obtain the result

$$\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-0.7B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

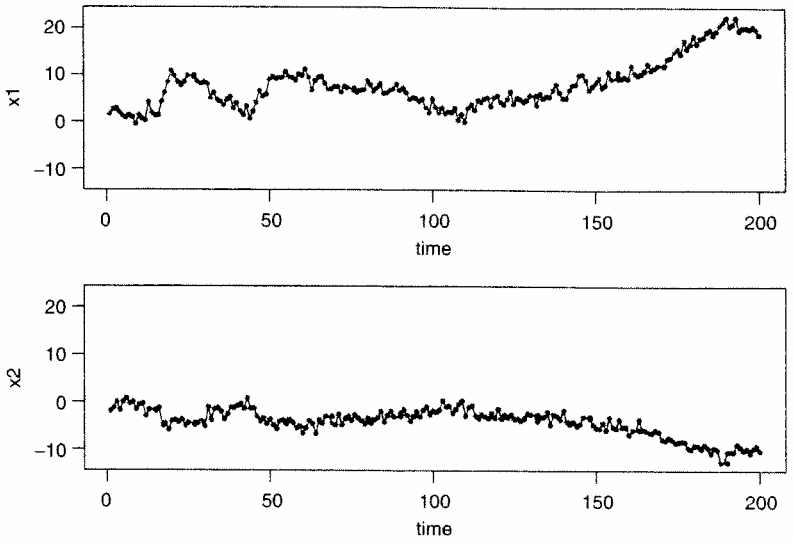

Figure 8.10. Time plots of a simulated series based on model (8.31) with identity covariance matrix for the shocks.

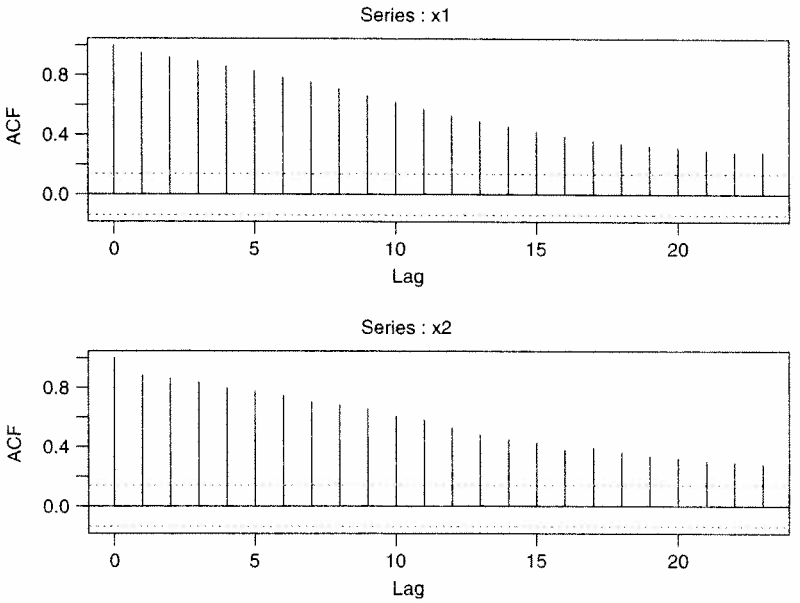

Figure 8.11. Sample autocorrelation functions of two simulated component series. There are 200 observations, and the model is given by Eq. (8.31) with identity covariance matrix for the shocks.

Therefore, each component x_{it} of the model is unit-root nonstationary and follows an ARIMA(0,1,1) model.

However, we can consider a linear transformation by defining

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{x}_t,$$

$$\begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{a}_t.$$

The VARMA model of the transformed series $\boldsymbol{y}_t$ can be obtained as follows:

$$\begin{aligned} \boldsymbol{L}\boldsymbol{x}_t &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{x}_{t-1}) + \boldsymbol{b}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{b}_{t-1}. \end{aligned}$$

Thus, the model for $\boldsymbol{y}_t$ is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{1,t-1} \\ b_{2,t-1} \end{bmatrix}. \tag{8.32}$$

From the prior model, we see that (a) y_{1t} and y_{2t} are uncoupled series with concurrent correlation equal to that between the shocks b_{1t} and b_{2t}, (b) y_{1t} follows a univariate ARIMA(0,1,1) model, and (c) y_{2t} is a white noise series (i.e., $y_{2t} = b_{2t}$). In particular, the model in Eq. (8.32) shows that there is *only* a single unit root in the system. Consequently, the unit roots of x_{1t} and x_{2t} are introduced by the unit root of y_{1t}. In the literature, y_{1t} is referred to as the *common trend* of x_{1t} and x_{2t}.

The phenomenon that both x_{1t} and x_{2t} are unit-root nonstationary, but there is only a single unit root in the vector series, is referred to as *cointegration* in the econometric and time series literature. Another way to define cointegration is to focus on linear transformations of unit-root nonstationary series. For the simulated example of model (8.31), the transformation shows that the linear combination $y_{2t} = 0.5x_{1t} + x_{2t}$ does not have a unit root. Consequently, x_{1t} and x_{2t} are cointegrated if (a) both of them are unit-root nonstationary, and (b) they have a linear combination that is unit-root stationary.

Generally speaking, for a k-dimensional unit-root nonstationary time series, cointegration exists if there are less than k unit roots in the system. Let h be the number of unit roots in the k-dimensional series $\boldsymbol{x}_t$. Cointegration exists if $0 < h < k$, and the quantity $k - h$ is called the number of cointegrating factors. Alternatively, the number of cointegrating factors is the number of different linear combinations that are unit-root stationary. The linear combinations are called the cointegrating vectors. For the prior simulated example, $y_{2t} = (0.5, 1)\boldsymbol{x}_t$ so that $(0.5, 1)'$ is a

cointegrating vector for the system. For more discussions on cointegration and cointegration tests, see Box and Tiao (1977), Engle and Granger (1987), Stock and Watson (1988), and Johansen (1988). We discuss cointegrated VAR models in Section 8.6.

The concept of cointegration is interesting and has attracted a lot of attention in the literature. However, there are difficulties in testing for cointegration in a real application. The main source of difficulties is that cointegration tests overlook the scaling effects of the component series. Interested readers are referred to Cochrane (1988) and Tiao, Tsay, and Wang (1993) for further discussion.

While I have some misgivings on the practical value of cointegration tests, the idea of cointegration is highly relevant in financial study. For example, consider the stock of Finnish Nokia Corporation. Its price on the Helsinki Stock Market must move in unison with the price of its American Depositary Receipts on the New York Stock Exchange; otherwise there exists some arbitrage opportunity for investors. If the stock price has a unit root, then the two price series must be cointegrated. In practice, such a cointegration can exist after adjusting for transaction costs and exchange-rate risk. We discuss issues like this later in Section 8.7.

8.5.1 An Error-Correction Form

Because there are more unit-root nonstationary components than the number of unit roots in a cointegrated system, differencing individual components to achieve stationarity results in overdifferencing. Overdifferencing leads to the problem of unit roots in the MA matrix polynomial, which in turn may encounter difficulties in parameter estimation. If the MA matrix polynomial contains unit roots, the vector time series is said to be noninvertible.

Engle and Granger (1987) discuss an error-correction representation for a cointegrated system that overcomes the difficulty of estimating noninvertible VARMA models. Consider the cointegrated system in Eq. (8.31). Let $\Delta\boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$ be the differenced series. Subtracting $\boldsymbol{x}_{t-1}$ from both sides of the equation, we obtain a model for $\Delta\boldsymbol{x}_t$ as

$$\begin{aligned} \begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} &= \begin{bmatrix} -0.5 & -1.0 \\ -0.25 & -0.5 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ -0.5 \end{bmatrix} [0.5, 1.0] \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \end{aligned}$$

This is a stationary model because both $\Delta\boldsymbol{x}_t$ and $[0.5, 1.0]\boldsymbol{x}_t = y_{2t}$ are unit-root stationary. Because $\boldsymbol{x}_{t-1}$ is used on the right-hand side of the previous equation, the MA matrix polynomial is the same as before and, hence, the model does not encounter the problem of noninvertibility. Such a formulation is referred to as an error-correction model for $\Delta\boldsymbol{x}_t$, and it can be extended to the general cointegrated VARMA model. For a cointegrated VARMA(p, q) model with m cointegrating

factors ($m < k$), an error-correction representation is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \boldsymbol{a}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{a}_{t-j}, \tag{8.33}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $k \times m$ full-rank matrices. The AR coefficient matrices $\boldsymbol{\Phi}_i^*$ are functions of the original coefficient matrices $\boldsymbol{\Phi}_j$. Specifically, we have

$$\begin{aligned} \boldsymbol{\Phi}_j^* &= -\sum_{i=j+1}^{p} \boldsymbol{\Phi}_i, \quad j = 1, \ldots, p-1, \\ \boldsymbol{\alpha}\boldsymbol{\beta}' &= \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_{p-1} + \cdots + \boldsymbol{\Phi}_1 - \boldsymbol{I} = -\boldsymbol{\Phi}(1). \end{aligned} \tag{8.34}$$

These results can be obtained by equating coefficient matrices of the AR matrix polynomials. The time series $\boldsymbol{\beta}'\boldsymbol{x}_t$ is unit-root stationary, and the columns of $\boldsymbol{\beta}$ are the cointegrating vectors of $\boldsymbol{x}_t$.

Existence of the stationary series $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$ in the error-correction representation (8.33) is natural. It can be regarded as a "compensation" term for the overdifferenced system $\Delta\boldsymbol{x}_t$. The stationarity of $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$ can be justified as follows. The theory of unit-root time series shows that the sample correlation coefficient between a unit-root nonstationary series and a stationary series converges to zero as the sample size goes to infinity; see Tsay and Tiao (1990) and the references therein. In an error-correction representation, $\boldsymbol{x}_{t-1}$ is unit-root nonstationary, but $\Delta\boldsymbol{x}_t$ is stationary. Therefore, the only way that $\Delta\boldsymbol{x}_t$ can relate meaningfully to $\boldsymbol{x}_{t-1}$ is through a stationary series $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$.

Remark. Our discussion of cointegration assumes that all unit roots are of multiplicity 1, but the concept can be extended to cases in which the unit roots have different multiplicities. Also, if the number of cointegrating factors m is given, then the error-correction model in Eq. (8.33) can be estimated by likelihood methods. We discuss the simple case of cointegrated VAR models in the next section. Finally, there are many ways to construct an error-correction representation. In fact, one can use any $\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-v}$ for $1 \le v \le p$ in Eq. (8.33) with some modifications to the AR coefficient matrices $\boldsymbol{\Phi}_i^*$. □

8.6 COINTEGRATED VAR MODELS

To better understand cointegration, we focus on VAR models for their simplicity in estimation. Consider a k-dimensional VAR(p) time series $\boldsymbol{x}_t$ with possible time trend so that the model is

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1 \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{x}_{t-p} + \boldsymbol{a}_t, \tag{8.35}$$

where the innovation $\boldsymbol{a}_t$ is assumed to be Gaussian and $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are k-dimensional constant vectors. Write $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$. Recall that if all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ are outside the unit circle, then

$\boldsymbol{x}_t$ is unit-root stationary. In the literature, a unit-root stationary series is said to be an $I(0)$ process; that is, it is not integrated. If $|\boldsymbol{\Phi}(1)| = 0$, then $\boldsymbol{x}_t$ is unit-root nonstationary. For simplicity, we assume that $\boldsymbol{x}_t$ is at most an integrated process of order 1, that is, an $I(1)$ process. This means that $(1 - B)x_{it}$ is unit-root stationary if x_{it} itself is not.

An error-correction model (ECM) for the VAR(p) process $\boldsymbol{x}_t$ is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi}\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t, \tag{8.36}$$

where $\boldsymbol{\Phi}_j^*$ are defined in Eq. (8.34) and $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = -\boldsymbol{\Phi}(1)$. We refer to the term $\boldsymbol{\Pi}\boldsymbol{x}_{t-1}$ of Eq. (8.36) as the *error-correction term*, which plays a key role in cointegration study. Notice that $\boldsymbol{\Phi}_i$ can be recovered from the ECM representation via

$$\begin{aligned} \boldsymbol{\Phi}_1 &= \boldsymbol{I} + \boldsymbol{\Pi} + \boldsymbol{\Phi}_1^*, \\ \boldsymbol{\Phi}_i &= \boldsymbol{\Phi}_i^* - \boldsymbol{\Phi}_{i-1}^*, \quad i = 2, \ldots, p, \end{aligned}$$

where $\boldsymbol{\Phi}_p^* = \boldsymbol{0}$, the zero matrix. Based on the assumption that $\boldsymbol{x}_t$ is at most $I(1)$, $\Delta\boldsymbol{x}_t$ of Eq. (8.36) is an $I(0)$ process.

If $\boldsymbol{x}_t$ contains unit roots, then $|\boldsymbol{\Phi}(1)| = 0$ so that $\boldsymbol{\Pi} = -\boldsymbol{\Phi}(1)$ is singular. Therefore, three cases are of interest in considering the ECM in Eq. (8.36):

1. Rank($\boldsymbol{\Pi}$) = 0. This implies $\boldsymbol{\Pi} = \boldsymbol{0}$ and $\boldsymbol{x}_t$ is not cointegrated. The ECM of Eq. (8.36) reduces to

   $$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

   so that $\Delta\boldsymbol{x}_t$ follows a VAR($p - 1$) model with deterministic trend $\boldsymbol{\mu}_t$.
2. Rank($\boldsymbol{\Pi}$) = k. This implies that $|\boldsymbol{\Phi}(1)| \neq 0$ and $\boldsymbol{x}_t$ contains no unit roots; that is, $\boldsymbol{x}_t$ is $I(0)$. The ECM model is not informative and one studies $\boldsymbol{x}_t$ directly.
3. $0 <$ Rank($\boldsymbol{\Pi}$) = $m < k$. In this case, one can write $\boldsymbol{\Pi}$ as

   $$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}', \tag{8.37}$$

   where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $k \times m$ matrices with Rank($\boldsymbol{\alpha}$) = Rank($\boldsymbol{\beta}$) = m. The ECM of Eq. (8.36) becomes

   $$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t. \tag{8.38}$$

   This means that $\boldsymbol{x}_t$ is cointegrated with m linearly independent cointegrating vectors, $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$, and has $k - m$ unit roots that give $k - m$ common stochastic trends of $\boldsymbol{x}_t$.

If $\boldsymbol{x}_t$ is cointegrated with Rank($\boldsymbol{\Pi}$) = m, then a simple way to obtain a presentation of the $k - m$ common trends is to obtain an orthogonal complement matrix $\boldsymbol{\alpha}_\perp$ of $\boldsymbol{\alpha}$; that is, $\boldsymbol{\alpha}_\perp$ is a $k \times (k - m)$ matrix such that $\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha} = \boldsymbol{0}$, a $(k - m) \times m$ zero matrix, and use $\boldsymbol{y}_t = \boldsymbol{\alpha}_\perp'\boldsymbol{x}_t$. To see this, one can premultiply the ECM by $\boldsymbol{\alpha}_\perp'$ and use $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ to see that there would be no error-correction

term in the resulting equation. Consequently, the $(k-m)$-dimensional series $\boldsymbol{y}_t$ should have $k-m$ unit roots. For illustration, consider the bivariate example of Section 8.5.1. For this special series, $\boldsymbol{\alpha} = (-1, -0.5)'$ and $\boldsymbol{\alpha}_\perp = (1, -2)'$. Therefore, $y_t = (1, -2)\boldsymbol{x}_t = x_{1t} - 2x_{2t}$, which is precisely the unit-root nonstationary series y_{1t} in Eq. (8.32).

Note that the factorization in Eq. (8.37) is not unique, because for any $m \times m$ orthogonal matrix $\boldsymbol{\Omega}$ satisfying $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}$, we have

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{\Omega}\boldsymbol{\Omega}'\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{\Omega})(\boldsymbol{\beta}\boldsymbol{\Omega})' \equiv \boldsymbol{\alpha}_*\boldsymbol{\beta}_*',$$

where both $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}_*$ are also of rank m. Additional constraints are needed to uniquely identify $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. It is common to require that $\boldsymbol{\beta}' = [\boldsymbol{I}_m, \boldsymbol{\beta}_1']$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix and $\boldsymbol{\beta}_1$ is a $(k-m) \times m$ matrix. In practice, this may require reordering of the elements of $\boldsymbol{x}_t$ such that the first m components all have a unit root. The elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ must also satisfy other constraints for the process $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ to be unit-root stationary. For example, consider the case of a bivariate VAR(1) model with one cointegrating vector. Here $k = 2$, $m = 1$, and the ECM is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [1, \beta_1]\boldsymbol{x}_{t-1} + \boldsymbol{a}_t.$$

Premultiplying the prior equation by $\boldsymbol{\beta}'$, using $w_{t-i} = \boldsymbol{\beta}'\boldsymbol{x}_{t-i}$, and moving w_{t-1} to the right-hand side of the equation, we obtain

$$w_t = \boldsymbol{\beta}'\boldsymbol{\mu}_t + (1 + \alpha_1 + \alpha_2\beta_1)w_{t-1} + b_t,$$

where $b_t = \boldsymbol{\beta}'\boldsymbol{a}_t$. This implies that w_t is a stationary AR(1) process. Consequently, α_i and β_1 must satisfy the stationarity constraint $|1 + \alpha_1 + \alpha_2\beta_1| < 1$.

The prior discussion shows that the rank of $\boldsymbol{\Pi}$ in the ECM of Eq. (8.36) is the number of cointegrating vectors. Thus, to test for cointegration, one can examine the rank of $\boldsymbol{\Pi}$. This is the approach taken by Johansen (1988, 1995) and Reinsel and Ahn (1992).

8.6.1 Specification of the Deterministic Function

Similar to the univariate case, the limiting distributions of cointegration tests depend on the deterministic function $\boldsymbol{\mu}_t$. In this subsection, we discuss some specifications of $\boldsymbol{\mu}_t$ that have been proposed in the literature. To understand some of the statements made below, keep in mind that $\boldsymbol{\alpha}_\perp'\boldsymbol{x}_t$ provides a presentation for the common stochastic trends of $\boldsymbol{x}_t$ if it is cointegrated.

1. $\boldsymbol{\mu}_t = \boldsymbol{0}$: In this case, all the component series of $\boldsymbol{x}_t$ are $I(1)$ without drift and the stationary series $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ has mean zero.
2. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{c}_0$, where $\boldsymbol{c}_0$ is an m-dimensional nonzero constant vector. The ECM becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{c}_0) + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

so that the components of $\boldsymbol{x}_t$ are $I(1)$ without drift, but $\boldsymbol{w}_t$ have a nonzero mean $-\boldsymbol{c}_0$. This is referred to as the case of restricted constant.

3. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$, which is nonzero. Here the component series of $\boldsymbol{x}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ and $\boldsymbol{w}_t$ may have a nonzero mean.
4. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{c}_1 t$, where $\boldsymbol{c}_1$ is a nonzero vector. The ECM becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{c}_1 t) + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

so that the components of $\boldsymbol{x}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ and $\boldsymbol{w}_t$ has a linear time trend related to $\boldsymbol{c}_1 t$. This is the case of restricted trend.

5. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, where $\boldsymbol{\mu}_i$ are nonzero. Here both the constant and trend are unrestricted. The components of $\boldsymbol{x}_t$ are $I(1)$ and have a quadratic time trend and $\boldsymbol{w}_t$ have a linear trend.

Obviously, the last case is not common in empirical work. The first case is not common for economic time series but may represent the log price series of some assets. The third case is also useful in modeling asset prices.

8.6.2 Maximum Likelihood Estimation

In this subsection, we briefly outline the maximum likelihood estimation of a cointegrated VAR(p) model. Suppose that the data are $\{\boldsymbol{x}_t | t = 1, \ldots, T\}$. Without loss of generality, we write $\boldsymbol{\mu}_t = \boldsymbol{\mu}\boldsymbol{d}_t$, where $\boldsymbol{d}_t = [1, t]'$, and it is understood that $\boldsymbol{\mu}_t$ depends on the specification of the previous subsection. For a given m, which is the rank of $\boldsymbol{\Pi}$, the ECM model becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t, \tag{8.39}$$

where $t = p+1, \ldots, T$. A key step in the estimation is to concentrate the likelihood function with respect to the deterministic term and the stationary effects. This is done by considering the following two multivariate linear regressions:

$$\Delta \boldsymbol{x}_t = \boldsymbol{\gamma}_0\boldsymbol{d}_t + \boldsymbol{\Omega}_1\Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Omega}_{p-1}\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{u}_t, \tag{8.40}$$

$$\boldsymbol{x}_{t-1} = \boldsymbol{\gamma}_1\boldsymbol{d}_t + \boldsymbol{\Xi}_1\Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Xi}_{p-1}\Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{v}_t. \tag{8.41}$$

Let $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$ be the residuals of Eqs. (8.40) and (8.41), respectively. Define the sample covariance matrices

$$\boldsymbol{S}_{00} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t', \quad \boldsymbol{S}_{01} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{v}}_t', \quad \boldsymbol{S}_{11} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\boldsymbol{v}}_t\hat{\boldsymbol{v}}_t'.$$

Next, compute the eigenvalues and eigenvectors of $\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ with respect to $\boldsymbol{S}_{11}$. This amounts to solving the eigenvalue problem

$$|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}| = 0.$$

Denote the eigenvalue and eigenvector pairs by $(\hat{\lambda}_i, \boldsymbol{e}_i)$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_k$. Here the eigenvectors are normalized so that $\boldsymbol{e}' \boldsymbol{S}_{11} \boldsymbol{e} = \boldsymbol{I}$, where $\boldsymbol{e} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k]$ is the matrix of eigenvectors.

The unnormalized maximum likelihood estimate (MLE) of the cointegrating vector $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$, from which we can obtain a MLE for $\boldsymbol{\beta}$ that satisfies the identifying constraint and normalization condition. Denote the resulting estimate by $\hat{\boldsymbol{\beta}}_c$ with the subscript c signifying constraints. The MLE of other parameters can then be obtained by the multivariate linear regression

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu} \boldsymbol{d}_t + \boldsymbol{\alpha} \hat{\boldsymbol{\beta}}_c' \boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t.$$

The maximized value of the likelihood function based on m cointegrating vectors is

$$L_{\max}^{-2/T} \propto |\boldsymbol{S}_{00}| \prod_{i=1}^{m} (1 - \hat{\lambda}_i).$$

This value is used in the maximum likelihood ratio test for testing $\text{Rank}(\boldsymbol{\Pi}) = m$. Finally, estimates of the orthogonal complements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained using

$$\hat{\boldsymbol{\alpha}}_\perp = \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{11} [\boldsymbol{e}_{m+1}, \ldots, \boldsymbol{e}_k], \quad \hat{\boldsymbol{\beta}}_\perp = \boldsymbol{S}_{11} [\boldsymbol{e}_{m+1}, \ldots, \boldsymbol{e}_k].$$

8.6.3 A Cointegration Test

For a specified deterministic term $\boldsymbol{\mu}_t$, we now discuss the maximum likelihood test for testing the rank of the $\boldsymbol{\Pi}$ matrix in Eq. (8.36). Let $H(m)$ be the null hypothesis that the rank of $\boldsymbol{\Pi}$ is m. For example, under $H(0)$, $\text{Rank}(\boldsymbol{\Pi}) = 0$ so that $\boldsymbol{\Pi} = \boldsymbol{0}$ and there is no cointegration. The hypotheses of interest are

$$H(0) \subset \cdots \subset H(m) \subset \cdots \subset H(k).$$

For testing purpose, the ECM in Eq. (8.39) becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu} \boldsymbol{d}_t + \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

where $t = p + 1, \ldots, T$. Our goal is to test the rank of $\boldsymbol{\Pi}$. Mathematically, the rank of $\boldsymbol{\Pi}$ is the number of nonzero eigenvalues of $\boldsymbol{\Pi}$, which can be obtained if a consistent estimate of $\boldsymbol{\Pi}$ is available. Based on the prior equation, which is in the form of a multivariate linear regression, we see that $\boldsymbol{\Pi}$ is related to the covariance matrix between $\boldsymbol{x}_{t-1}$ and $\Delta \boldsymbol{x}_t$ after adjusting for the effects of $\boldsymbol{d}_t$ and $\Delta \boldsymbol{x}_{t-i}$ for $i = 1, \ldots, p-1$. The necessary adjustments can be achieved by the techniques of multivariate linear regression shown in the previous subsection. Indeed, the adjusted series of $\boldsymbol{x}_{t-1}$ and $\Delta \boldsymbol{x}_t$ are $\hat{\boldsymbol{v}}_t$ and $\hat{\boldsymbol{u}}_t$, respectively. The equation of interest for the cointegration test then becomes

$$\hat{\boldsymbol{u}}_t = \boldsymbol{\Pi} \hat{\boldsymbol{v}}_t + \boldsymbol{a}_t.$$

Under the normality assumption, the likelihood ratio test for testing the rank of $\boldsymbol{\Pi}$ in the prior equation can be done by using the canonical correlation analysis

between $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$. See Johnson and Wichern (1998) for information on canonical correlation analysis. The associated canonical correlations are the *partial canonical correlations* between $\Delta \boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_{t-1}$ because the effects of $\boldsymbol{d}_t$ and $\Delta \boldsymbol{x}_{t-i}$ have been adjusted. The quantities $\{\hat{\lambda}_i\}$ are the squared canonical correlations between $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$.

Consider the hypotheses

$$H_o : \text{Rank}(\boldsymbol{\Pi}) = m \quad \text{versus} \quad H_a : \text{Rank}(\boldsymbol{\Pi}) > m.$$

Johansen (1988) proposes the likelihood ratio (LR) statistic

$$LK_{\text{tr}}(m) = -(T - p) \sum_{i=m+1}^{k} \ln(1 - \hat{\lambda}_i) \tag{8.42}$$

to perform the test. If $\text{Rank}(\boldsymbol{\Pi}) = m$, then $\hat{\lambda}_i$ should be small for $i > m$ and hence $LK_{\text{tr}}(m)$ should be small. This test is referred to as the *trace* cointegration test. Due the presence of unit roots, the asymptotic distribution of $LK_{\text{tr}}(m)$ is not chi-squared, but a function of standard Brownian motions. Thus, critical values of $LK_{\text{tr}}(m)$ must be obtained via simulation.

Johansen (1988) also considers a sequential procedure to determine the number of cointegrating vectors. Specifically, the hypotheses of interest are

$$H_o : \text{Rank}(\boldsymbol{\Pi}) = m \quad \text{versus} \quad H_a : \text{Rank}(\boldsymbol{\Pi}) = m + 1.$$

The LK ratio test statistic, called the maximum eigenvalue statistic, is

$$LK_{\max}(m) = -(T - p) \ln(1 - \hat{\lambda}_{m+1}).$$

Again, critical values of the test statistics are nonstandard and must be evaluated via simulation.

8.6.4 Forecasting of Cointegrated VAR Models

The fitted ECM model can be used to produce forecasts. First, conditioned on the estimated parameters, the ECM equation can be used to produce forecasts of the differenced series $\Delta \boldsymbol{x}_t$. Such forecasts can in turn be used to obtain forecasts of $\boldsymbol{x}_t$. A difference between ECM forecasts and the traditional VAR forecasts is that the ECM approach imposes the cointegration relationships in producing the forecasts.

8.6.5 An Example

To demonstrate the analysis of cointegrated VAR models, we consider two weekly U.S. short-term interest rates. The series are the 3-month Treasury bill (TB) rate and 6-month Treasury bill rate from December 12, 1958 to August 6, 2004 for 2383 observations. The TB rates are from the secondary market and obtained from the Federal Reserve Bank of St. Loius. Figure 8.12 shows the time plots of the interest rates. As expected, the two series move closely together.

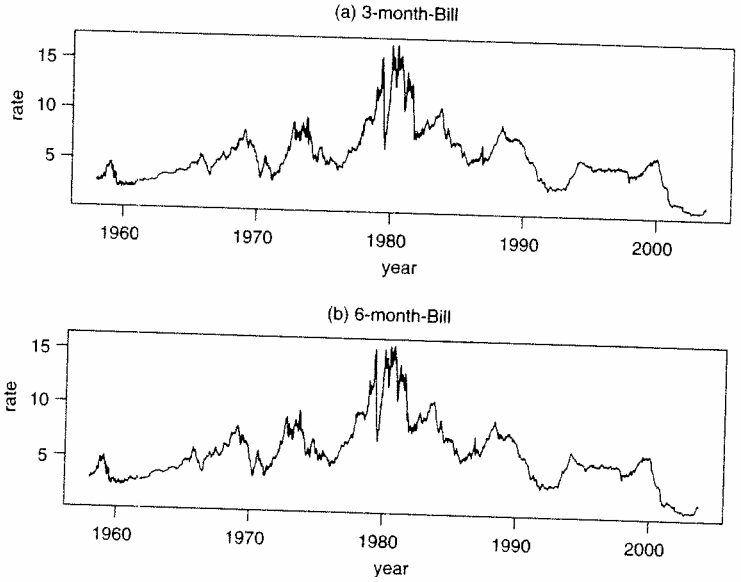

Figure 8.12. Time plots of weekly U.S. interest rate from December 12, 1958 to August 6, 2004. (a) The 3-month Treasury bill rate and (b) the 6-month Treasury bill rate. The rates are from the secondary market.

Our analysis uses the S-Plus software with commands `VAR` for VAR analysis, `coint` for cointegration test, and `VECM` for vector error-correction estimation. Denote the two series by `tb3m` and `tb6m` and define the vector series $\boldsymbol{x}_t = (\text{tb3m}_t, \text{tb6m}_t)'$. The augmented Dickey–Fuller unit-root tests fail to reject the hypothesis of a unit root in the individual series; see Chapter 2. Indeed, the test statistics are -2.34 and -2.33 with p-value about 0.16 for the 3-month and 6-month interest rate when an AR(3) model is used. Thus, we proceed to VAR modeling.

For the bivariate series $\boldsymbol{x}_t$, the BIC criterion selects a VAR(3) model.

```
> x=cbind(tb3m,tb6m)
> y=data.frame(x)
> ord.choice$ar.order
[1] 3
```

To perform a cointegration test, we choose a restricted constant for $\boldsymbol{\mu}_t$ because there is no reason a priori to believe the existence of a drift in the U.S. interest rate. Both Johansen's tests confirm that the two series are cointegrated with one cointegrating vector when a VAR(3) model is entertained.

```
> cointst.rc=coint(x,trend='rc', lags=2)  % lags = p-1.
```



```
> cointst.rc
Call:
coint(Y = x, lags = 2, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests sign. at the 5% level are flagged by ' +'.
Trace tests sign. at the 1% level are flagged by '++'.
Max Eig. tests sign. at the 5% level are flagged by ' *'.
Max Eig. tests sign. at the 1% level are flagged by '**'.

Tests for Cointegration Rank:
        Eigenvalue Trace Stat  95% CV  99% CV
H(0)++** 0.0322     83.2712    19.96   24.60
H(1)     0.0023      5.4936     9.24   12.97

        Max Stat  95% CV  99% CV
H(0)++** 77.7776  15.67   20.20
H(1)      5.4936   9.24   12.97
```

Next, we perform the maximum likelihood estimation of the specified cointegrated VAR(3) model using an ECM presentation. The results are given below:

```
> vecm.fit=VECM(cointst.rc)
> summary(vecm.fit)
Call:
VECM(test = cointst.rc)

Cointegrating Vectors:
             coint.1
              1.0000

      tb6m   -1.0124
 (std.err)    0.0086
  (t.stat) -118.2799

Intercept*    0.2254
 (std.err)    0.0545
  (t.stat)    4.1382

VECM Coefficients:
             tb3m     tb6m
  coint.1 -0.0949 -0.0211
(std.err)  0.0199  0.0179
 (t.stat) -4.7590 -1.1775

tb3m.lag1  0.0466 -0.0419
(std.err)  0.0480  0.0432
```

```
 (t.stat)   0.9696 -0.9699

tb6m.lag1   0.2650  0.3164
(std.err)   0.0538  0.0484
 (t.stat)   4.9263  6.5385

tb3m.lag2  -0.2067 -0.0346
(std.err)   0.0481  0.0433
 (t.stat)  -4.2984 -0.8005

tb6m.lag2   0.2547  0.0994
(std.err)   0.0543  0.0488
 (t.stat)   4.6936  2.0356

Regression Diagnostics:
                   tb3m   tb6m
     R-squared 0.1081 0.0913
Adj. R-squared 0.1066 0.0898
  Resid. Scale 0.2009 0.1807

> plot(vecm.fit)
Make a plot selection (or 0 to exit):

1: plot: All
2: plot: Response and Fitted Values
3: plot: Residuals
...
13: plot: PACF of Squared Cointegrating Residuals
Selection:
```

As expected, the output shows that the stationary series is $w_t \approx \text{tb3m}_t - \text{tb6m}_t$ and the mean of w_t is about -0.225. The fitted ECM model is

$$\Delta \boldsymbol{x}_t = \begin{bmatrix} -0.09 \\ -0.02 \end{bmatrix} (w_{t-1} + 0.23) + \begin{bmatrix} 0.05 & 0.27 \\ -0.04 & 0.32 \end{bmatrix} \Delta \boldsymbol{x}_{t-1} + \begin{bmatrix} -0.21 & 0.25 \\ -0.03 & 0.10 \end{bmatrix} \Delta \boldsymbol{x}_{t-2} + \boldsymbol{a}_t,$$

and the estimated standard errors of a_{it} are 0.20 and 0.18, respectively. Adequacy of the fitted ECM model can be examined via various plots. For illustration, Figure 8.13 shows the cointegrating residuals. Some large residuals are shown in the plot, which occurred in the early 1980s when the interest rates were high and volatile.

Finally, we use the fitted ECM model to produce 1-step to 10-step ahead forecasts for both $\Delta \boldsymbol{x}_t$ and $\boldsymbol{x}_t$. The forecast origin is August 6, 2004.

```
> vecm.fst=predict(vecm.fit, n.predict=10)
> summary(vecm.fst)
```


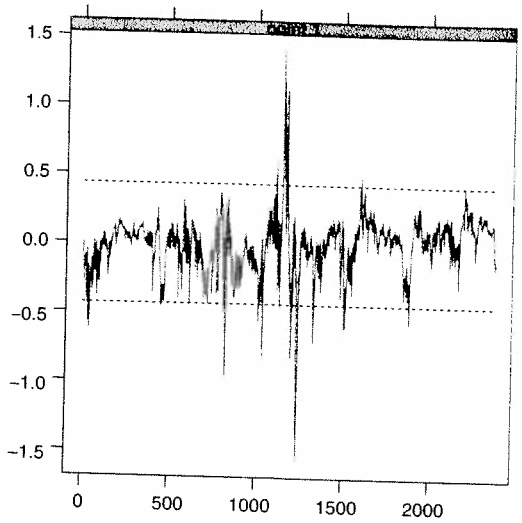

Figure 8.13. Time plot of cointegrating residuals for an ECM fit to the weekly U.S. interest rate series. The data span is from December 12, 1958 to August 6, 2004.

```
Predicted Values with Standard Errors:

                 tb3m    tb6m
 1-step-ahead -0.0378 -0.0642
    (std.err)  0.2009  0.1807
 2-step-ahead -0.0870 -0.0864
    (std.err)  0.3222  0.2927
 ...
10-step-ahead -0.2276 -0.1314
    (std.err)  0.8460  0.8157
> plot(vecm.fst,xold=diff(x),n.old=12)

> vecm.fit.level=VECM(cointst.rc,levels=T)
> vecm.fst.level=predict(vecm.fit.level, n.predict=10)
> summary(vecm.fst.level)

Predicted Values with Standard Errors:
                tb3m   tb6m
 1-step-ahead 1.4501 1.7057
    (std.err) 0.2009 0.1807
 2-step-ahead 1.4420 1.7017
    (std.err) 0.3222 0.2927
 ...
10-step-ahead 1.4722 1.7078
    (std.err) 0.8460 0.8157
> plot(vecm.fst.level, xold=x, n.old=50)
```

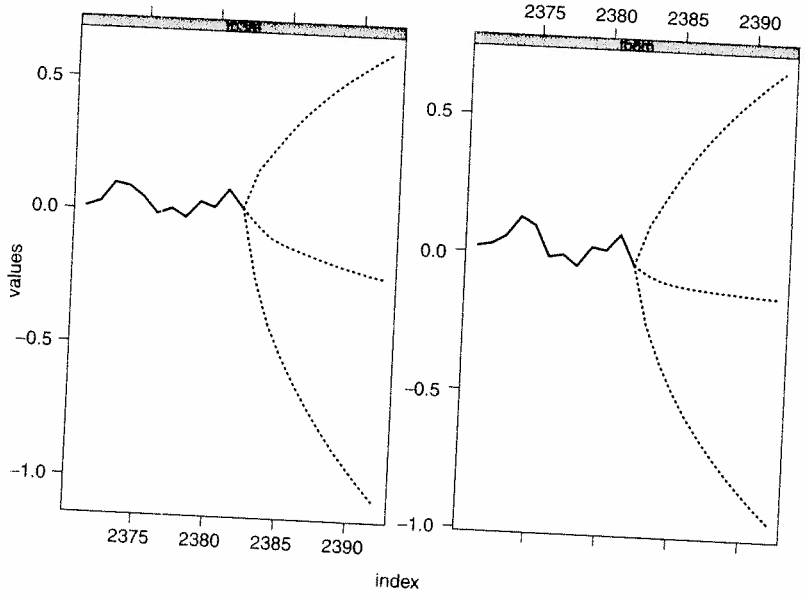

Figure 8.14. Forecasting plots of a fitted ECM model for the weekly U.S. interest rate series. The forecasts are for the differenced series and the forecast origin is August 6, 2004.

The forecasts are shown in Figures 8.14 and 8.15 for the differenced data and the original series, respectively, along with some observed data points. The dashed lines in the plots are pointwise 95% confidence intervals. Because of unit-root nonstationarity, the intervals are wide and not informative.

8.7 THRESHOLD COINTEGRATION AND ARBITRAGE

In this section, we focus on detecting arbitrage opportunities in index trading by using multivariate time series methods. We also demonstrate that simple univariate nonlinear models of Chapter 4 can be extended naturally to the multivariate case in conjunction with the idea of cointegration.

Our study considers the relationship between the price of the S&P 500 index futures and the price of the shares underlying the index on the cash market. Let $f_{t,\ell}$ be the log price of the index futures at time t with maturity ℓ, and let s_t be the log price of the shares underlying the index on the cash market at time t. A version of the *cost-of-carry model* in the finance literature states

$$f_{t,\ell} - s_t = (r_{t,\ell} - q_{t,\ell})(\ell - t) + z_t^*, \tag{8.43}$$

where $r_{t,\ell}$ is the risk-free interest rate, $q_{t,\ell}$ is the dividend yield with respect to the cash price at time t, and $(\ell - t)$ is the time to maturity of the futures contract;

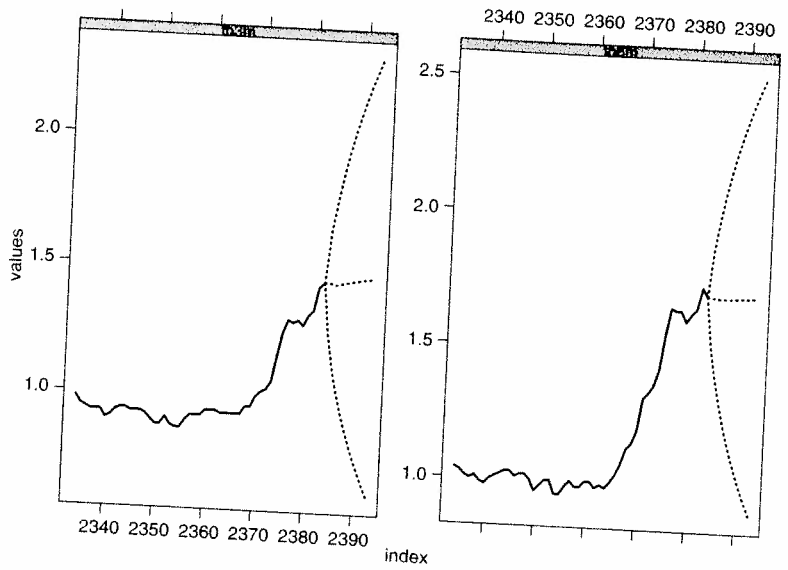

Figure 8.15. Forecasting plots of a fitted ECM model for the weekly U.S. interest rate series. The forecasts are for the interest rates and the forecast origin is August 6, 2004.

see Brenner and Kroner (1995), Dwyer, Locke, and Yu (1996), and the references therein.

The z_t^* process of model (8.43) must be unit-root stationary; otherwise there exist *persistent* arbitrage opportunities. Here an arbitrage trading consists of simultaneously buying (short-selling) the security index and selling (buying) the index futures whenever the log prices diverge by more than the cost of carrying the index over time until maturity of the futures contract. Under the weak stationarity of z_t^*, for arbitrage to be profitable, z_t^* must exceed a certain value in modulus determined by transaction costs and other economic and risk factors.

It is commonly believed that the $f_{t,\ell}$ and s_t series of the S&P 500 index contain a unit root, but Eq. (8.43) indicates that they are cointegrated after adjusting for the effect of interest rate and dividend yield. The cointegrating vector is $(1, -1)$ after the adjustment, and the cointegrated series is z_t^*. Therefore, one should use an error-correction form to model the return series $\boldsymbol{r}_t = (\Delta f_t, \Delta s_t)'$, where $\Delta f_t = f_{t,\ell} - f_{t-1,\ell}$ and $\Delta s_t = s_t - s_{t-1}$, where for ease in notation we drop the maturity time ℓ from the subscript of Δf_t.

8.7.1 Multivariate Threshold Model

In practice, arbitrage tradings affect the dynamics of the market, and hence the model for $\boldsymbol{r}_t$ may vary over time depending on the presence or absence of arbitrage

tradings. Consequently, the prior discussions lead naturally to the model

$$\boldsymbol{r}_t = \begin{cases} \boldsymbol{c}_1 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(1)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_1 z_{t-1} + \boldsymbol{a}_t^{(1)} & \text{if } z_{t-1} \le \gamma_1, \\ \boldsymbol{c}_2 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(2)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_2 z_{t-1} + \boldsymbol{a}_t^{(2)} & \text{if } \gamma_1 < z_{t-1} \le \gamma_2, \\ \boldsymbol{c}_3 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(3)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_3 z_{t-1} + \boldsymbol{a}_t^{(3)} & \text{if } \gamma_2 < z_{t-1}, \end{cases} \tag{8.44}$$

where $z_t = 100 z_t^*$, $\gamma_1 < 0 < \gamma_2$ are two real numbers, and $\{\boldsymbol{a}_t^{(i)}\}$ are sequences of two-dimensional white noises and are independent of each other. Here we use $z_t = 100 z_t^*$ because the actual value of z_t^* is relatively small.

The model in Eq. (8.44) is referred to as a multivariate threshold model with three regimes. The two real numbers γ_1 and γ_2 are the thresholds and z_{t-1} is the threshold variable. The threshold variable z_{t-1} is supported by the data; see Tsay (1998). In general, one can select z_{t-d} as a threshold variable by considering $d \in \{1, \ldots, d_0\}$, where d_0 is a prespecified positive integer.

Model (8.44) is a generalization of the threshold autoregressive model of Chapter 4. It is also a generalization of the error-correlation model of Eq. (8.33). As mentioned earlier, an arbitrage trading is profitable only when z_t^* or, equivalently, z_t is large in modulus. Therefore, arbitrage tradings only occurred in regimes 1 and 3 of model (8.44). As such, the dynamic relationship between $f_{t,\ell}$ and s_t in regime 2 is determined mainly by the normal market force, and hence the two series behave more or less like a random walk. In other words, the two log prices in the middle regime should be free from arbitrage effects and, hence, free from the cointegration constraint. From an econometric viewpoint, this means that the estimate of $\boldsymbol{\beta}_2$ in the middle regime should be insignificant.

In summary, we expect that the cointegration effects between the log price of the futures and the log price of security index on the cash market are significant in regimes 1 and 3, but insignificant in regime 2. This phenomenon is referred to as a *threshold cointegration*; see Balke and Fomby (1997).

8.7.2 The Data

The data used in this case study are the intraday transaction data of the S&P 500 index in May 1993 and its June futures contract traded at the Chicago Mercantile Exchange; see Forbes, Kalb, and Kofman (1999), who used the data to construct a minute-by-minute bivariate price series with 7060 observations. To avoid the undue influence of unusual returns, I replaced 10 extreme values (5 on each side) by the simple average of their two nearest neighbors. This step does not affect the qualitative conclusion of the analysis but may affect the conditional heteroscedasticity in the data. For simplicity, we do not consider conditional heteroscedasticity in the study. Figure 8.16 shows the time plots of the log returns of the index futures and cash prices and the associated threshold variable $z_t = 100 z_t^*$ of model (8.43).

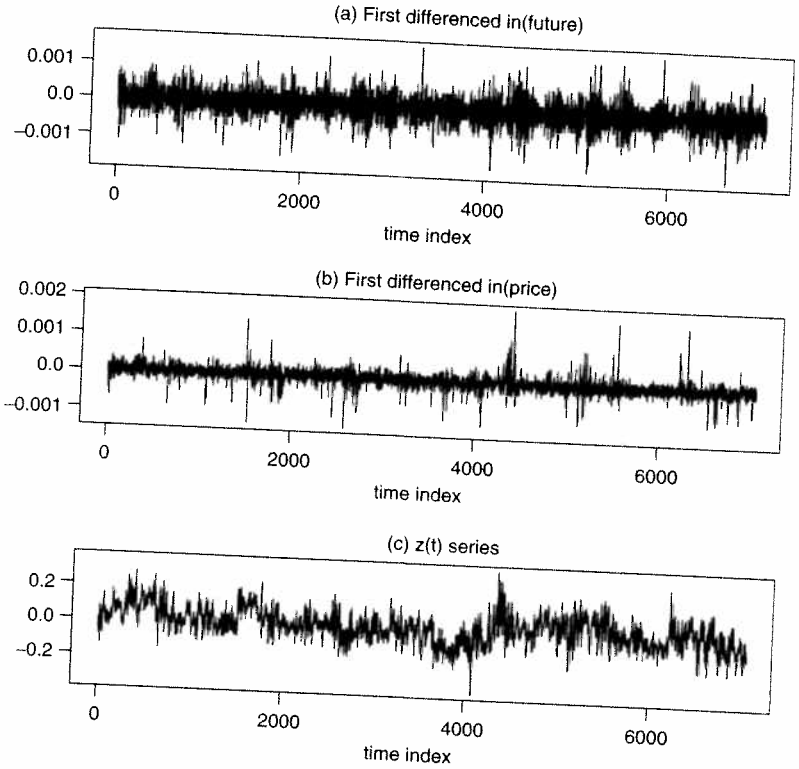

Figure 8.16. Time plots of 1-minute log returns of the S&P 500 index futures and cash prices and the associated threshold variable in May 1993: (a) log returns of the index futures, (b) log returns of the index cash prices, and (c) the z_t series.

8.7.3 Estimation

A formal specification of the multivariate threshold model in Eq. (8.44) includes selecting the threshold variable, determining the number of regimes, and choosing the order p for each regime. Interested readers are referred to Tsay (1998) and Forbes, Kalb, and Kofman (1999). The thresholds γ_1 and γ_2 can be estimated by using some information criteria (e.g., the Akaike information criterion [AIC] or the sum of squares of residuals). Assuming $p = 8$, $d \in \{1, 2, 3, 4\}$, $\gamma_1 \in [-0.15, -0.02]$, and $\gamma_2 \in [0.025, 0.145]$, and using a grid search method with 300 points on each of the two intervals, the AIC selects z_{t-1} as the threshold variable with thresholds $\hat{\gamma}_1 = -0.0226$ and $\hat{\gamma}_2 = 0.0377$. Details of the parameter estimates are given in Table 8.8.

From Table 8.8, we make the following observations. First, the t-ratios of $\widehat{\boldsymbol{\beta}}_2$ in the middle regime show that, as expected, the estimates are insignificant at the 5% level, confirming that there is no cointegration between the two log prices in

Table 8.8. Least Squares Estimates and Their t-Ratios of the Multivariate Threshold Model in Eq. (8.44) for the S&P 500 Index Data in May 1993[a]

	Regime 1		Regime 2		Regime 3	
	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t
ϕ_0	0.00002	0.00005	0.00000	0.00000	−0.00001	−0.00005
t	(1.47)	(7.64)	(−0.07)	(0.53)	(−0.74)	(−6.37)
Δf_{t-1}	−0.08468	0.07098	−0.03861	0.04037	−0.04102	0.02305
t	(−3.83)	(6.15)	(−1.53)	(3.98)	(−1.72)	(1.96)
Δf_{t-2}	−0.00450	0.15899	0.04478	0.08621	−0.02069	0.09898
t	(−0.20)	(13.36)	(1.85)	(8.88)	(−0.87)	(8.45)
Δf_{t-3}	0.02274	0.11911	0.07251	0.09752	0.00365	0.08455
t	(0.95)	(9.53)	(3.08)	(10.32)	(0.15)	(7.02)
Δf_{t-4}	0.02429	0.08141	0.01418	0.06827	−0.02759	0.07699
t	(0.99)	(6.35)	(0.60)	(7.24)	(−1.13)	(6.37)
Δf_{t-5}	0.00340	0.08936	0.01185	0.04831	−0.00638	0.05004
t	(0.14)	(7.10)	(0.51)	(5.13)	(−0.26)	(4.07)
Δf_{t-6}	0.00098	0.07291	0.01251	0.03580	−0.03941	0.02615
t	(0.04)	(5.64)	(0.54)	(3.84)	(−1.62)	(2.18)
Δf_{t-7}	−0.00372	0.05201	0.02989	0.04837	−0.02031	0.02293
t	(−0.15)	(4.01)	(1.34)	(5.42)	(−0.85)	(1.95)
Δf_{t-8}	0.00043	0.00954	0.01812	0.02196	−0.04422	0.00462
t	(0.02)	(0.76)	(0.85)	(2.57)	(−1.90)	(0.40)
Δs_{t-1}	−0.08419	0.00264	−0.07618	−0.05633	0.06664	0.11143
t	(−2.01)	(0.12)	(−1.70)	(−3.14)	(1.49)	(5.05)
Δs_{t-2}	−0.05103	0.00256	−0.10920	−0.01521	0.04099	−0.01179
t	(−1.18)	(0.11)	(−2.59)	(−0.90)	(0.92)	(−0.53)
Δs_{t-3}	0.07275	−0.03631	−0.00504	0.01174	−0.01948	−0.01829
t	(1.65)	(−1.58)	(−0.12)	(0.71)	(−0.44)	(−0.84)
Δs_{t-4}	0.04706	0.01438	0.02751	0.01490	0.01646	0.00367
t	(1.03)	(0.60)	(0.71)	(0.96)	(0.37)	(0.17)
Δs_{t-5}	0.08118	0.02111	0.03943	0.02330	−0.03430	−0.00462
t	(1.77)	(0.88)	(0.97)	(1.43)	(−0.83)	(−0.23)
Δs_{t-6}	0.04390	0.04569	0.01690	0.01919	0.06084	−0.00392
t	(0.96)	(1.92)	(0.44)	(1.25)	(1.45)	(−0.19)
Δs_{t-7}	−0.03033	0.02051	−0.08647	0.00270	−0.00491	0.03597
t	(−0.70)	(0.91)	(−2.09)	(0.16)	(−0.13)	(1.90)
Δs_{t-8}	−0.02920	0.03018	0.01887	−0.00213	0.00030	0.02171
t	(−0.68)	(1.34)	(0.49)	(−0.14)	(0.01)	(1.14)
z_{t-1}	0.00024	0.00097	−0.00010	0.00012	0.00025	0.00086
t	(1.34)	(10.47)	(−0.30)	(0.86)	(1.41)	(9.75)

[a]The number of data points for the three regimes are 2234, 2410, and 2408, respectively.

the absence of arbitrage opportunities. Second, Δf_t depends negatively on Δf_{t-1} in all three regimes. This is in agreement with the bid–ask bounce discussed in Chapter 5. Third, past log returns of the index futures seem to be more informative than the past log returns of the cash prices because there are more significant t-ratios in Δf_{t-i} than in Δs_{t-i}. This is reasonable because futures series are in general more liquid. For more information on index arbitrage, see Dwyer, Locke, and Yu (1996).

APPENDIX A: REVIEW OF VECTORS AND MATRICES

In this appendix, we briefly review some algebra and properties of vectors and matrices. No proofs are given as they can be found in standard textbooks on matrices (e.g., Graybill, 1969).

An $m \times n$ real-valued matrix is an m by n array of real numbers. For example,

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 3 & 4 \end{bmatrix}$$

is a 2×3 matrix. This matrix has two rows and three columns. In general, an $m \times n$ matrix is written as

$$\boldsymbol{A} \equiv [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{mn} \end{bmatrix}. \tag{8.45}$$

The positive integers m and n are the *row dimension* and *column dimension* of $\boldsymbol{A}$. The real number a_{ij} is referred to as the (i, j)th element of $\boldsymbol{A}$. In particular, the elements a_{ii} are the *diagonal elements* of the matrix.

An $m \times 1$ matrix forms an m-dimensional column vector, and a $1 \times n$ matrix is an n-dimensional row vector. In the literature, a vector is often meant to be a column vector. If $m = n$, then the matrix is a square matrix. If $a_{ij} = 0$ for $i \neq j$ and $m = n$, then the matrix $\boldsymbol{A}$ is a *diagonal matrix*. If $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = 1$ for all i, then $\boldsymbol{A}$ is the $m \times m$ *identity matrix*, which is commonly denoted by $\boldsymbol{I}_m$ or simply $\boldsymbol{I}$ if the dimension is clear.

The $n \times m$ matrix

$$\boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m-1,1} & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m-1,2} & a_{m2} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{m-1,n} & a_{mn} \end{bmatrix}$$

is the *transpose* of the matrix $\boldsymbol{A}$. For example,

$$\begin{bmatrix} 2 & -1 \\ 5 & 3 \\ 8 & 4 \end{bmatrix} \quad \text{is the transpose of} \quad \begin{bmatrix} 2 & 5 & 8 \\ -1 & 3 & 4 \end{bmatrix}.$$

We use the notation $\boldsymbol{A}' = [a'_{ij}]$ to denote the transpose of $\boldsymbol{A}$. From the definition, $a'_{ij} = a_{ji}$ and $(\boldsymbol{A}')' = \boldsymbol{A}$. If $\boldsymbol{A}' = \boldsymbol{A}$, then $\boldsymbol{A}$ is a *symmetric matrix*.

Basic Operations

Suppose that $\boldsymbol{A} = [a_{ij}]_{m\times n}$ and $\boldsymbol{C} = [c_{ij}]_{p\times q}$ are two matrices with dimensions given in the subscript. Let b be a real number. Some basic matrix operations are defined next:

- Addition: $\boldsymbol{A} + \boldsymbol{C} = [a_{ij} + c_{ij}]_{m\times n}$ if $m = p$ and $n = q$.
- Subtraction: $\boldsymbol{A} - \boldsymbol{C} = [a_{ij} - c_{ij}]_{m\times n}$ if $m = p$ and $n = q$.
- Scalar multiplication: $b\boldsymbol{A} = [ba_{ij}]_{m\times n}$.
- Multiplication: $\boldsymbol{AC} = [\sum_{v=1}^{n} a_{iv}c_{vj}]_{m\times q}$ provided that $n = p$.

When the dimensions of matrices satisfy the condition for multiplication to take place, the two matrices are said to be *conformable*. An example of matrix multiplication is

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & -4 \end{bmatrix} = \begin{bmatrix} 2\cdot 1 - 1\cdot 1 & 2\cdot 2 + 1\cdot 2 & 2\cdot 3 - 1\cdot 4 \\ 1\cdot 1 - 1\cdot 1 & 1\cdot 2 + 1\cdot 2 & 1\cdot 3 - 1\cdot 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 & 2 \\ 0 & 4 & -1 \end{bmatrix}.$$

Important rules of matrix operations include (a) $(\boldsymbol{AC})' = \boldsymbol{C}'\boldsymbol{A}'$ and (b) $\boldsymbol{AC} \neq \boldsymbol{CA}$ in general.

Inverse, Trace, Eigenvalue, and Eigenvector

A square matrix $\boldsymbol{A}_{m\times m}$ is *nonsingular* or *invertible* if there exists a unique matrix $\boldsymbol{C}_{m\times m}$ such that $\boldsymbol{AC} = \boldsymbol{CA} = \boldsymbol{I}_m$, the $m \times m$ identity matrix. In this case, $\boldsymbol{C}$ is called the *inverse* matrix of $\boldsymbol{A}$ and is denoted by $\boldsymbol{C} = \boldsymbol{A}^{-1}$.

The trace of $\boldsymbol{A}_{m\times m}$ is the sum of its diagonal elements (i.e., $tr(\boldsymbol{A}) = \sum_{i=1}^{m} a_{ii}$). It is easy to see that (a) $tr(\boldsymbol{A} + \boldsymbol{C}) = tr(\boldsymbol{A}) + tr(\boldsymbol{C})$, (b) $tr(\boldsymbol{A}) = tr(\boldsymbol{A}')$, and (c) $tr(\boldsymbol{AC}) = tr(\boldsymbol{CA})$ provided that the two matrices are conformable.

A number λ and an $m \times 1$ vector $\boldsymbol{b}$, possibly complex-valued, are a right *eigenvalue* and *eigenvector* pair of the matrix $\boldsymbol{A}$ if $\boldsymbol{Ab} = \lambda\boldsymbol{b}$. There are m possible eigenvalues for the matrix $\boldsymbol{A}$. For a real-valued matrix $\boldsymbol{A}$, complex eigenvalues occur in conjugated pairs. The matrix $\boldsymbol{A}$ is nonsingular if and only if all of its eigenvalues are nonzero. Denote the eigenvalues by $\{\lambda_i | i = 1, \ldots, m\}$: we have $tr(\boldsymbol{A}) = \sum_{i=1}^{m} \lambda_i$. In addition, the *determinant* of the matrix $\boldsymbol{A}$ can be defined as $|\boldsymbol{A}| = \prod_{i=1}^{m} \lambda_i$. For a general definition of determinant of a matrix, see a standard textbook on matrices (e.g., Graybill, 1969).

Finally, the rank of the matrix $\boldsymbol{A}_{m\times n}$ is the number of nonzero eigenvalues of the symmetric matrix $\boldsymbol{AA}'$. Also, for a nonsingular matrix $\boldsymbol{A}$, $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$.

Positive-Definite Matrix

A square matrix $\boldsymbol{A}$ $(m \times m)$ is a *positive-definite* matrix if (a) $\boldsymbol{A}$ is symmetric, and (b) all eigenvalues of $\boldsymbol{A}$ are positive. Alternatively, $\boldsymbol{A}$ is a positive-definite matrix if for any nonzero m-dimensional vector $\boldsymbol{b}$, we have $\boldsymbol{b}'\boldsymbol{Ab} > 0$.

Useful properties of a positive-definite matrix $\boldsymbol{A}$ include (a) all eigenvalues of $\boldsymbol{A}$ are real and positive, and (b) the matrix can be decomposed as

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}',$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of all eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an $m \times m$ matrix consisting of the m right eigenvectors of $\boldsymbol{A}$. It is common to write the eigenvalues as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and the eigenvectors as $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ such that $\boldsymbol{Ae}_i = \lambda_i\boldsymbol{e}_i$ and $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$. In addition, these eigenvectors are orthogonal to each other—namely, $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$ if $i \neq j$—if the eigenvalues are distinct. The matrix $\boldsymbol{P}$ is an *orthogonal* matrix and the decomposition is referred to as the *spectral decomposition* of the matrix $\boldsymbol{A}$. Consider, for example, the simple 2×2 matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

which is positive definite. Simple calculations show that

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Therefore, 3 and 1 are eigenvalues of $\boldsymbol{\Sigma}$ with normalized eigenvectors $(1/\sqrt{2}, 1/\sqrt{2})'$ and $(1/\sqrt{2}, -1/\sqrt{2})'$, respectively. It is easy to verify that the spectral decomposition holds—that is,

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

For a symmetric matrix $\boldsymbol{A}$, there exists a lower triangular matrix $\boldsymbol{L}$ with diagonal elements being 1 and a diagonal matrix $\boldsymbol{G}$ such that $\boldsymbol{A} = \boldsymbol{LGL}'$; see Chapter 1 of Strang (1980). If $\boldsymbol{A}$ is positive definite, then the diagonal elements of $\boldsymbol{G}$ are positive. In this case, we have

$$\boldsymbol{A} = \boldsymbol{L}\sqrt{\boldsymbol{G}}\sqrt{\boldsymbol{G}}\boldsymbol{L}' = (\boldsymbol{L}\sqrt{\boldsymbol{G}})(\boldsymbol{L}\sqrt{\boldsymbol{G}})',$$

where $\boldsymbol{L}\sqrt{\boldsymbol{G}}$ is again a lower triangular matrix and the square root is taken element by element. Such a decomposition is called the *Cholesky decomposition* of $\boldsymbol{A}$. This decomposition shows that a positive-definite matrix $\boldsymbol{A}$ can be diagonalized as

$$\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}')^{-1} = \boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{-1})' = \boldsymbol{G}.$$

Since $\boldsymbol{L}$ is a lower triangular matrix with unit diagonal elements, $\boldsymbol{L}^{-1}$ is also a lower triangular matrix with unit diagonal elements. Consider again the prior 2×2 matrix $\boldsymbol{\Sigma}$. It is easy to verify that

$$\boldsymbol{L} = \begin{bmatrix} 1.0 & 0.0 \\ 0.5 & 1.0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 1.5 \end{bmatrix}$$

satisfy $\boldsymbol{\Sigma} = \boldsymbol{LGL}'$. In addition,

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^{-1})' = \boldsymbol{G}.$$

Vectorization and Kronecker Product

Writing an $m \times n$ matrix $\boldsymbol{A}$ in its columns as $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$, we define the stacking operation as $\text{vec}(\boldsymbol{A}) = (\boldsymbol{a}_1', \boldsymbol{a}_2', \ldots, \boldsymbol{a}_m')'$, which is an $mn \times 1$ vector. For two matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{C}_{p\times q}$, the Kronecker product between $\boldsymbol{A}$ and $\boldsymbol{C}$ is

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{bmatrix} a_{11}\boldsymbol{C} & a_{12}\boldsymbol{C} & \cdots & a_{1n}\boldsymbol{C} \\ a_{21}\boldsymbol{C} & a_{22}\boldsymbol{C} & \cdots & a_{2n}\boldsymbol{C} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{C} & a_{m2}\boldsymbol{C} & \cdots & a_{mn}\boldsymbol{C} \end{bmatrix}_{mp\times nq}.$$

For example, assume that

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 4 & -1 & 3 \\ -2 & 5 & 2 \end{bmatrix}.$$

Then $\text{vec}(\boldsymbol{A}) = (2, -1, 1, 3)'$, $\text{vec}(\boldsymbol{C}) = (4, -2, -1, 5, 3, 2)'$, and

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{bmatrix} 8 & -2 & 6 & 4 & -1 & 3 \\ -4 & 10 & 4 & -2 & 5 & 2 \\ -4 & 1 & -3 & 12 & -3 & 9 \\ 2 & -5 & -2 & -6 & 15 & 6 \end{bmatrix}.$$

Assuming that the dimensions are appropriate, we have the following useful properties for the two operators:

1. $\boldsymbol{A} \otimes \boldsymbol{C} \neq \boldsymbol{C} \otimes \boldsymbol{A}$ in general.
2. $(\boldsymbol{A} \otimes \boldsymbol{C})' = \boldsymbol{A}' \otimes \boldsymbol{C}'$.
3. $\boldsymbol{A} \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D}$.
4. $(\boldsymbol{A} \otimes \boldsymbol{C})(\boldsymbol{F} \otimes \boldsymbol{G}) = (\boldsymbol{AF}) \otimes (\boldsymbol{CG})$.
5. If $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible, then $(\boldsymbol{A} \otimes \boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{C}^{-1}$.
6. For square matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, $tr(\boldsymbol{A} \otimes \boldsymbol{C}) = tr(\boldsymbol{A})tr(\boldsymbol{C})$.
7. $\text{vec}(\boldsymbol{A} + \boldsymbol{C}) = \text{vec}(\boldsymbol{A}) + \text{vec}(\boldsymbol{C})$.
8. $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$.
9. $tr(\boldsymbol{AC}) = \text{vec}(\boldsymbol{C}')'\text{vec}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A}')'\text{vec}(\boldsymbol{C})$.
10. $tr(\boldsymbol{ABC}) = \text{vec}(\boldsymbol{A}')'(\boldsymbol{C}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{B}) = \text{vec}(\boldsymbol{A}')'(\boldsymbol{I} \otimes \boldsymbol{B})\text{vec}(\boldsymbol{C})$
$= \text{vec}(\boldsymbol{B}')'(\boldsymbol{A}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{C}) = \text{vec}(\boldsymbol{B}')'(\boldsymbol{I} \otimes \boldsymbol{C})\text{vec}(\boldsymbol{A})$
$= \text{vec}(\boldsymbol{C}')'(\boldsymbol{B}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{A}) = \text{vec}(\boldsymbol{C}')'(\boldsymbol{I} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$.

In multivariate statistical analysis, we often deal with symmetric matrices. It is therefore convenient to generalize the stacking operation to the *half-stacking* operation, which consists of elements on or below the main diagonal. Specifically, for a symmetric square matrix $\boldsymbol{A} = [a_{ij}]_{k\times k}$, define

$$\text{vech}(\boldsymbol{A}) = (\boldsymbol{a}_{1\cdot}', \boldsymbol{a}_{2*}', \ldots, \boldsymbol{a}_{k*}')',$$

where $\boldsymbol{a}_{1\cdot}$ is the first column of $\boldsymbol{A}$, and $\boldsymbol{a}_{i*} = (a_{ii}, a_{i+1,i}, \ldots, a_{ki})'$ is a $(k - i + 1)$-dimensional vector. The dimension of $\text{vech}(\boldsymbol{A})$ is $k(k + 1)/2$. For example, suppose that $k = 3$. Then we have $\text{vech}(\boldsymbol{A}) = (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})'$, which is a six-dimensional vector.

APPENDIX B: MULTIVARIATE NORMAL DISTRIBUTIONS

A k-dimensional random vector $\boldsymbol{x} = (x_1, \ldots, x_k)'$ follows a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and positive-definite covariance matrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$ if its probability density function (pdf) is

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp[-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})]. \tag{8.46}$$

We use the notation $\boldsymbol{x} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote that $\boldsymbol{x}$ follows such a distribution. This normal distribution plays an important role in multivariate statistical analysis and it has several nice properties. Here we consider only those properties that are relevant to our study. Interested readers are referred to Johnson and Wichern (1998) for details.

To gain insight into multivariate normal distributions, consider the bivariate case (i.e., $k = 2$). In this case, we have

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}.$$

Using the correlation coefficient $\rho = \sigma_{12}/(\sigma_1\sigma_2)$, where $\sigma_i = \sqrt{\sigma_{ii}}$ is the standard deviation of x_i, we have $\sigma_{12} = \rho\sqrt{\sigma_{11}\sigma_{22}}$ and $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22}(1 - \rho^2)$. The pdf of $\boldsymbol{x}$ then becomes

$$f(x_1, x_2|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)}[Q(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})]\right),$$

where

$$Q(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right).$$

Chapter 4 of Johnson and Wichern (1998) contains some plots of this pdf function.

Let $\boldsymbol{c} = (c_1, \ldots, c_k)'$ be a nonzero k-dimensional vector. Partition the random vector as $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, where $\boldsymbol{x}_1 = (x_1, \ldots, x_p)'$ and $\boldsymbol{x}_2 = (x_{p+1}, \ldots, x_k)'$ with $1 \leq p < k$. Also partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly as

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right).$$

Some properties of $\boldsymbol{x}$ are as follows:

1. $\boldsymbol{c}'\boldsymbol{x} \sim N(\boldsymbol{c}'\boldsymbol{\mu}, \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c})$. That is, any nonzero linear combination of $\boldsymbol{x}$ is univariate normal. The inverse of this property also holds. Specifically, if $\boldsymbol{c}'\boldsymbol{x}$ is univariate normal for any nonzero vector $\boldsymbol{c}$, then $\boldsymbol{x}$ is multivariate normal.

2. The marginal distribution of $\boldsymbol{x}_i$ is normal. In fact, $\boldsymbol{x}_i \sim N_{k_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$ for $i = 1$ and 2, where $k_1 = p$ and $k_2 = k - p$.
3. $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ if and only if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent.
4. The random variable $y = (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ follows a chi-squared distribution with m degrees of freedom.
5. The conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2 = \boldsymbol{b}$ is also normally distributed as

$$(\boldsymbol{x}_1 | \boldsymbol{x}_2 = \boldsymbol{b}) \sim N_p[\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}].$$

The last property is useful in many scientific areas. For instance, it forms the basis for time series forecasting under the normality assumption and for recursive least squares estimation.

APPENDIX C: SOME SCA COMMANDS

The following SCA commands are used in the analysis of Example 8.6.

```
input x1,x2. file 'm-gsln3-5301.txt'  % Load data
 --
r1=ln(x1)  % Take log transformation
 --
r2=ln(x2)
 --
miden r1,r2. no ccm. arfits 1 to 8.
 --  % Denote the model by v21.
mtsm v21. series r1,r2.  @
model (i-p1*b-p2*b**2)series=c+(i-t1*b)noise.
 --
mestim v21.    % Initial estimation
 --
p1(2,1)=0      % Set zero constraints
 --
cp1(2,1)=1
 --
p2(2,1)=0
 --
cp2(2,1)=1
 --
p2(2,2)=0
 --
cp2(2,2)=1
 --
t1(2,1)=0
 --
ct1(2,1)=1
 -- % Refine estimation and store residuals
mestim v21. method exact. hold resi(res1,res2)
 --
miden res1,res2.
```


EXERCISES

8.1. Consider the monthly log stock returns, in percentages and including dividends, of Merck & Company, Johnson & Johnson, General Electric, General Motors, Ford Motor Company, and value-weighted index from January 1960 to December 1999; see the file `m-mrk2vw.txt`, which has six columns in the order listed before.

(a) Compute the sample mean, covariance matrix, and correlation matrix of the data.

(b) Test the hypothesis $H_o : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_6 = \mathbf{0}$, where $\boldsymbol{\rho}_i$ is the lag-i cross-correlation matrix of the data. Draw conclusions based on the 5% significance level.

(c) Is there any lead–lag relationship among the six return series?

8.2. The Federal Reserve Bank of St. Louis publishes selected interest rates and U.S. financial data on its Web site: `http://research.stlouisfed.org/fred2/`. Consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2000 for 571 observations; see the file `m-gsln10.txt`. The rates are in percentages.

(a) Let $c_t = r_t - r_{t-1}$ be the change series of the monthly interest rate r_t. Build a bivariate autoregressive model for the two change series. Discuss the implications of the model. Transform the model into a structural form.

(b) Build a bivariate moving-average model for the two change series. Discuss the implications of the model and compare it with the bivariate AR model built earlier.

8.3. Again consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2000. Consider the log series of the data and build a VARMA model for the series. Discuss the implications of the model obtained.

8.4. Again consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2000. Are the two interest rate series threshold-cointegrated? Use the interest spread $s_t = r_{10,t} - r_{1,t}$ as the threshold variable, where r_{it} is the i-year Treasury constant maturity rate. If they are threshold-cointegrated, build a multivariate threshold model for the two series.

8.5. The bivariate AR(4) model $\boldsymbol{x}_t - \boldsymbol{\Phi}_4\boldsymbol{x}_{t-4} = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$ is a special seasonal model with periodicity 4, where $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Such a seasonal model may be useful in studying quarterly earnings of a company. (a) Assume that $\boldsymbol{x}_t$ is weakly stationary. Derive the mean vector

and covariance matrix of $\boldsymbol{x}_t$. (b) Derive the necessary and sufficient condition of weak stationarity for $\boldsymbol{x}_t$. (c) Show that $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_4 \boldsymbol{\Gamma}_{\ell-4}$ for $\ell > 0$, where $\boldsymbol{\Gamma}_\ell$ is the lag-ℓ autocovariance matrix of $\boldsymbol{x}_t$.

8.6. The bivariate MA(4) model $\boldsymbol{x}_t = \boldsymbol{a}_t - \boldsymbol{\Theta}_4 \boldsymbol{a}_{t-4}$ is another seasonal model with periodicity 4, where $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Derive the covariance matrices $\boldsymbol{\Gamma}_\ell$ of $\boldsymbol{x}_t$ for $\ell = 0, \ldots, 5$.

8.7. Consider the monthly U.S. 1-year and 3-year Treasury constant maturity rates from April 1953 to March 2004. The data can be obtained from the Federal Reserve Bank of St. Louis or from the file `m-gsln3-5304.txt` (1-year, 3-year, dates). See also Example 8.6 that uses a shorter data span. Here we use the interest rates directly without the log transformation and define $\boldsymbol{x}_t = (x_{1t}, x_{2t})'$, where x_{1t} is the 1-year maturity rate and x_{2t} is the 3-year maturity rate.

(a) Identify a VAR model for the bivariate interest rate series. Write down the fitted model.

(b) Compute the impulse response functions of the fitted VAR model. It suffices to use the first 6 lags.

(c) Use the fitted VAR model to produce 1-step to 12-step ahead forecasts of the interest rates, assuming that the forecast origin is March 2004.

(d) Are the two interest rate series cointegrated, when a restricted constant term is used? Use 5% significance level to perform the test.

(e) If the series are cointegrated, build an ECM for the series. Write down the fitted model.

(f) Use the fitted ECM to produce 1-step to 12-step ahead forecasts of the interest rates, assuming that the forecast origin is March 2004.

(g) Compare the forecasts produced by the VAR model and the ECM.

REFERENCES

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review* **38**: 627–645.

Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**: 355–366.

Brenner, R. J. and Kroner, K. F. (1995). Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. *Journal of Financial and Quantitative Analysis* **30**: 23–42.

Cochrane, J. H. (1988). How big is the random walk in the GNP? *Journal of Political Economy* **96**: 893–920.

Dwyer, G. P. Jr., Locke, P., and Yu, W. (1996). Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash. *Review of Financial Studies* **9**: 301–332.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction representation, estimation and testing. *Econometrica* **55**: 251–276.

Forbes, C. S., Kalb, G. R. J., and Kofman, P. (1999). Bayesian arbitrage threshold analysis. *Journal of Business & Economic Statistics* **17**: 364–372.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, Hoboken, NJ.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth, Belmont, CA.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society Series B* **41**: 190–195.

Hillmer, S. C. and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association* **74**: 652–660.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**: 602–608.

Hosking, J. R. M. (1981). Lagrange-multiplier tests of multivariate time series models. *Journal of the Royal Statistical Society Series B* **43**: 219–230.

Johansen, S. (1988). Statistical analysis of co-integration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.

Johansen, S. (1995). *Likelihood Based Inference in Cointegrated Vector Error Correction Models*. Oxford University Press, Oxford, UK.

Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th edition. Prentice Hall, Upper Saddle River, NJ.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society Series B* **43**: 231–239.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.

Reinsel, G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis* **13**: 353–375.

Stock, J. H. and Watson, M.W. (1988). Testing for common trends. *Journal of the American Statistical Association* **83**: 1097–1107.

Strang, G. (1980). *Linear Algebra and Its Applications*, 2nd edition. Harcourt Brace Jovanovich, Chicago.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association* **76**: 802–816.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series (with discussions). *Journal of the Royal Statistical Society Series B* **51**: 157–213.

Tiao, G. C., Tsay, R. S., and Wang, T. (1993). Usefulness of linear transformations in multivariate time series analysis. *Empirical Economics* **18**: 567–593.

Tsay, R. S. (1991). Two canonical forms for vector ARMA processes. *Statistica Sinica* **1**: 247–269.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1202.

Tsay, R. S., and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics* **18**: 220–250.

Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer-Verlag, New York.

CHAPTER 9

Principal Component Analysis and Factor Models

Most financial portfolios consist of multiple assets, and their returns depend concurrently and dynamically on many economic and financial variables. Therefore, it is important to use proper multivariate statistical analyses to study the behavior and properties of portfolio returns. However, as demonstrated in the previous chapter, analysis of multiple asset returns often requires high-dimensional statistical models that are complicated and hard to apply. To simplify the task of modeling multiple returns, we discuss in this chapter some dimension reduction methods to search for the underlying structure of the assets. *Principal component analysis* (PCA) is perhaps the most commonly used statistical method in dimension reduction, and we start our discussion with the method. In practice, observed return series often exhibit similar characteristics leading to the belief that they might be driven by some common sources, often referred to as common factors. To study the common pattern in asset returns and to simplify portfolio analysis, various factor models have been proposed in the literature to analyze multiple asset returns. The second goal of this chapter is to introduce some useful factor models and demonstrate their applications in finance.

Three types of factor models are available for studying asset returns; see Connor (1995) and Campbell, Lo, and MacKinlay (1997). The first type is the *macroeconomic factor models* that use macroeconomic variables such as growth rate of GDP, interest rates, inflation rate, and unemployment numbers to describe the common behavior of asset returns. Here the factors are observable and the model can be estimated via linear regression methods. The second type is the *fundamental factor models* that use firm or asset specific attributes such as firm size, book and market values, and industrial classification to construct common factors. The third type is the *statistical factor models* that treat the common factors as unobservable or latent variables to be estimated from the returns series. In this chapter, we discuss all three types of factor models and their applications in finance. Principal component

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

CHAPTER 11

State-Space Models and Kalman Filter

The state-space model provides a flexible approach to time series analysis, especially for simplifying maximum likelihood estimation and handling missing values. In this chapter, we discuss the relationship between the state-space model and the ARIMA model, the Kalman filter algorithm, various smoothing methods, and some applications. We begin with a simple model that shows the basic ideas of the state-space approach to time series analysis before introducing the general state-space model. For demonstrations, we use the model to analyze realized volatility series of asset returns, the time-varying coefficient market models, and the quarterly earnings per share of a company.

There are many books on statistical analysis using the state-space model. Durbin and Koopman (2001) provide a recent treatment of the approach, Kim and Nelson (1999) focus on economic applications and regime switching, and Anderson and Moore (1979) give a nice summary of theory and applications of the approach for engineering and optimal control. Many time series textbooks include the Kalman filter and state-space model. For example, Chan (2002), Shumway and Stoffer (2000), Hamilton (1994), and Harvey (1993) all have chapters on the topic. West and Harrison (1997) provide a Bayesian treatment with emphasis on forecasting, and Kitagawa and Gersch (1996) use a smoothing prior approach.

The derivation of Kalman filter and smoothing algorithms necessarily involves heavy notation. Therefore, Section 11.4 could be dry for readers who are interested mainly in the concept and applications of state-space models and can be skipped on the first read.

11.1 LOCAL TREND MODEL

Consider the univariate time series y_t satisfying

$$y_t = \mu_t + e_t, \quad e_t \sim N(0, \sigma_e^2), \tag{11.1}$$

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \tag{11.2}$$

where $\{e_t\}$ and $\{\eta_t\}$ are two independent Gaussian white noise series and $t = 1, \ldots, T$. The initial value μ_1 is either given or follows a known distribution, and it is independent of $\{e_t\}$ and $\{\eta_t\}$ for $t > 0$. Here μ_t is a pure *random walk* of Chapter 2 with initial value μ_1 and y_t is an observed version of μ_t with added noise a_t. In the literature, μ_t is referred to as the *trend* of the series, which is not directly observable, and y_t is the observed data with observational noise e_t. The dynamic dependence of y_t is governed by that of μ_t because $\{e_t\}$ is not serially correlated.

The model in Eqs. (11.1) and (11.2) can readily be used to analyze realized volatility of an asset price; see Example 11.1 below. Here μ_t represents the underlying log volatility of the asset price and y_t is the logarithm of realized volatility. The true log volatility is not directly observed but evolves over time according to a random-walk model. On the other hand, y_t is constructed from high-frequency transactions data and subjected to the influence of market microstructure. The standard deviation of e_t denotes the scale used to measure the impact of market microstructure.

The model in Eqs. (11.1) and (11.2) is a special *linear Gaussian state-space model*. The variable μ_t is called the *state* of the system at time t and is not directly observed. Equation (11.1) provides the link between the data y_t and the state μ_t and is called the *observation equation* with *measurement error* e_t. Equation (11.2) governs the time evolution of the state variable and is the *state equation* (or *state transition equation*) with innovation η_t. The model is also called a *local level model* in Durbin and Koopman (2001, Chapter 2), which is a simple case of the *structural time series model* of Harvey (1993).

Relationship to ARIMA Model

If there is no measurement error in Eq. (11.1), that is, $\sigma_e = 0$, then $y_t = \mu_t$, which is an ARIMA(0,1,0) model. If $\sigma_e > 0$, that is, there exist measurement errors, then y_t is an ARIMA(0,1,1) model satisfying

$$(1 - B)y_t = (1 - \theta B)a_t, \tag{11.3}$$

where $\{a_t\}$ is a Gaussian white noise with mean zero and variance σ_a^2. The values of θ and σ_a^2 are determined by σ_e and σ_η. This result can be derived as follows.

From Eq. (11.2), we have

$$(1 - B)\mu_{t+1} = \eta_t, \quad \text{or} \quad \mu_{t+1} = \frac{1}{1 - B}\eta_t.$$

Using this result, Eq. (11.1) can be written as

$$y_t = \frac{1}{1 - B}\eta_{t-1} + e_t.$$

Multiplying by $(1 - B)$, we have

$$(1 - B)y_t = \eta_{t-1} + e_t - e_{t-1}.$$

Let $(1 - B)y_t = w_t$. We have $w_t = \eta_{t-1} + e_t - e_{t-1}$. Under the assumptions, it is easy to see that (a) w_t is Gaussian, (b) $\text{Var}(w_t) = 2\sigma_e^2 + \sigma_\eta^2$, (c) $\text{Cov}(w_t, w_{t-1}) = -\sigma_e^2$, and (d) $\text{Cov}(w_t, w_{t-j}) = 0$ for $j > 1$. Consequently, w_t follows an MA(1) model and can be written as $w_t = (1 - \theta B)a_t$. By equating the variance and lag-1 autocovariance of $w_t = (1 - \theta B)a_t = \eta_{t-1} + e_t - e_{t-1}$, we have

$$(1 + \theta^2)\sigma_a^2 = 2\sigma_e^2 + \sigma_\eta^2,$$
$$\theta\sigma_a^2 = \sigma_e^2.$$

For given σ_e^2 and σ_η^2, one considers the ratio of the prior two equations to form a quadratic function of θ. This quadratic form has two solutions so one should select the one that satisfies $|\theta| < 1$. The value of σ_a^2 can then be easily obtained. Thus, the state-space model in Eqs. (11.1) and (11.2) is also an ARIMA(0,1,1) model, which is the simple exponential smoothing model of Chapter 2.

On the other hand, for an ARIMA(0,1,1) model with positive θ, one can use the prior two identities to solve for σ_e^2 and σ_η^2, and obtain a local trend model. If θ is negative, then the model can still be put in a state-space form without the observational error, that is, $\sigma_e = 0$. In fact, as will be seen later, an ARIMA model can be transformed into state-space models in many ways. Thus, the linear state-space model is closely related to the ARIMA model.

In practice, what one observes is the y_t series. Thus, based on the data alone, the decision of using ARIMA models or linear state-space models is not critical. Both model representations have pros and cons. The objective of data analysis, substantive issues, and experience all play a role in choosing a statistical model.

Example 11.1. To illustrate the ideas of the state-space model and Kalman filter, we consider the intradaily realized volatility of Alcoa stock from January 2, 2003 to May 7, 2004 for 340 observations. The daily realized volatility used is the sum of squares of intraday 10-minute log returns measured in percentage. No overnight returns or the first 10-minute intraday returns are used. See Chapter 3 for more information about realized volatility. The series used in the demonstration is the logarithm of the daily realized volatility.

Figure 11.1 shows the time plot of the logarithms of the realized volatility of Alcoa stock from January 2, 2003 to May 7, 2004. The transactions data are obtained from the TAQ database of the NYSE. If ARIMA models are entertained, we obtain an ARIMA(0,1,1) model

$$(1 - B)y_t = (1 - 0.855B)a_t, \quad \hat{\sigma}_a = 0.5184, \tag{11.4}$$

where y_t is the log realized volatility, and the standard error of $\hat{\theta}$ is 0.029. The residuals show $Q(12) = 12.4$ with p-value 0.33, indicating that there is no significant

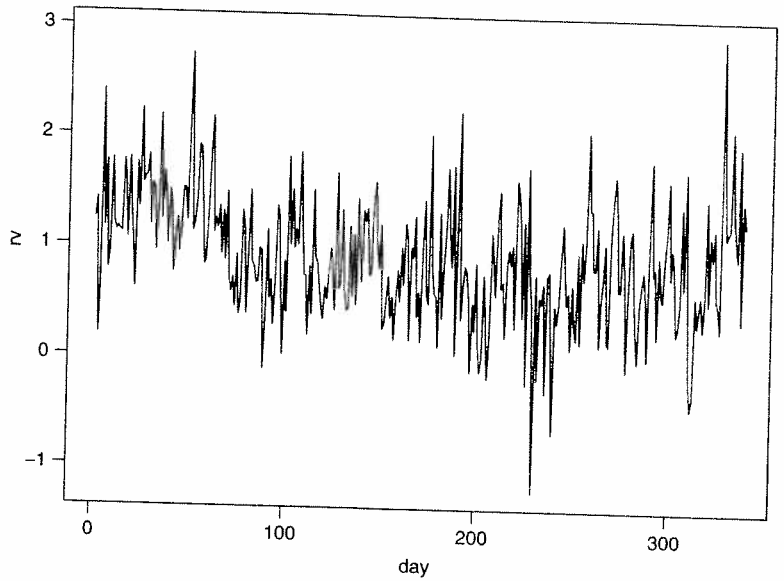

Figure 11.1. Time plot of the logarithms of intradaily realized volatility of Alcoa stock from January 2, 2003 to May 7, 2004. The realized volatility is computed from the intraday 10-minute log returns measured in percentage.

serial correlation in the residuals. Similarly, the squared residuals give $Q(12) = 8.2$ with p-value 0.77, suggesting no ARCH effects in the series.

Since $\hat{\theta}$ is positive, we can transform the ARIMA(0,1,1) model into a local trend model in Eqs. (11.1) and (11.2). The maximum likelihood estimates of the two parameters are $\hat{\sigma}_\eta = 0.0735$ and $\hat{\sigma}_e = 0.4803$. The measurement errors have a larger variance than the state innovations, confirming that intraday high-frequency returns are subject to measurement errors. Details of estimation will be discussed in Section 11.1.7. Here we treat the two estimates as given and use the model to demonstrate application of the Kalman filter.

11.1.1 Statistical Inference

Return to the state-space model in Eqs. (11.1) and (11.2). The aim of the analysis is to infer properties of the state μ_t from the data $\{y_t | t = 1, \ldots, T\}$ and the model. Three types of inference are commonly discussed in the literature. They are *filtering*, *prediction*, and *smoothing*. Let $F_t = \{y_1, \ldots, y_t\}$ be the information available at time t (inclusive) and assume that the model is known, including all parameters. The three types of inference can briefly be described as follows:

- *Filtering*. Filtering means to recover the state variable μ_t given F_t, that is, to remove the measurement errors from the data.

- *Prediction*. Prediction means to forecast μ_{t+h} or y_{t+h} for $h > 0$ given F_t, where t is the forecast origin.
- *Smoothing*. Smoothing is to estimate μ_t given F_T, where $T > t$.

A simple analogy of the three types of inference is reading a handwritten note. Filtering is figuring out the word you are reading based on knowledge accumulated from the beginning of the note, predicting is to guess the next word, and smoothing is deciphering a particular word once you have read through the note.

To describe the inference more precisely, we introduce some notation. Let $\mu_{t|j} = E(\mu_t|F_j)$ and $\Sigma_{t|j} = \text{Var}(\mu_t|F_j)$ be, respectively, the conditional mean and variance of μ_t given F_j. Similarly, $y_{t|j}$ denotes the conditional mean of y_t given F_j. Furthermore, let $v_t = y_t - y_{t|t-1}$ and $V_t = \text{Var}(v_t|F_{t-1})$ be the 1-step ahead forecast error and its variance of y_t given F_{t-1}. Note that the forecast error v_t is independent of F_{t-1} so that the conditional variance is the same as the unconditional variance; that is, $\text{Var}(v_t|F_{t-1}) = \text{Var}(v_t)$. From Eq. (11.1),

$$y_{t|t-1} = E(y_t|F_{t-1}) = E(\mu_t + e_t|F_{t-1}) = E(\mu_t|F_{t-1}) = \mu_{t|t-1}.$$

Consequently,

$$v_t = y_t - y_{t|t-1} = y_t - \mu_{t|t-1} \tag{11.5}$$

and

$$\begin{aligned} V_t &= \text{Var}(y_t - \mu_{t|t-1}|F_{t-1}) = \text{Var}(\mu_t + e_t - \mu_{t|t-1}|F_{t-1}) \\ &= \text{Var}(\mu_t - \mu_{t|t-1}|F_{t-1}) + \text{Var}(e_t|F_{t-1}) = \Sigma_{t|t-1} + \sigma_e^2. \end{aligned} \tag{11.6}$$

It is also easy to see that

$$\begin{aligned} E(v_t) &= E[E(v_t|F_{t-1})] = E[E(y_t - y_{t|t-1}|F_{t-1})] = E[y_{t|t-1} - y_{t|t-1}] = 0, \\ \text{Cov}(v_t, y_j) &= E(v_t y_j) = E[E(v_t y_j|F_{t-1})] = E[y_j E(v_t|F_{t-1})] = 0, \quad j < t. \end{aligned}$$

Thus, as expected, the 1-step ahead forecast error is uncorrelated (hence, independent) with y_j for $j < t$. Furthermore, for the linear model in Eqs. (11.1) and (11.2), $\mu_{t|t} = E(\mu_t|F_t) = E(\mu_t|F_{t-1}, v_t)$ and $\Sigma_{t|t} = \text{Var}(\mu_t|F_t) = \text{Var}(\mu_t|F_{t-1}, v_t)$. In other words, the information set F_t can be written as $F_t = \{F_{t-1}, y_t\} = \{F_{t-1}, v_t\}$.

The following properties of multivariate normal distribution are useful in studying the Kalman filter under normality. They can be shown via the multivariate linear regression method or factorization of the joint density. See, also, Appendix B of Chapter 8. For random vectors $\boldsymbol{w}$ and $\boldsymbol{m}$, denote the mean vectors and covariance matrix as $E(\boldsymbol{w}) = \boldsymbol{\mu}_w$, $E(\boldsymbol{m}) = \boldsymbol{\mu}_m$, and $\text{Cov}(\boldsymbol{m}, \boldsymbol{w}) = \boldsymbol{\Sigma}_{mw}$, respectively.

Theorem 11.1. *Suppose that $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are three random vectors such that their joint distribution is multivariate normal. In addition, assume that the diagonal block covariance matrix $\boldsymbol{\Sigma}_{ww}$ is nonsingular for $w = x, y, z$, and $\boldsymbol{\Sigma}_{yz} = \boldsymbol{0}$. Then,*

1. $E(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$.
2. $\text{Var}(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}$.

3. $E(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) = E(\boldsymbol{x}|\boldsymbol{y}) + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z)$.
4. $\text{Var}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) = \text{Var}(\boldsymbol{x}|\boldsymbol{y}) - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx}$.

11.1.2 Kalman Filter

The goal of the *Kalman filter* is to update knowledge of the state variable recursively when a new data point becomes available. That is, knowing the conditional distribution of μ_t given F_{t-1} and the new data y_t, we would like to obtain the conditional distribution of μ_t given F_t, where, as before, $F_j = \{y_1, \ldots, y_j\}$. Since $F_t = \{F_{t-1}, v_t\}$, giving y_t and F_{t-1} is equivalent to giving v_t and F_{t-1}. Consequently, to derive the Kalman filter, it suffices to consider the joint conditional distribution of $(\mu_t, v_t)'$ given F_{t-1} before applying Theorem 11.1.

The conditional distribution of v_t given F_{t-1} is normal with mean zero and variance given in Eq. (11.6), and that of μ_t given F_{t-1} is also normal with mean $\mu_{t|t-1}$ and variance $\Sigma_{t|t-1}$. Furthermore, the joint distribution of $(\mu_t, v_t)'$ given F_{t-1} is also normal. Thus, what remains to be solved is the conditional covariance between μ_t and v_t given F_{t-1}. From the definition,

$$\begin{aligned} \text{Cov}(\mu_t, v_t|F_{t-1}) &= E(\mu_t v_t|F_{t-1}) = E[\mu_t(y_t - \mu_{t|t-1})|F_{t-1}] \quad \text{(by Eq.(11.5))} \\ &= E[\mu_t(\mu_t + e_t - \mu_{t|t-1})|F_{t-1}] \\ &= E[\mu_t(\mu_t - \mu_{t|t-1})|F_{t-1}] + E(\mu_t e_t|F_{t-1}) \\ &= E[(\mu_t - \mu_{t|t-1})^2|F_{t-1}] = \text{Var}(\mu_t|F_{t-1}) = \Sigma_{t|t-1}, \end{aligned} \tag{11.7}$$

where we have used the fact that $E[\mu_{t|t-1}(\mu_t - \mu_{t|t-1})|F_{t-1}] = 0$. Putting the results together, we have

$$\begin{bmatrix} \mu_t \\ v_t \end{bmatrix}_{F_{t-1}} \sim N\left(\begin{bmatrix} \mu_{t|t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{bmatrix}\right).$$

By Theorem 11.1, the conditional distribution of μ_t given F_t is normal with mean and variance

$$\mu_{t|t} = \mu_{t|t-1} + \frac{\Sigma_{t|t-1}v_t}{V_t} = \mu_{t|t-1} + K_t v_t, \tag{11.8}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \frac{\Sigma_{t|t-1}^2}{V_t} = \Sigma_{t|t-1}(1 - K_t), \tag{11.9}$$

where $K_t = \Sigma_{t|t-1}/V_t$ is commonly referred to as the *Kalman gain*, which is the regression coefficient of μ_t on v_t. From Eq. (11.8), Kalman gain is the factor that governs the contribution of the new shock v_t to the state variable μ_t.

Next, one can make use of the knowledge of μ_t given F_t to predict μ_{t+1} via Eq. (11.2). Specifically, we have

$$\mu_{t+1|t} = E(\mu_t + \eta_t|F_t) = E(\mu_t|F_t) = \mu_{t|t}, \tag{11.10}$$

$$\Sigma_{t+1|t} = \text{Var}(\mu_{t+1}|F_t) = \text{Var}(\mu_t|F_t) + \text{Var}(\eta_t) = \Sigma_{t|t} + \sigma_\eta^2. \tag{11.11}$$

Once the new data y_{t+1} is observed, one can repeat the above procedure to update knowledge of μ_{t+1}. This is the famous *Kalman filter* algorithm proposed by Kalman (1960).

In summary, putting Eqs. (11.5)–(11.11) together and conditioning on the initial assumption that μ_1 is distributed as $N(\mu_{1|0}, \Sigma_{1|0})$, the Kalman filter for the local trend model is as follows:

$$\begin{aligned} v_t &= y_t - \mu_{t|t-1}, \\ V_t &= \Sigma_{t|t-1} + \sigma_e^2, \\ K_t &= \Sigma_{t|t-1}/V_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + K_t v_t, \\ \Sigma_{t+1|t} &= \Sigma_{t|t-1}(1 - K_t) + \sigma_\eta^2, \quad t = 1, \ldots, T. \end{aligned} \tag{11.12}$$

There are many ways to derive the Kalman filter. We use Theorem 11.1, which describes some properties of multivariate normal distribution, for its simplicity. In practice, the choice of initial values $\Sigma_{1|0}$ and $\mu_{1|0}$ requires some attention and we shall discuss it later in Section 11.1.6. For the local trend model in Eqs. (11.1) and (11.2), the two parameters σ_e and σ_η can be estimated via the maximum likelihood method. Again, the Kalman filter is useful in evaluating the likelihood function of the data in estimation. We shall discuss estimation in Section 11.1.7.

Example 11.1 (Continued). To illustrate application of the Kalman filter, we use the fitted state-space model for daily realized volatility of Alcoa stock returns and apply the Kalman filter algorithm to the data with $\Sigma_{1|0} = \infty$ and $\mu_{1|0} = 0$. The choice of these initial values will be discussed in Section 11.1.6. Figure 11.2a shows the time plot of the filtered state variable $\mu_{t|t}$ and Figure 11.2b is the time plot of the 1-step ahead forecast error v_t. Compared with Figure 11.1, the filtered states are smoother. The forecast errors appear to be stable and center around zero. These forecast errors are out-of-sample 1-step ahead prediction errors.

11.1.3 Properties of Forecast Error

The one-step ahead forecast errors $\{v_t\}$ are useful in many applications, hence it pays to study carefully their properties. Given the initial values $\Sigma_{1|0}$ and $\mu_{1|0}$, which are independent of y_t, the Kalman filter enables us to compute v_t recursively as a linear function of $\{y_1, \ldots, y_t\}$. Specifically, by repeated substitutions,

$$\begin{aligned} v_1 &= y_1 - \mu_{1|0}, \\ v_2 &= y_2 - \mu_{2|1} = y_2 - \mu_{1|0} - K_1(y_1 - \mu_{1|0}), \\ v_3 &= y_3 - \mu_{3|2} = y_3 - \mu_{1|0} - K_2(y_2 - \mu_{1|0}) - K_1(1 - K_2)(y_1 - \mu_{1|0}), \end{aligned}$$

and so on. This transformation can be written in matrix form as

$$\boldsymbol{v} = \boldsymbol{K}(\boldsymbol{y} - \mu_{1|0}\mathbf{1}_T), \tag{11.13}$$

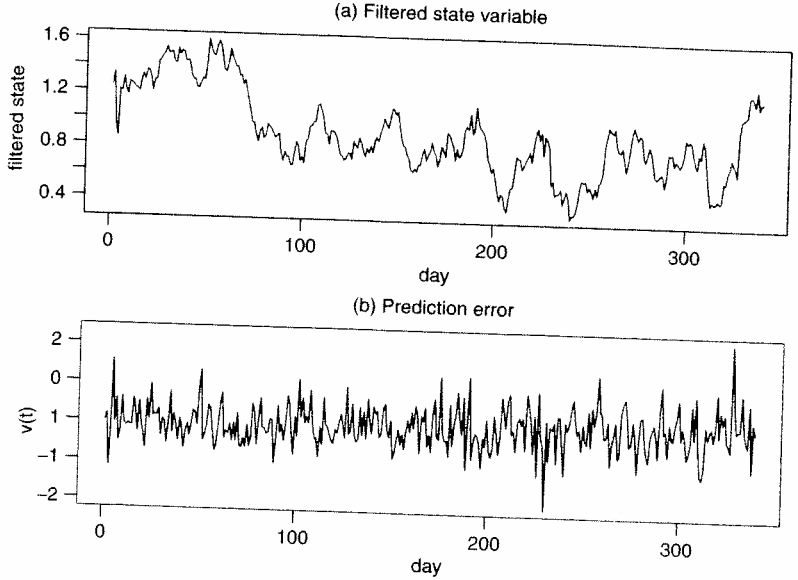

Figure 11.2. Time plots of output of the Kalman filter applied to the daily realized log volatility of Alcoa stock based on the local trend state-space model: (a) the filtered state $\mu_{t|t}$ and (b) the one-step ahead forecast error v_t.

where $\boldsymbol{v} = (v_1, \ldots, v_T)'$, $\boldsymbol{y} = (y_1, \ldots, y_T)'$, $\mathbf{1}_T$ is the T-dimensional vector of ones, and $\boldsymbol{K}$ is a lower triangular matrix defined as

$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ k_{21} & 1 & 0 & \cdots & 0 \\ k_{31} & k_{32} & 1 & & 0 \\ \vdots & \vdots & & & \vdots \\ k_{T1} & k_{T2} & k_{T3} & \cdots & 1 \end{bmatrix},$$

where $k_{i,i-1} = -K_{i-1}$ and $k_{ij} = -(1 - K_{i-1})(1 - K_{i-2}) \cdots (1 - K_{j+1})K_j$ for $i = 2, \ldots, T$ and $j = 1, \ldots, i-2$. It should be noted that, from the definition, the Kalman gain K_t does not depend on $\mu_{1|0}$ or the data $\{y_1, \ldots, y_t\}$; it depends on $\Sigma_{1|0}$ and σ_e^2 and σ_η^2.

The transformation in Eq. (11.13) has several important implications. First, $\{v_t\}$ are mutually independent under the normality assumption. To show this, consider the joint probability density function of the data

$$p(y_1, \ldots, y_T) = p(y_1) \prod_{j=2}^{T} p(y_j | F_{j-1}).$$

Equation (11.13) indicates that the transformation from y_t to v_t has a unit Jacobian so that $p(\boldsymbol{v}) = p(\boldsymbol{y})$. Furthermore, since $\mu_{1|0}$ is given, $p(v_1) = p(y_1)$. Consequently, the joint probability density function of $\boldsymbol{v}$ is

$$p(\boldsymbol{v}) = p(\boldsymbol{y}) = p(y_1)\prod_{j=2}^{T} p(y_j|F_{j-1}) = p(v_1)\prod_{j}^{T} p(v_j) = \prod_{j=1}^{T} p(v_j).$$

This shows that $\{v_t\}$ are mutually independent.

Second, the Kalman filter provides a Cholesky decomposition of the covariance matrix of $\boldsymbol{y}$. To see this, let $\boldsymbol{\Omega} = \text{Cov}(\boldsymbol{y})$. Equation (11.13) shows that $\text{Cov}(\boldsymbol{v}) = \boldsymbol{K}\boldsymbol{\Omega}\boldsymbol{K}'$. On the other hand, $\{v_t\}$ are mutually independent with $\text{Var}(v_t) = V_t$. Therefore, $\boldsymbol{K}\boldsymbol{\Omega}\boldsymbol{K}' = \text{diag}\{V_1, \ldots, V_T\}$, which is precisely a Cholesky decomposition of $\boldsymbol{\Omega}$. The elements k_{ij} of the matrix $\boldsymbol{K}$ thus have some nice interpretations; see Chapter 10.

State Error Recursion

Turn to the estimation error of the state variable μ_t. Define

$$x_t = \mu_t - \mu_{t|t-1}$$

as the forecast error of the state variable μ_t given data F_{t-1}. From Section 11.1.1, $\text{Var}(x_t|F_{t-1}) = \Sigma_{t|t-1}$. From the Kalman filter in Eq. (11.12),

$$v_t = y_t - \mu_{t|t-1} = \mu_t + e_t - \mu_{t|t-1} = x_t + e_t,$$

and

$$\begin{aligned} x_{t+1} &= \mu_{t+1} - \mu_{t+1|t} = \mu_t + \eta_t - (\mu_{t|t-1} + K_t v_t) \\ &= x_t + \eta_t - K_t v_t = x_t + \eta_t - K_t(x_t + e_t) = L_t x_t + \eta_t - K_t e_t, \end{aligned}$$

where $L_t = 1 - K_t = 1 - \Sigma_{t|t-1}/V_t = (V_t - \Sigma_{t|t-1})/V_t = \sigma_e^2/V_t$. Consequently, for the state errors, we have

$$v_t = x_t + e_t, \quad x_{t+1} = L_t x_t + \eta_t - K_t e_t, \quad t = 1, \ldots, T, \tag{11.14}$$

where $x_1 = \mu_1 - \mu_{1|0}$. Equation (11.14) is in the form of a time-varying state-space model with x_t being the state variable and v_t the observation.

11.1.4 State Smoothing

Next we consider the estimation of the state variables $\{\mu_1, \ldots, \mu_T\}$ given the data F_T and the model. That is, given the state-space model in Eqs. (11.1) and (11.2), we wish to obtain the conditional distribution $\mu_t|F_T$ for all t. To this end, we first recall some facts available about the model:

- All distributions involved are normal so that we can write the conditional distribution of μ_t given F_T as $N(\mu_{t|T}, \Sigma_{t|T})$, where $t \le T$. We refer to $\mu_{t|T}$ as the *smoothed state* at time t and $\Sigma_{t|T}$ as the *smoothed state variance*.

- Based on the properties of $\{v_t\}$ shown in Section 11.1.3, $\{v_1, \ldots, v_T\}$ are mutually independent and are linear functions of $\{y_1, \ldots, y_T\}$.
- If $y_1, \ldots, y_T$ are fixed, then F_{t-1} and $\{v_t, \ldots, v_T\}$ are fixed, and vice versa.
- $\{v_t, \ldots, v_T\}$ are independent of F_{t-1} with mean zero and variance $\text{Var}(v_j) = V_j$ for $j \ge t$.

Applying Theorem 11.1(3) to the conditional joint distribution of $(\mu_t, v_t, \ldots, v_T)$ given F_{t-1}, we have

$$\begin{aligned} \mu_{t|T} &= E(\mu_t|F_T) = E(\mu_t|F_{t-1}, v_t, \ldots, v_T) \\ &= E(\mu_t|F_{t-1}) + \text{Cov}[\mu_t, (v_t, \ldots, v_T)']\text{Cov}[(v_t, \ldots, v_T)']^{-1}(v_t, \ldots, v_T)' \\ &= \mu_{t|t-1} + \begin{bmatrix} \text{Cov}(\mu_t, v_t) \\ \text{Cov}(\mu_t, v_{t+1}) \\ \vdots \\ \text{Cov}(\mu_t, v_T) \end{bmatrix}' \begin{bmatrix} V_t & 0 & \cdots & 0 \\ 0 & V_{t+1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & V_T \end{bmatrix}^{-1} \begin{bmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_T \end{bmatrix} \\ &= \mu_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(\mu_t, v_j)V_j^{-1}v_j. \end{aligned} \tag{11.15}$$

From the definition and independence of $\{v_t\}$, $\text{Cov}(\mu_t, v_j) = \text{Cov}(x_t, v_j)$ for $j = t, \ldots, T$, and

$$\begin{aligned} \text{Cov}(x_t, v_t) &= E[x_t(x_t + e_t)] = \text{Var}(x_t) = \Sigma_{t|t-1}, \\ \text{Cov}(x_t, v_{t+1}) &= E[x_t(x_{t+1} + e_{t+1})] = E[x_t(L_t x_t + \eta_t - K_t e_t)] = \Sigma_{t|t-1}L_t. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \text{Cov}(x_t, v_{t+2}) &= E[x_t(x_{t+2} + e_{t+2})] = \cdots = \Sigma_{t|t-1}L_t L_{t+1}, \\ &\vdots = \vdots \\ \text{Cov}(x_t, v_T) &= E[x_t(x_T + e_T)] = \cdots = \Sigma_{t|t-1}\prod_{j=t}^{T-1} L_j. \end{aligned}$$

Consequently, Eq. (11.15) becomes

$$\begin{aligned} \mu_{t|T} &= \mu_{t|t-1} + \Sigma_{t|t-1}\frac{v_t}{V_t} + \Sigma_{t|t-1}L_t\frac{v_{t+1}}{V_{t+1}} + \Sigma_{t|t-1}L_t L_{t+1}\frac{v_{t+2}}{V_{t+2}} + \cdots \\ &\equiv \mu_{t|t-1} + \Sigma_{t|t-1}q_{t-1}, \end{aligned}$$

where

$$q_{t-1} = \frac{v_t}{V_t} + L_t\frac{v_{t+1}}{V_{t+1}} + L_t L_{t+1}\frac{v_{t+2}}{V_{t+2}} + \cdots + \left(\prod_{j=t}^{T-1} L_j\right)\frac{v_T}{V_T} \tag{11.16}$$

is a weighted linear combination of the innovations $\{v_t, \ldots, v_T\}$. This weighted sum satisfies

$$q_{t-1} = \frac{v_t}{V_t} + L_t \left[\frac{v_{t+1}}{V_{t+1}} + L_{t+1}\frac{v_{t+2}}{V_{t+2}} + \cdots + \left(\prod_{j=t+1}^{T-1} L_j \right) \frac{v_T}{V_T} \right]$$
$$= \frac{v_t}{V_t} + L_t q_t.$$

Therefore, using the initial value $q_T = 0$, we have the backward recursion

$$q_{t-1} = \frac{v_t}{V_t} + L_t q_t, \quad t = T, T-1, \ldots, 1. \tag{11.17}$$

Putting Eqs. (11.15)–(11.17) together, we have a backward recursive algorithm to compute the smoothed state variables:

$$q_{t-1} = V_t^{-1} v_t + L_t q_t, \quad \mu_{t|T} = \mu_{t|t-1} + \Sigma_{t|t-1} q_{t-1}, \quad t = T, \ldots, 1, \tag{11.18}$$

where $q_T = 0$, and $\mu_{t|t-1}$, $\Sigma_{t|t-1}$ and L_t are available from the Kalman filter in Eq. (11.12).

Smoothed State Variance

The variance of the smoothed state variable $\mu_{t|T}$ can be derived in a similar manner via Theorem 11.1(4). Specifically, letting $\boldsymbol{v}_t^T = (v_t, \ldots, v_T)'$, we have

$$\Sigma_{t|T} = \mathrm{Var}(\mu_t | F_T) = \mathrm{Var}(\mu_t | F_{t-1}, v_t, \ldots, v_T)$$
$$= \mathrm{Var}(\mu_t | F_{t-1}) - \mathrm{Cov}[\mu_t, (\boldsymbol{v}_t^T)'][\mathrm{Cov}[(\boldsymbol{v}_t^T)]]^{-1}\mathrm{Cov}[\mu_t, (\boldsymbol{v}_t^T)]$$
$$= \Sigma_{t|t-1} - \sum_{j=t}^{T} [\mathrm{Cov}(\mu_t, v_j)]^2 V_j^{-1}, \tag{11.19}$$

where $\mathrm{Cov}(\mu_t, v_j) = \mathrm{Cov}(x_t, v_j)$ are given earlier after Eq. (11.15). Thus,

$$\Sigma_{t|T} = \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 \frac{1}{V_t} - \Sigma_{t|t-1}^2 L_t^2 \frac{1}{V_{t+1}} - \cdots - \Sigma_{t|t-1}^2 \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T}$$
$$\equiv \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1}, \tag{11.20}$$

where

$$M_{t-1} = \frac{1}{V_t} + L_t^2 \frac{1}{V_{t+1}} + L_t^2 L_{t+1}^2 \frac{1}{V_{t+2}} + \cdots + \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T}$$

is a weighted linear combination of the inverses of variances of the 1-step ahead forecast errors after time $t-1$. Let $M_T = 0$ because no 1-step ahead forecast error

is available after time index T. The statistic M_{t-1} can be written as

$$M_{t-1} = \frac{1}{V_t} + L_t^2 \left[\frac{1}{V_{t+1}} + L_{t+1}^2 \frac{1}{V_{t+2}} + \cdots + \left(\prod_{j=t+1}^{T-1} L_j^2 \right) \frac{1}{V_T} \right]$$
$$= \frac{1}{V_t} + L_t^2 M_t, \quad t = T, T-1, \ldots, 1.$$

Note that from the independence of $\{v_t\}$ and Eq. (11.16), we have

$$\mathrm{Var}(q_{t-1}) = \frac{1}{V_t} + L_t^2 \frac{1}{V_{t+1}} + \cdots + \left(\prod_{j=t}^{T-1} L_j^2 \right) \frac{1}{V_T} = M_{t-1}.$$

Combining the results, variances of the smoothed state variables can be computed efficiently via the backward recursion

$$M_{t-1} = V_t^{-1} + L_t^2 M_t, \quad \Sigma_{t|T} = \Sigma_{t|t-1} - \Sigma_{t|t-1}^2 M_{t-1}, \quad t = T, \ldots, 1, \tag{11.21}$$

where $M_T = 0$.

Example 11.1 (Continued). Applying the Kalman filter and state-smoothing algorithms in Eqs. (11.18) and (11.21) to the daily realized volatility of Alcoa stock using the fitted state-space model, we can easily compute the filtered state $\mu_{t|t}$ and the smoothed state $\mu_{t|T}$ and their variances. Figure 11.3 shows the filtered state variable and its 95% pointwise confidence interval, whereas Figure 11.4 provides the time plot of smoothed state variable and its 95% pointwise confidence interval. As expected, the smoothed state variables are smoother than the filtered state variables. The confidence intervals for the smoothed state variables are also narrower than those of the filtered state variables. Note that the width of the 95% confidence interval of $\mu_{1|1}$ depends on the initial value $\Sigma_{1|0}$.

11.1.5 Missing Values

An advantage of the state-space model is in handling missing values. Suppose that the observations $\{y_t\}_{t=\ell+1}^{\ell+h}$ are missing, where $h \geq 1$ and $1 \leq \ell < T$. There are several ways to handle missing values in state-space formulation. Here we discuss a method that keeps the original time scale and model form. For $t \in \{\ell + 1, \ldots, \ell + h\}$, we can use Eq. (11.2) to express μ_t as a linear combination of $\mu_{\ell+1}$ and $\{\eta_j\}_{j=\ell+1}^{t-1}$. Specifically,

$$\mu_t = \mu_{t-1} + \eta_{t-1} = \cdots = \mu_{\ell+1} + \sum_{j=\ell+1}^{t-1} \eta_j,$$

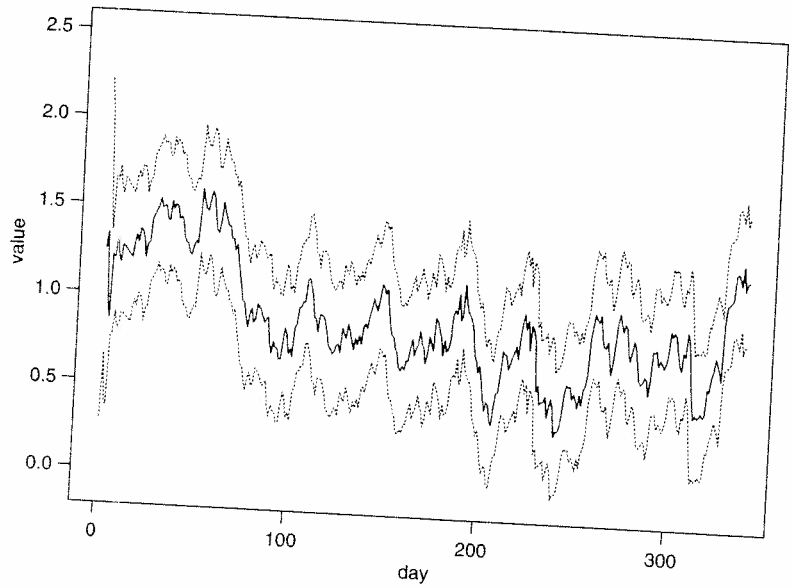

Figure 11.3. Filtered state variable $\mu_{t|t}$ and its 95% pointwise confidence interval for the daily log realized volatility of Alcoa stock returns based on the fitted local trend state-space model.

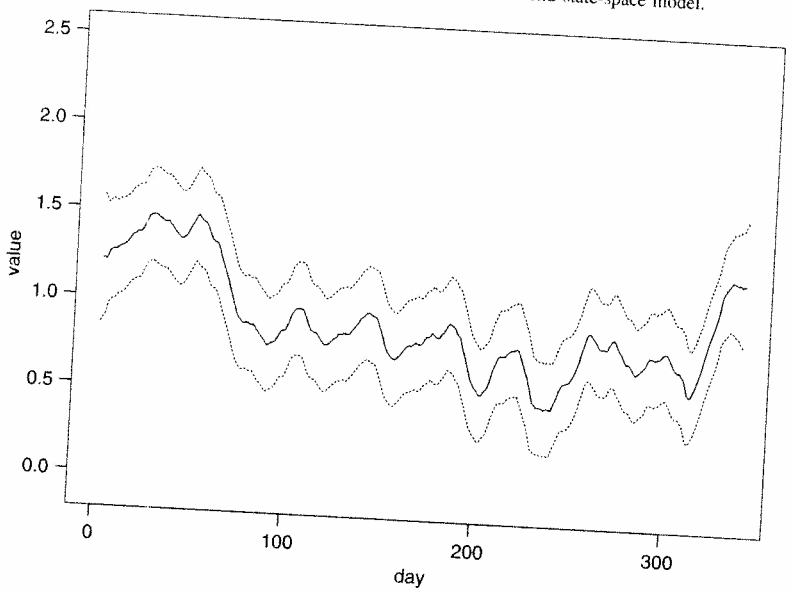

Figure 11.4. Smoothed state variable $\mu_{t|T}$ and its 95% pointwise confidence interval for the daily log realized volatility of Alcoa stock returns based on the fitted local trend state-space model.

where it is understood that the summation term is zero if its lower limit is greater than its upper limit. Therefore, for $t \in \{\ell+1, \ldots, \ell+h\}$,

$$E(\mu_t|F_{t-1}) = E(\mu_t|F_\ell) = \mu_{\ell+1|\ell},$$
$$\text{Var}(\mu_t|F_{t-1}) = \text{Var}(\mu_t|F_\ell) = \Sigma_{\ell+1|\ell} + (t-\ell-1)\sigma_\eta^2.$$

Consequently, we have

$$\mu_{t|t-1} = \mu_{t-1|t-2}, \quad \Sigma_{t|t-1} = \Sigma_{t-1|t-2} + \sigma_\eta^2, \tag{11.22}$$

for $t = \ell+2, \ldots, \ell+h$. These results show that we can continue to apply the Kalman filter algorithm in Eq. (11.12) by taking $v_t = 0$ and $K_t = 0$ for $t = \ell+1, \ldots, \ell+h$. This is rather natural because when y_t is missing, there is no new innovation or new Kalman gain so that $v_t = 0$ and $K_t = 0$.

11.1.6 Effect of Initialization

In this subsection, we consider the effects of initial condition $\mu_1 \sim N(\mu_{1|0}, \Sigma_{1|0})$ on the Kalman filter and state smoothing. From the Kalman filter in Eq. (11.12),

$$v_1 = y_1 - \mu_{1|0}, \quad V_1 = \Sigma_{1|0} + \sigma_e^2,$$

and, by Eqs. (11.8)–(11.11),

$$\mu_{2|1} = \mu_{1|0} + \frac{\Sigma_{1|0}}{V_1} v_1 = \mu_{1|0} + \frac{\Sigma_{1|0}}{\Sigma_{1|0}+\sigma_e^2}(y_1 - \mu_{1|0}),$$
$$\Sigma_{2|1} = \Sigma_{1|0}\left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0}+\sigma_e^2}\right) + \sigma_\eta^2 = \frac{\Sigma_{1|0}}{\Sigma_{1|0}+\sigma_e^2}\sigma_e^2 + \sigma_\eta^2.$$

Therefore, letting $\Sigma_{1|0}$ increase to infinity, we have $\mu_{2|1} = y_1$ and $\Sigma_{2|1} = \sigma_e^2 + \sigma_\eta^2$. This is equivalent to treating y_1 as fixed and assuming $\mu_1 \sim N(y_1, \sigma_e^2)$. In the literature, this approach to initializing the Kalman filter is called *diffuse initialization* because a very large $\Sigma_{1|0}$ means one is uncertain about the initial condition.

Next, turn to the effect of diffuse initialization on state smoothing. It is obvious that based on the results of Kalman filtering, state smoothing is not affected by the diffuse initialization for $t = T, \ldots, 2$. Thus, we focus on μ_1 given F_T. From Eq. (11.18) and the definition of $L_1 = 1 - K_1 = V_1^{-1}\sigma_e^2$,

$$\begin{aligned}\mu_{1|T} &= \mu_{1|0} + \Sigma_{1|0} q_0 \\ &= \mu_{1|0} + \Sigma_{1|0}\left[\frac{1}{\Sigma_{1|0}+\sigma_e^2} v_1 + \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0}+\sigma_e^2}\right) q_1\right] \\ &= \mu_{1|0} + \frac{\Sigma_{1|0}}{\Sigma_{1|0}+\sigma_e^2}(v_1 + \sigma_e^2 q_1).\end{aligned}$$

Letting $\Sigma_{1|0} \to \infty$, we have $\mu_{1|T} = \mu_{1|0} + v_1 + \sigma_e^2 q_1 = y_1 + \sigma_e^2 q_1$. Furthermore, from Eq. (11.21) and using $V_1 = \Sigma_{1|0} + \sigma_e^2$, we have

$$\begin{aligned}\Sigma_{1|T} &= \Sigma_{1|0} - \Sigma_{1|0}^2 \left[\frac{1}{\Sigma_{1|0} + \sigma_e^2} + \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 M_1 \right] \\ &= \Sigma_{1|0} \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right) - \left(1 - \frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 \Sigma_{1|0}^2 M_1 \\ &= \left(\frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right) \sigma_e^2 - \left(\frac{\Sigma_{1|0}}{\Sigma_{1|0} + \sigma_e^2}\right)^2 \sigma_e^4 M_1.\end{aligned}$$

Thus, letting $\Sigma_{1|0} \to \infty$, we obtain $\Sigma_{1|T} = \sigma_e^2 - \sigma_e^4 M_1$.

Based on the prior discussion, we suggest using diffuse initialization when little is known about the initial value μ_1. However, it might be hard to justify the use of a random variable with infinite variance in real applications. If necessary, one can treat μ_1 as an additional parameter of the state-space model and estimate it jointly with other parameters. This latter approach is closely related to the exact maximum likelihood estimation of Chapters 2 and 8.

11.1.7 Estimation

In this subsection, we consider the estimation of σ_e and σ_η of the local trend model in Eqs. (11.1) and (11.2). Based on properties of forecast errors discussed in Section 11.1.3, the Kalman filter provides an efficient way to evaluate the likelihood function of the data for estimation. Specifically, the likelihood function under normality is

$$\begin{aligned}p(y_1, \ldots, y_T | \sigma_e, \sigma_\eta) &= p(y_1 | \sigma_e, \sigma_\eta) \prod_{t=2}^{T} (y_t | F_{t-1}, \sigma_e, \sigma_\eta) \\ &= p(y_1 | \sigma_e, \sigma_\eta) \prod_{t=2}^{T} (v_t | F_{t-1}, \sigma_e, \sigma_\eta),\end{aligned}$$

where $y_1 \sim N(\mu_{1|0}, V_1)$ and $v_t = (y_t - \mu_{t|t-1}) \sim N(0, V_t)$. Consequently, assuming $\mu_{1|0}$ and $\Sigma_{1|0}$ are known, and taking the logarithms, we have

$$\ln[L(\sigma_e, \sigma_\eta)] = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left(\ln(V_t) + \frac{v_t^2}{V_t} \right), \qquad (11.23)$$

which involves v_t and V_t. Therefore, the log likelihood function, including cases with missing values, can be evaluated recursively via the Kalman filter. Many software packages perform state-space model estimation via a Kalman filter algorithm such as Matlab, RATS, and S-Plus. In this chapter, we use the *SsfPack* program developed by Koopman, Shephard, and Doornik (1999) and available in S-Plus and OX. Both *SsfPack* and OX are free and can be downloaded from their Web sites.

Table 11.1. State-Space Form and Notation in S-Plus

State-Space Parameter	S-Plus Name
δ	`mDelta`
Φ	`mPhi`
Ω	`mOmega`
Σ	`mSigma`

Table 11.2. Some Commands of *SsfPack* Package

Command	Function
`SsfFit`	Maximum likelihood estimation
`CheckSsf`	Create ``Ssf'' object in S-Plus
`KalmanFil`	Perform Kalman filtering
`KalmanSmo`	Perform state smoothing
`SsfMomentEst` with task ``STFIL''	Compute filtered state and variance
`SsfMomentEst` with task ``STSMO''	Compute smoothed state and variance
`SsfCondDens` with task ``STSMO''	Compute smoothed state without variance

11.1.8 S-Plus Commands Used

We provide here the *SsfPack* commands used to perform analysis of the daily realized volatility of Alcoa stock returns. Only brief explanations are given. For further details of the commands used, see Durbin and Koopman (2001, Section 6.6). S-Plus uses specific notation to specify a state-space model; see Table 11.1. The notation must be followed closely. In Table 11.2, we give some commands and their functions.

In our analysis, we first perform maximum likelihood estimation of the state-space model in Eqs. (11.1) and (11.2) to obtain estimates of σ_e and σ_η. The initial values used are $\Sigma_{1|0} = -1$ and $\mu_{1|0} = 0$, where "-1" signifies diffuse initialization, that is, $\Sigma_{1|0}$ is very large. We then treat the fitted model as given to perform Kalman filtering and state smoothing.

SsfPack and S-Plus Commands for State-Space Model

```
> da = matrix(scan(file='aa-rv-0304.txt'),2) % load data
> y = log(da[1,]) % log(RV)

> ltm.start=c(3,1)  % Initial parameter values
> P1 = -1   % Initialization of Kalman filter
> a1 = 0

> ltm.m=function(parm){    % Specify a function for the
+ sigma.eta=parm[1]        % local trend model.
+ sigma.e=parm[2]
```

```
+ ssf.m=list(mPhi=as.matrix(c(1,1)),
+ mOmega=diag(c(sigma.eta^2,sigma.e^2)),
+ mSigma=as.matrix(c(P1,a1)))
+ CheckSsf(ssf.m)
+ }
% perform estimation
> ltm.mle=SsfFit(ltm.start,y,"ltm.m",lower=c(0,0),
+ upper=c(100,100))
> ltm.mle$parameters
[1] 0.07350827 0.48026284
> sigma.eta=ltm.mle$parameter[1]
> sigma.eta
[1] 0.07350827
> sigma.e=ltm.mle$parameters[2]
> sigma.e
[1] 0.4802628
% Specify a state-space model in S-Plus.
> ssf.ltm.list=list(mPhi=as.matrix(c(1,1)),
+ mOmega=diag(c(sigma.eta^2,sigma.e^2)),
+ mSigma=as.matrix(c(P1,a1)))
% check validity of the specified model.
> ssf.ltm=CheckSsf(ssf.ltm.list)
> ssf.ltm
$mPhi:
     [,1]
[1,]    1
[2,]    1
$mOmega:
           [,1]      [,2]
[1,] 0.0054035 0.0000000
[2,] 0.0000000 0.2306524
$mSigma:
     [,1]
[1,]   -1
[2,]    0
$mDelta:
     [,1]
[1,]    0
[2,]    0
$mJPhi:
[1] 0
$mJOmega:
[1] 0
$mJDelta:
[1] 0
$mX:
[1] 0
$cT:
[1] 0
```



```
$cX:
[1] 0
$cY:
[1] 1
$cSt:
[1] 1
attr(, "class"):
[1] "ssf"
% Apply Kalman filter
> KalmanFil.ltm=KalmanFil(y,ssf.ltm,task="STFIL")
> names(KalmanFil.ltm)
 [1] "mOut"         "innov" "std.innov" "mGain" "loglike"
 [6] "loglike.conc" "dVar"  "mEst"      "mOffP" "task"
[11] "err"          "call"
> par(mfcol=c(2,1))    % Obtain plot
> plot(1:340,KalmanFil.ltm$mEst[,1],xlab='day',
+ ylab='filtered state',type='l')
> title(main='(a) Filtered state variable')
> plot(1:340,KalmanFil.ltm$mOut[,1],xlab='day',
+ ylab='v(t)',type='l')
> title(main='(b) Prediction error')
% Obtain residuals and their variances
> KalmanSmo.ltm=KalmanSmo(KalmanFil.ltm,ssf.ltm)
> names(KalmanSmo.ltm)
[1] "state.residuals"    "response.residuals" "state.variance"
[4] "response.variance" "aux.residuals"      "scores"
[7] "call"
% Next, filtered states
> FiledEst.ltm=SsfMomentEst(y,ssf.ltm,task="STFIL")
> names(FiledEst.ltm)
[1] "state.moment"      "state.variance"  "response.moment"
[4] "response.variance" "task"
% Smoothed states
> SmoedEst.ltm=SsfMomentEst(y,ssf.ltm,task="STSMO")
> names(SmoedEst.ltm)
[1] "state.moment"      "state.variance"  "response.moment"
[4] "response.variance" "task"
% Obtain plots of filtered and smoothed states with 95% C.I.
> up=FiledEst.ltm$state.moment +
+  2*sqrt(FiledEst.ltm$state.variance)
> lw=FiledEst.ltm$state.moment -
+  2*sqrt(FiledEst.ltm$state.variance)
> par(mfcol=c(1,1))
> plot(1:340,FiledEst.ltm$state.moment,type='l',xlab='day',
+ ylab='value',ylim=c(-0.1,2.5))
> lines(1:340,up,lty=2)
> lines(1:340,lw,lty=2)
> title(main='Filed state variable')
> up=SmoedEst.ltm$state.moment +
```

```
+   2*sqrt(SmoedEst.ltm$state.variance)
> lw=SmoedEst.ltm$state.moment -
+   2*sqrt(SmoedEst.ltm$state.variance)
> plot(1:340,SmoedEst.ltm$state.moment,type='l',xlab='day',
+   ylab='value',ylim=c(-0.1,2.5))
> lines(1:340,up,lty=2)
> lines(1:340,lw,lty=2)
> title(main='Smoothed state variable')
% Model checking
> resi=KalmanFil.ltm$mOut[,1]*sqrt(KalmanFil.ltm$mOut[,3])
> archTest(resi)
> autocorTest(resi)
```

For the daily realized volatility of Alcoa stock returns, the fitted local trend model is adequate based on residual analysis. Specifically, given the parameter estimates, we use the Kalman filter to obtain the 1-step ahead forecast error v_t and its variance V_t. We then compute the standardized forecast error $\tilde{v}_t = v_t/\sqrt{V_t}$ and check the serial correlations and ARCH effects of $\{\tilde{v}_t\}$. We found that $Q(25) = 23.37(0.56)$ for the standardized forecast errors and the LM test statistic for ARCH effect is 18.48(0.82) for 25 lags, where the number in parentheses denotes p-value.

11.2 LINEAR STATE-SPACE MODELS

We now consider the general state-space model. Many dynamic time series models in economics and finance can be represented in state-space form. Examples include the ARIMA models, dynamic linear models with unobserved components, time-varying regression models, and stochastic volatility models. A general Gaussian linear state-space model assumes the form

$$\boldsymbol{s}_{t+1} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, \tag{11.24}$$

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t \boldsymbol{s}_t + \boldsymbol{e}_t, \tag{11.25}$$

where $\boldsymbol{s}_t = (s_{1t}, \ldots, s_{mt})'$ is an m-dimensional state vector, $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ is a k-dimensional observation vector, $\boldsymbol{d}_t$ and $\boldsymbol{c}_t$ are m- and k-dimensional deterministic vectors, $\boldsymbol{T}_t$ and $\boldsymbol{Z}_t$ are $m \times m$ and $k \times m$ coefficient matrices, $\boldsymbol{R}_t$ is an $m \times n$ matrix often consisting of a subset of columns of the $m \times m$ identity matrix, and $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{e}_t\}$ are n- and k-dimensional Gaussian white noise series such that

$$\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \quad \boldsymbol{e}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t),$$

where $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ are positive-definite matrices. We assume that $\{\boldsymbol{e}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are independent, but this condition can be relaxed if necessary. The initial state $\boldsymbol{s}_1$ is

$N(\boldsymbol{\mu}_{1|0}, \boldsymbol{\Sigma}_{1|0})$, where $\boldsymbol{\mu}_{1|0}$ and $\boldsymbol{\Sigma}_{1|0}$ are given, and is independent of $\boldsymbol{e}_t$ and $\boldsymbol{\eta}_t$ for $t > 0$.

Equation (11.25) is the *measurement* or *observation* equation that relates the vector of observations $\boldsymbol{y}_t$ to the state vector $\boldsymbol{s}_t$, the explanatory variable $\boldsymbol{c}_t$, and the measurement error $\boldsymbol{e}_t$. Equation (11.24) is the *state* or *transition* equation that describes a first-order Markov chain to govern the state transition with innovation $\boldsymbol{\eta}_t$. The matrices $\boldsymbol{T}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Q}_t$, $\boldsymbol{Z}_t$, and $\boldsymbol{H}_t$ are known and referred to as *system matrices*. These matrices are often sparse, and they can be functions of some parameters $\boldsymbol{\theta}$, which can be estimated by the maximum likelihood method.

The state-space model in Eqs. (11.24) and (11.25) can be rewritten in a compact form as

$$\begin{bmatrix} \boldsymbol{s}_{t+1} \\ \boldsymbol{y}_t \end{bmatrix} = \boldsymbol{\delta}_t + \boldsymbol{\Phi}_t \boldsymbol{s}_t + \boldsymbol{u}_t, \tag{11.26}$$

where

$$\boldsymbol{\delta}_t = \begin{bmatrix} \boldsymbol{d}_t \\ \boldsymbol{c}_t \end{bmatrix}, \quad \boldsymbol{\Phi}_t = \begin{bmatrix} \boldsymbol{T}_t \\ \boldsymbol{Z}_t \end{bmatrix}, \quad \boldsymbol{u}_t = \begin{bmatrix} \boldsymbol{R}_t \boldsymbol{\eta}_t \\ \boldsymbol{e}_t \end{bmatrix},$$

and $\{\boldsymbol{u}_t\}$ is a sequence of Gaussian white noises with mean zero and covariance matrix

$$\boldsymbol{\Omega}_t = \text{Cov}(\boldsymbol{u}_t) = \begin{bmatrix} \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_t \end{bmatrix}.$$

The case of diffuse initialization is achieved by using

$$\boldsymbol{\Sigma}_{1|0} = \boldsymbol{\Sigma}_* + \lambda \boldsymbol{\Sigma}_\infty,$$

where $\boldsymbol{\Sigma}_*$ and $\boldsymbol{\Sigma}_\infty$ are $m \times m$ symmetric positive-definite matrices and λ is a large real number, which can approach infinity. In S-Plus and *SsfPack*, the notation

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1|0} \\ \boldsymbol{\mu}_{1|0}' \end{bmatrix}_{(m+1)\times m}$$

is used; see the notation in Table 11.1.

In many applications, the system matrices are time-invariant. However, these matrices can be time-varying, making the state-space model flexible.

11.3 MODEL TRANSFORMATION

To appreciate the flexibility of the state-space model, we rewrite some well-known econometric and financial models in state-space form.

11.3.1 CAPM with Time-Varying Coefficients

First, consider the capital asset pricing model (CAPM) with time-varying intercept and slope. The model is

$$\begin{aligned} r_t &= \alpha_t + \beta_t r_{M,t} + e_t, \quad e_t \sim N(0, \sigma_e^2), \\ \alpha_{t+1} &= \alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \\ \beta_{t+1} &= \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \end{aligned} \tag{11.27}$$

where r_t is the excess return of an asset, $r_{M,t}$ is the excess return of the market, and the innovations $\{e_t, \eta_t, \epsilon_t\}$ are mutually independent. This CAPM allows for time-varying α and β that evolve as a random walk over time. We can easily rewrite the model as

$$\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix},$$

$$r_t = [1, r_{M,t}] \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + e_t.$$

Thus, the time-varying CAPM is a special case of the state-space model with $\boldsymbol{s}_t = (\alpha_t, \beta_t)'$, $\boldsymbol{T}_t = \boldsymbol{R}_t = \boldsymbol{I}_2$, the 2×2 identity matrix, $\boldsymbol{d}_t = \boldsymbol{0}$, $c_t = 0$, $\boldsymbol{Z}_t = (1, r_{M,t})$, $\boldsymbol{H}_t = \sigma_e^2$, and $\boldsymbol{Q}_t = \text{diag}\{\sigma_\eta^2, \sigma_\epsilon^2\}$. Furthermore, in the form of Eq. (11.26), we have $\boldsymbol{\delta}_t = \boldsymbol{0}$, $\boldsymbol{u}_t = (\eta_t, \epsilon_t, e_t)'$,

$$\boldsymbol{\Phi}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & r_{M,t} \end{bmatrix}, \quad \boldsymbol{\Omega}_t = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}.$$

If diffuse initialization is used, then

$$\boldsymbol{\Sigma} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}.$$

SsfPack/S-Plus Specification of Time-Varying Models

For the CAPM in Eq. (11.27), $\boldsymbol{\Phi}_t$ contains $r_{M,t}$, which is time-varying. Some special input is required to specify such a model in *SsfPack*. Basically, it requires two additional variables: (a) a data matrix $\boldsymbol{X}$ that stores $\boldsymbol{Z}_t$ and (b) an index matrix for $\boldsymbol{\Phi}_t$ that identifies $\boldsymbol{Z}_t$ from the data matrix. The notation for index matrices of the state-space model in Eq. (11.26) is given in Table 11.3. Note that the matrix $\boldsymbol{J}_\Phi$ must have the same dimension as $\boldsymbol{\Phi}_t$. The elements of $\boldsymbol{J}_\Phi$ are all set to "-1" except the elements for which the corresponding elements of $\boldsymbol{\Phi}_t$ are time-varying. The non-negative index value of $\boldsymbol{J}_\Phi$ indicates the column of the data matrix $\boldsymbol{X}$, which contains the time-varying values.

Table 11.3. Notation and Name Used in *SsfPack*/S-Plus for Time-Varying State-Space Model

Index Matrix	Name Used in *SsfPack*/S-Plus
$\boldsymbol{J}_\delta$	mJDelta
$\boldsymbol{J}_\Phi$	mJPhi
$\boldsymbol{J}_\Omega$	mJOmega
Time-Varying Data Matrix	**Name Used in *SsfPack*/S-Plus**
$\boldsymbol{X}$	mX

To illustrate, consider the monthly simple excess returns of General Motors stock from January 1990 to December 2003 used in Chapter 9. The monthly simple excess return of the S&P 500 composite index is used as the market return. The specification of a time-varying CAPM requires values of the variances σ_η^2, σ_ϵ^2, and σ_e^2. Suppose that $(\sigma_\eta, \sigma_\epsilon, \sigma_e) = (0.02, 0.04, 0.1)$. The state-space specification for the CAPM under *SsfPack*/S-Plus is given below:

```
> X.mtx=cbind(1,sp) % Here 'sp' is the market excess returns.
> Phi.t = rbind(diag(2),rep(0,2))
> Sigma=-Phi.t
> sigma.eta=.02
> sigma.ep=.04
> sigma.e=.1
> Omega=diag(c(sigma.eta^2,sigma.ep^2,sigma.e^2))
> JPhi = matrix(-1,3,2) % Create a 3-by-2 matrix of -1.
> JPhi[3,1]=1
> JPhi[3,2]=2
> ssf.tv.capm=list(mPhi=Phi.t,
+ mOmega=Omega,
+ mJPhi=JPhi,
+ mSigma=Sigma,
+ mX=X.mtx)
> ssf.tv.capm
$mPhi:
     [,1] [,2]
[1,]    1    0
[2,]    0    1
[3,]    0    0
$mOmega:
      [,1]    [,2] [,3]
[1,] 4e-04 0.0000 0.00
[2,] 0e+00 0.0016 0.00
[3,] 0e+00 0.0000 0.01
$mJPhi:
     [,1] [,2]
```

```
[1,]   -1   -1
[2,]   -1   -1
[3,]    1    2
 $mSigma:
     [,1] [,2]
[1,]   -1    0
[2,]    0   -1
[3,]    0    0
 $mX:
numeric matrix: 168 rows, 2 columns.
              sp
 [1,] 1 -0.075187
  ...
[168,] 1  0.05002
```

11.3.2 ARMA Models

Consider a zero-mean ARMA(p, q) process y_t of Chapter 2,

$$\phi(B)y_t = \theta(B)a_t, \quad a_t \sim N(0, \sigma_a^2), \tag{11.28}$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$, and p and q are nonnegative integers. There are many ways to transform such an ARMA model into a state-space form. We discuss three methods available in the literature. Let $m = \max(p, q+1)$ and rewrite the ARMA model in Eq. (11.28) as

$$y_t = \sum_{i=1}^{m} \phi_i y_{t-i} + a_t - \sum_{j=1}^{m-1} \theta_j a_{t-j}, \tag{11.29}$$

where $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$. In particular, $\theta_m = 0$ because $m > q$.

Akaike's Approach

Akaike (1975) defines the state vector $\boldsymbol{s}_t$ as the minimum collection of variables that contains all the information needed to produce forecasts at the forecast origin t. It turns out that, for the ARMA process in Eq. (11.28) with $m = \max(p, q+1)$, $\boldsymbol{s}_t = (y_{t|t}, y_{t+1|t}, \ldots, y_{t+m-1|t})'$, where $y_{t+j|t} = E(y_{t+j}|F_t)$ is the conditional expectation of y_{t+j} given $F_t = \{y_1, \ldots, y_t\}$. Since $y_{t|t} = y_t$, the first element of $\boldsymbol{s}_t$ is y_t. Thus, the observation equation is

$$y_t = \boldsymbol{Z}\boldsymbol{s}_t, \tag{11.30}$$

where $\boldsymbol{Z} = (1, 0, \ldots, 0)_{1\times m}$. We derive the transition equation in several steps. First, from the definition,

$$s_{1,t+1} = y_{t+1} = y_{t+1|t} + (y_{t+1} - y_{t+1|t}) = s_{2t} + a_{t+1}, \tag{11.31}$$

where s_{it} is the ith element of $\boldsymbol{s}_t$. Next, consider the MA representation of ARMA models given in Chapter 2. That is,

$$y_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

where $\psi_0 = 1$ and other ψ-weights can be obtained by equating coefficients of B^i in $1 + \sum_{i=1}^{\infty} \psi_i B^i = \theta(B)/\phi(B)$. In particular, we have

$$\begin{aligned}
\psi_1 &= \phi_1 - \theta_1, \\
\psi_2 &= \phi_1\psi_1 + \phi_2 - \theta_2, \\
&\vdots = \vdots \\
\psi_{m-1} &= \phi_1\psi_{m-2} + \phi_2\psi_{m-3} + \cdots + \phi_{m-2}\psi_1 + \phi_{m-1} - \theta_{m-1} \\
&= \sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1}.
\end{aligned} \tag{11.32}$$

Using the MA representation, we have, for $j > 0$,

$$\begin{aligned}
y_{t+j|t} = E(y_{t+j}|F_t) &= E\left(\sum_{i=0}^{\infty} \psi_i a_{t+j-i}|F_t\right) \\
&= \psi_j a_t + \psi_{j+1} a_{t-1} + \psi_{j+2} a_{t-2} + \cdots
\end{aligned}$$

and

$$\begin{aligned}
y_{t+j|t+1} = E(y_{t+j}|F_{t+1}) &= \psi_{j-1} a_{t+1} + \psi_j a_t + \psi_{j+1} a_{t-1} + \cdots \\
&= \psi_{j-1} a_{t+1} + y_{t+j|t}.
\end{aligned}$$

Thus, for $j > 0$, we have

$$y_{t+j|t+1} = y_{t+j|t} + \psi_{j-1} a_{t+1}. \tag{11.33}$$

This result is referred to as the forecast updating formula of ARMA models. It provides a simple way to update the forecast from origin t to origin $t+1$ when y_{t+1} becomes available. The new information of y_{t+1} is contained in the innovation a_{t+1}, and the time-t forecast is revised based on this new information with weight ψ_{j-1} to compute the time-$(t+1)$ forecast.

Finally, from Eq. (11.29) and using $E(a_{t+j}|F_{t+1}) = 0$ for $j > 1$, we have

$$y_{t+m|t+1} = \sum_{i=1}^{m} \phi_i y_{t+m-i|t+1} - \theta_{m-1} a_{t+1}.$$

Taking Eq. (11.33), the prior equation becomes

$$\begin{aligned} y_{t+m|t+1} &= \sum_{i=1}^{m-1} \phi_i (y_{t+m-i|t} + \psi_{m-i-1} a_{t+1}) + \psi_m y_{t|t} - \theta_{m-1} a_{t+1} \\ &= \sum_{i=1}^{m} \phi_i y_{t+m-i|t} + \left(\sum_{i=1}^{m-1} \phi_i \psi_{m-1-i} - \theta_{m-1} \right) a_{t+1} \\ &= \sum_{i=1}^{m} \phi_i y_{t+m-i|t} + \psi_{m-1} a_{t+1}, \end{aligned} \tag{11.34}$$

where the last equality uses Eq. (11.32). Combining Eqs. (11.31), (11.33) for $j = 2, \ldots, m-1$, and (11.34) together, we have

$$\begin{bmatrix} y_{t+1} \\ y_{t+2|t+1} \\ \vdots \\ y_{t+m-1\,t+1} \\ y_{t+m|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+m-2|t} \\ y_{t+m-1|t} \end{bmatrix} + \begin{bmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-2} \\ \psi_{m-1} \end{bmatrix} a_{t+1}. \tag{11.35}$$

Thus, the transition equation of Akaike's approach is

$$\boldsymbol{s}_{t+1} = \boldsymbol{T}\boldsymbol{s}_t + \boldsymbol{R}\eta_t, \quad \eta_t \sim N(0, \sigma_a^2), \tag{11.36}$$

where $\eta_t = a_{t+1}$, and $\boldsymbol{T}$ and $\boldsymbol{R}$ are the coefficient matrices in Eq. (11.35).

Harvey's Approach

Harvey (1993, Section 4.4) provides a state-space form with an m-dimensional state vector s_t, the first element of which is y_t, that is, $s_{1t} = y_t$. The other elements of s_t are obtained recursively. From the ARMA$(m, m-1)$ model, we have

$$\begin{aligned} y_{t+1} &= \phi_1 y_t + \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j} + a_{t+1} \\ &\equiv \phi_1 s_{1t} + s_{2t} + \eta_t, \end{aligned}$$

where $s_{2t} = \sum_{i=2}^{m} \phi_i y_{t+1-i} - \sum_{j=1}^{m-1} \theta_j a_{t+1-j}$, $\eta_t = a_{t+1}$, and as defined earlier $s_{1t} = y_t$. Focusing on $s_{2,t+1}$, we have

$$\begin{aligned} s_{2,t+1} &= \sum_{i=2}^{m} \phi_i y_{t+2-i} - \sum_{j=1}^{m-1} \theta_j a_{t+2-j} \\ &= \phi_2 y_t + \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j} - \theta_1 a_{t+1} \\ &\equiv \phi_2 s_{1t} + s_{3t} + (-\theta_1)\eta_t, \end{aligned}$$

where $s_{3t} = \sum_{i=3}^{m} \phi_i y_{t+2-i} - \sum_{j=2}^{m-1} \theta_j a_{t+2-j}$. Next, considering $s_{3,t+1}$, we have

$$\begin{aligned} s_{3,t+1} &= \sum_{i=3}^{m} \phi_i y_{t+3-i} - \sum_{j=2}^{m-1} \theta_j a_{t+3-j} \\ &= \phi_3 y_t + \sum_{i=4}^{m} \phi_i y_{t+3-i} - \sum_{j=3}^{m-1} \theta_j a_{t+3-j} + (-\theta_2) a_{t+1} \\ &\equiv \phi_3 s_{1t} + s_{4t} + (-\theta_2)\eta_t, \end{aligned}$$

where $s_{4t} = \sum_{i=4}^{m} \phi_i y_{t+3-i} - \sum_{j=3}^{m-1} \theta_j a_{t+3-j}$. Repeating the procedure, we have $s_{mt} = \sum_{i=m}^{m} \phi_i y_{t+m-1-i} - \sum_{j=m-1}^{m-1} \theta_j a_{t+m-1-j} = \phi_m y_{t-1} - \theta_{m-1} a_t$. Finally,

$$\begin{aligned} s_{m,t+1} &= \phi_m y_t - \theta_{m-1} a_{t+1} \\ &= \phi_m s_{1t} + (-\theta_{m-1})\eta_t. \end{aligned}$$

Putting the prior equations together, we have a state-space form

$$\boldsymbol{s}_{t+1} = \boldsymbol{T}\boldsymbol{s}_t + \boldsymbol{R}\eta_t, \quad \eta_t \sim N(0, \sigma_a^2), \tag{11.37}$$

$$y_t = \boldsymbol{Z}\boldsymbol{s}_t, \tag{11.38}$$

where the system matrices are time-invariant defined as $\boldsymbol{Z} = (1, 0, \ldots, 0)_{1\times m}$,

$$\boldsymbol{T} = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ \phi_{m-1} & 0 & 0 & \cdots & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{m-1} \end{bmatrix},$$

and $\boldsymbol{d}_t$, $\boldsymbol{c}_t$, and $\boldsymbol{H}_t$ are all zero. The model in Eqs. (11.37) and (11.38) has no measurement errors. It has an advantage that the AR and MA coefficients are directly used in the system matrices.

Aoki's Approach

Aoki (1987, Chapter 4) discusses several ways to convert an ARMA model into a state-space form. First, consider the MA model, that is, $y_t = \theta(B)a_t$. In this case, we can simply define $\boldsymbol{s}_t = (a_{t-q}, a_{t-q+2}, \ldots, a_{t-1})'$ and obtain the state-space form

$$\begin{bmatrix} a_{t-q+1} \\ a_{t-q+2} \\ \vdots \\ a_{t-1} \\ a_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{t-q} \\ a_{t-q+1} \\ \vdots \\ a_{t-2} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_t,$$

$$y_t = (-\theta_q, -\theta_{q-1}, \ldots, -\theta_1)\boldsymbol{s}_t + a_t, \tag{11.39}$$

Note that, in this particular case, a_t appears in both state and measurement equations.

Next, consider the AR model, that is, $\phi(B)z_t = a_t$. Aoki (1987) introduces two methods. The first method is a straightforward one by defining $\boldsymbol{s}_t = (z_{t-p+1}, \ldots, z_t)'$ to obtain

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_{t+2} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t+1} \\ z_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} a_{t+1},$$

$$z_t = (0, 0, \cdots, 0, 1)\boldsymbol{s}_t. \tag{11.40}$$

The second method defines the state vector in the same way as the first method except that a_t is removed from the last element; that is, $\boldsymbol{s}_t = z_t - a_t$ if $p = 1$ and $\boldsymbol{s}_t = (z_{t-p+1}, \ldots, z_{t-1}, z_t - a_t)'$ if $p > 1$. Simple algebra shows that

$$\begin{bmatrix} z_{t-p+2} \\ z_{t-p+3} \\ \vdots \\ z_t \\ z_{t+1} - a_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 \end{bmatrix} \begin{bmatrix} z_{t-p+1} \\ z_{t-p+2} \\ \vdots \\ z_{t-1} \\ z_t - a_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \phi_1 \end{bmatrix} a_t,$$

$$z_t = (0, 0, \ldots, 0, 1)\boldsymbol{s}_t + a_t. \tag{11.41}$$

Again, a_t appears in both transition and measurement equations.

Turn to the ARMA(p, q) model $\phi(B)y_t = \theta(B)a_t$. For simplicity, we assume $q < p$ and introduce an auxiliary variable $z_t = [1/\phi(B)]a_t$. Then, we have

$$\phi(B)z_t = a_t, \quad y_t = \theta(B)z_t.$$

Since z_t is an AR(p) model, we can use the transition equation in Eq. (11.40) or Eq. (11.41). If Eq. (11.40) is used, we can use $y_t = \theta(B)z_t$ to construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \ldots, -\theta_1, 1)\boldsymbol{s}_t, \tag{11.42}$$

where it is understood that $p > q$ and $\theta_j = 0$ for $j > q$. On the other hand, if Eq. (11.41) is used as the transition equation, we construct the measurement equation as

$$y_t = (-\theta_{p-1}, -\theta_{p-2}, \ldots, -\theta_1, 1)\boldsymbol{s}_t + a_t. \tag{11.43}$$

In summary, there are many state-space representations for an ARMA model. Each representation has its pros and cons. For estimation and forecasting purposes, one can choose any one of those representations. On the other hand, for a time-invariant coefficient state-space model in Eqs. (11.24) and (11.25), one can use the Cayley–Hamilton theorem to show that the observation y_t follows an ARMA(m, m) model, where m is the dimension of the state vector.

SsfPack Command

In *SsfPack*/S-Plus, a command `GetSsfArma` can be used to transform an ARMA model into a state-space form. Harvey's approach is used. To illustrate, consider the AR(1) model

$$y_t = 0.6y_{t-1} + a_t, \quad a_t \sim N(0, 0.4^2).$$

The state-space form of the model is

```
> ssf.ar1 = GetSsfArma(ar=0.6,sigma=0.4)
> ssf.ar1
$mPhi:
     [,1]
[1,]  0.6
[2,]  1.0
$mOmega:
     [,1] [,2]
[1,] 0.16    0
[2,] 0.00    0
$mSigma:
     [,1]
[1,] 0.25
[2,] 0.00
```

Since the AR(1) model is stationary, the program uses $\boldsymbol{\Sigma}_{1|0} = \text{Var}(y_t) = (0.4)^2/(1 - 0.6^2) = 0.25$ and $\boldsymbol{\mu}_{1|0} = 0$. These values appear in the matrix `mSigma`.

As a second example, consider the ARMA(2,1) model

$$y_t = 1.2y_{t-1} - 0.35y_{t-2} + a_t - 0.25a_{t-1}, \quad a_t \sim N(0, 1.1^2).$$

The state-space form of the model is

```
> arma21.m = list(ar=c(1.2,-0.35),ma=c(-0.25),sigma=1.1)
> ssf.arma21= GetSsfArma(model=arma21.m)
> ssf.arma21
$mPhi:
     [,1] [,2]
[1,]  1.20    1
[2,] -0.35    0
[3,]  1.00    0
$mOmega:
        [,1]      [,2] [,3]
[1,]  1.2100 -0.302500    0
[2,] -0.3025  0.075625    0
[3,]  0.0000  0.000000    0
$mSigma:
          [,1]        [,2]
```

```
[1,]  4.060709 -1.4874057
[2,] -1.487406  0.5730618
[3,]  0.000000  0.0000000
```

As expected, the output shows that

$$\boldsymbol{T} = \begin{bmatrix} 1.2 & 1 \\ -0.35 & 0 \end{bmatrix}, \quad \boldsymbol{Z} = (1, 0),$$

and `mPhi` and `mOmega` follow the format of Eq. (11.26), and the covariance matrix of $(y_t, y_{t-1})'$ is used in `mSigma`. Note that in *SsfPack*, the MA polynomial of an ARMA model assumes the form $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, not the form $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ commonly used in the literature.

11.3.3 Linear Regression Model

Multiple linear regression models can also be represented in state-space form. Consider the model

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t, \quad e_t \sim N(0, \sigma_e^2),$$

where $\boldsymbol{x}_t$ is a p-dimensional explanatory variable and $\boldsymbol{\beta}$ is a p-dimensional parameter vector. Let $s_t = \boldsymbol{\beta}$ for all t. Then the model can be written as

$$\begin{bmatrix} s_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_p \\ \boldsymbol{x}_t' \end{bmatrix} s_t + \begin{bmatrix} \boldsymbol{0}_p \\ e_t \end{bmatrix}. \tag{11.44}$$

Thus, the system matrices are $\boldsymbol{T}_t = \boldsymbol{I}_p$, $\boldsymbol{Z}_t = \boldsymbol{x}_t'$, $\boldsymbol{d}_t = \boldsymbol{0}$, $c_t = 0$, $\boldsymbol{H}_t = \boldsymbol{0}$, and $\boldsymbol{Q}_t = \sigma_e^2$. Since the state vector is fixed, a diffuse initialization should be used.

One can extend the regression model so that $\boldsymbol{\beta}_t$ is random, say,

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \boldsymbol{R}_t \eta_t, \quad \eta_t \sim N(0, 1),$$

and $\boldsymbol{R}_t = (\sigma_1, \ldots, \sigma_p)'$ with $\sigma_i \geq 0$. If $\sigma_i = 0$, then β_i is time-invariant.

SsfPack Command

In *SsfPack*, the command `GetSsfReg` creates a state-space form for the multiple linear regression model. The command has an input argument that contains the data matrix of explanatory variables. To illustrate, consider the simple market model

$$r_t = \beta_0 + \beta_1 r_{M,t} + e_t, \quad t = 1, \ldots, 168,$$

where r_t is the return of an asset and $r_{M,t}$ is the market return, for example, the S&P 500 composite index return. The state-space form can be obtained as


```
> ssf.reg=GetSsfReg(cbind(1,sp))  % 'sp' is market return.
> ssf.reg
$mPhi:
     [,1] [,2]
[1,]    1    0
[2,]    0    1
[3,]    0    0
$mOmega:
     [,1] [,2] [,3]
[1,]    0    0    0
[2,]    0    0    0
[3,]    0    0    1
$mSigma:
     [,1] [,2]
[1,]   -1    0
[2,]    0   -1
[3,]    0    0
$mJPhi:
     [,1] [,2]
[1,]   -1   -1
[2,]   -1   -1
[3,]    1    2
$mX:
numeric matrix: 168 rows, 2 columns.
             sp
 [1,] 1 -0.075187
...
[168,] 1  0.05002
```

11.3.4 Linear Regression Models with ARMA Errors

Consider the regression model with ARMA(p, q) errors,

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + z_t, \quad \phi(B) z_t = \theta(B) a_t, \tag{11.45}$$

where $a_t \sim N(0, \sigma_a^2)$ and $\boldsymbol{x}_t$ is a k-dimensional vector of explanatory variables. A special case of this model is the nonzero mean ARMA(p, q) model in which $\boldsymbol{x}_t = 1$ for all t and $\boldsymbol{\beta}$ becomes a scalar parameter. Let s_t be a state vector for the z_t series, for example, that defined in Eq. (11.37). We can define a state vector s_t^* for y_t as

$$s_t^* = \begin{bmatrix} s_t \\ \boldsymbol{\beta}_t \end{bmatrix}, \tag{11.46}$$

where $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ for all t. Then, a state-space form for y_t is

$$s_{t+1}^* = \boldsymbol{T}^* s_t^* + \boldsymbol{R}^* \eta_t, \tag{11.47}$$

$$y_t = \boldsymbol{Z}_t^* s_t^*, \tag{11.48}$$

where $\boldsymbol{Z}_t^* = (1, 0, \ldots, 0, \boldsymbol{x}_t')_{1\times(m+k)}$, $m = \max(p, q+1)$, and

$$\boldsymbol{T}^* = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k \end{bmatrix}, \quad \boldsymbol{R}^* = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix},$$

where $\boldsymbol{T}$ and $\boldsymbol{R}$ are defined in Eq. (11.37). In a compact form, we have the state-space model

$$\begin{bmatrix} \boldsymbol{s}_{t+1}^* \\ y_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}^* \\ \boldsymbol{Z}_t^* \end{bmatrix} \boldsymbol{s}_t^* + \begin{bmatrix} \boldsymbol{R}^* \eta_t \\ 0 \end{bmatrix}.$$

SsfPack Command

SsfPack uses the command `GetSsfRegArma` to construct a state-space form for linear regression models with ARMA errors. The arguments of the command can be found using the command `args(GetSsfRegArma)`. They consist of a data matrix for the explanatory variables and ARMA model specification. To illustrate, consider the model

$$y_t = \beta_0 + \beta_1 x_t + z_t, \quad t = 1, \ldots, 168,$$
$$z_t = 1.2z_{t-1} - 0.35z_{t-2} + a_t - 0.25a_{t-1}, \quad a_t \sim N(0, \sigma_a^2).$$

We use the notation `X` to denote the $T \times 2$ matrix of regressors $(1, x_t)$. A state-space form for the prior model can be obtained as

```
> ssf.reg.arma21=GetSsfRegArma(X,ar=c(1.2,-0.35),
+ ma=c(-0.25))
> ssf.reg.arma21
$mPhi:
      [,1] [,2] [,3] [,4]
[1,]  1.20    1    0    0
[2,] -0.35    0    0    0
[3,]  0.00    0    1    0
[4,]  0.00    0    0    1
[5,]  1.00    0    0    0
$mOmega:
      [,1]    [,2] [,3] [,4] [,5]
[1,]  1.00 -0.2500    0    0    0
[2,] -0.25  0.0625    0    0    0
[3,]  0.00  0.0000    0    0    0
[4,]  0.00  0.0000    0    0    0
[5,]  0.00  0.0000    0    0    0
$mSigma:
         [,1]      [,2] [,3] [,4]
[1,]  3.35595 -1.229260    0    0
[2,] -1.22926  0.473604    0    0
[3,]  0.00000  0.000000   -1    0
[4,]  0.00000  0.000000    0   -1
```



```
[5,]  0.00000  0.000000    0    0
$mJPhi:
     [,1] [,2] [,3] [,4]
[1,]   -1   -1   -1   -1
[2,]   -1   -1   -1   -1
[3,]   -1   -1   -1   -1
[4,]   -1   -1   -1   -1
[5,]   -1   -1    1    2
$mX:
numeric matrix: 168 rows, 2 columns.
          xt
  [1,] 1 0.4993
...
[168,] 1 0.7561
```

11.3.5 Scalar Unobserved Component Model

The basic univariate unobserved component model, or the *structural time series model* (STSM), assumes the form

$$y_t = \mu_t + \gamma_t + \varpi_t + e_t, \tag{11.49}$$

where μ_t, γ_t, and ϖ_t represent the unobserved *trend, seasonal,* and *cycle* components, respectively, and e_t is the unobserved *irregular* component. In the literature, a nonstationary (possibly double-unit-root) model is commonly used for the trend component:

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \varsigma_t, \quad \varsigma_t \sim N(0, \sigma_\varsigma^2), \end{aligned} \tag{11.50}$$

where $\mu_1 \sim N(0, \xi)$ and $\beta_1 \sim N(0, \xi)$ with ξ a large real number, for example, $\xi = 10^8$. See, for instance, Kitagawa and Gersch (1996). If $\sigma_\varsigma = 0$, then μ_t follows a random walk with drift β_1. If $\sigma_\varsigma = \sigma_\eta = 0$, then μ_t represents a linear deterministic trend.

The seasonal component γ_t assumes the form

$$(1 + B + \cdots + B^{s-1})\gamma_t = \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2), \tag{11.51}$$

where s is the number of seasons in a year, that is, the period of the seasonality. If $\sigma_\omega = 0$, then the seasonal pattern is deterministic. The cycle component is postulated as

$$\begin{bmatrix} \varpi_{t+1} \\ \varpi_{t+1}^* \end{bmatrix} = \delta \begin{bmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{bmatrix} \begin{bmatrix} \varpi_t \\ \varpi_t^* \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon_t^* \end{bmatrix}, \tag{11.52}$$

Table 11.4. Arguments of the Command `GetSsfStsm` in *SsfPack*/S-Plus

Argument	STSM Parameter
`irregular`	σ_ε
`level`	σ_η
`slope`	σ_ζ
`seasonalDummy`	σ_ω, s
`seasonalTrig`	σ_ω, s
`seasonalHS`	σ_ω, s
`Cycle0`	σ_ε, λ_c, δ
⋮	⋮
`Cycle9`	σ_ε, λ_c, δ

where

$$\begin{bmatrix} \varepsilon_t \\ \varepsilon_t^* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_\varepsilon^2(1-\delta^2)\boldsymbol{I}_2 \right),$$

$\varpi_0 \sim N(0, \sigma_\varepsilon^2)$, $\varpi_0^* \sim N(0, \sigma_\varepsilon^2)$, and $\text{Cov}(\varpi_0, \varpi_0^*) = 0$, $\delta \in (0, 1]$ is called a *damping* factor, and the frequency of the cycle is $\lambda_c = 2\pi/q$ with q being the period. If $\delta = 1$, then the cycle becomes a deterministic sine–cosine wave.

***SsfPack*/S-Plus Command**

The command `GetSsfStsm` constructs a state-space form for the structural time series model. It allows for 10 cycle components; see the output of the command `args(GetSsfStsm)`. Table 11.4 provides a summary of the arguments and their corresponding symbols of the model. To illustrate, consider the local trend model in Eqs. (11.1) and (11.2) with $\sigma_\varepsilon = 0.4$ and $\sigma_\eta = 0.2$. This is a special case of the scalar unobserved component model. One can obtain a state-space form as

```
> ssf.stsm=GetSsfStsm(irregular=0.4,level=0.2)
> ssf.stsm
$mPhi:
     [,1]
[1,]    1
[2,]    1
$mOmega:
     [,1] [,2]
[1,] 0.04 0.00
[2,] 0.00 0.16
$mSigma:
     [,1]
[1,]   -1
[2,]    0
```


11.4 KALMAN FILTER AND SMOOTHING

In this section, we study the Kalman filter and various smoothing methods for the general state-space model in Eqs. (11.24) and (11.25). The derivation follows closely the steps taken in Section 11.1. For readers interested in applications, this section can be skipped at the first read. A good reference for this section is Durbin and Koopman (2001, Chapter 4).

11.4.1 Kalman Filter

Recall that the aim of the Kalman filter is to obtain recursively the conditional distribution of $\boldsymbol{s}_{t+1}$ given the data $F_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ and the model. Since the conditional distribution involved is normal, it suffices to study the conditional mean and covariance matrix. Let $\boldsymbol{s}_{j|i}$ and $\boldsymbol{\Sigma}_{j|i}$ be the conditional mean and covariance matrix of $\boldsymbol{s}_j$ given F_i, that is, $\boldsymbol{s}_j|F_i \sim N(\boldsymbol{s}_{j|i}, \boldsymbol{\Sigma}_{j|i})$. From Eq. (11.24),

$$\boldsymbol{s}_{t+1|t} = E(\boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t|F_t) = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t}, \tag{11.53}$$

$$\boldsymbol{\Sigma}_{t+1|t} = \text{Var}(\boldsymbol{T}_t\boldsymbol{s}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t|F_t) = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t'. \tag{11.54}$$

Similarly to that of Section 11.1, let $\boldsymbol{y}_{t|t-1}$ be the conditional mean of $\boldsymbol{y}_t$ given F_{t-1}. From Eq. (11.25),

$$\boldsymbol{y}_{t|t-1} = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}.$$

Let

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1} = \boldsymbol{y}_t - (\boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{s}_{t|t-1}) = \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t, \tag{11.55}$$

be the 1-step ahead forecast error of $\boldsymbol{y}_t$ given F_{t-1}. It is easy to see that (a) $E(\boldsymbol{v}_t|F_{t-1}) = \boldsymbol{0}$; (b) $\boldsymbol{v}_t$ is independent of F_{t-1}, that is, $\text{Cov}(\boldsymbol{v}_t, \boldsymbol{y}_j) = \boldsymbol{0}$ for $1 \le j < t$; and (c) $\{\boldsymbol{v}_t\}$ is a sequence of independent normal random vectors. Also, let $\boldsymbol{V}_t = \text{Var}(\boldsymbol{v}_t|F_{t-1}) = \text{Var}(\boldsymbol{v}_t)$ be the covariance matrix of the 1-step ahead forecast error. From Eq. (11.55), we have

$$\boldsymbol{V}_t = \text{Var}[\boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t] = \boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t. \tag{11.56}$$

Since $F_t = \{F_{t-1}, \boldsymbol{y}_t\} = \{F_{t-1}, \boldsymbol{v}_t\}$, we can apply Theorem 11.1 to obtain

$$\begin{aligned} \boldsymbol{s}_{t|t} &= E(\boldsymbol{s}_t|F_t) = E(\boldsymbol{s}_t|F_{t-1}, \boldsymbol{v}_t) \\ &= E(\boldsymbol{s}_t|F_{t-1}) + \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)[\text{Var}(\boldsymbol{v}_t)]^{-1}\boldsymbol{v}_t \\ &= \boldsymbol{s}_{t|t-1} + \boldsymbol{C}_t\boldsymbol{V}_t^{-1}\boldsymbol{v}_t, \end{aligned} \tag{11.57}$$

where $\boldsymbol{C}_t = \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t|F_{t-1})$ given by

$$\begin{aligned} \boldsymbol{C}_t &= \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t|F_{t-1}) = \text{Cov}[\boldsymbol{s}_t, \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t|F_{t-1}] \\ &= \text{Cov}[\boldsymbol{s}_t, \boldsymbol{Z}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1})|F_{t-1}] = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'. \end{aligned}$$

Here we assume that $\boldsymbol{V}_t$ is invertible, because $\boldsymbol{H}_t$ is. Using Eqs. (11.53) and (11.57), we obtain

$$\boldsymbol{s}_{t+1|t} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|t-1} + \boldsymbol{T}_t \boldsymbol{C}_t \boldsymbol{V}_t^{-1} \boldsymbol{v}_t = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|t-1} + \boldsymbol{K}_t \boldsymbol{v}_t, \tag{11.58}$$

where

$$\boldsymbol{K}_t = \boldsymbol{T}_t \boldsymbol{C}_t \boldsymbol{V}_t^{-1} = \boldsymbol{T}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1}, \tag{11.59}$$

which is the *Kalman gain* at time t. Applying Theorem 11.1(2), we have

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t} &= \mathrm{Var}(\boldsymbol{s}_t | F_{t-1}) \\
&= \mathrm{Var}(\boldsymbol{s}_t | F_{t-1}) - \mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)[\mathrm{Var}(\boldsymbol{v}_t)]^{-1} \mathrm{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_t)' \\
&= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{C}_t \boldsymbol{V}_t^{-1} \boldsymbol{C}_t' \\
&= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1} \boldsymbol{Z}_t \boldsymbol{\Sigma}_{t|t-1}.
\end{aligned} \tag{11.60}$$

Plugging Eq. (11.60) into Eq. (11.54) and using Eq. (11.59), we obtain

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{T}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{L}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t', \tag{11.61}$$

where

$$\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t.$$

Putting the prior equations together, we obtain the celebrated Kalman filter for the state-space model in Eqs. (11.24) and (11.25). Given the starting values $\boldsymbol{s}_{1|0}$ and $\boldsymbol{\Sigma}_{1|0}$, the Kalman filter algorithm is

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t \boldsymbol{s}_{t|t-1}, \\
\boldsymbol{V}_t &= \boldsymbol{Z}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t' + \boldsymbol{H}_t, \\
\boldsymbol{K}_t &= \boldsymbol{T}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1}, \\
\boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t, \\
\boldsymbol{s}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|t-1} + \boldsymbol{K}_t \boldsymbol{v}_t, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{L}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t', \quad t = 1, \ldots, T.
\end{aligned} \tag{11.62}$$

If the filtered quantities $\boldsymbol{s}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}$ are also of interest, then we modify the filter to include the contemporaneous filtering equations in Eqs. (11.57) and (11.60). The resulting algorithm is

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t \boldsymbol{s}_{t|t-1}, \\
\boldsymbol{C}_t &= \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t',
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{V}_t &= \boldsymbol{Z}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{Z}_t' + \boldsymbol{H}_t = \boldsymbol{Z}_t \boldsymbol{C}_t + \boldsymbol{H}_t, \\
\boldsymbol{s}_{t|t} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{C}_t \boldsymbol{V}_t^{-1} \boldsymbol{v}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{C}_t \boldsymbol{V}_t^{-1} \boldsymbol{C}_t', \\
\boldsymbol{s}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|t}, \\
\boldsymbol{\Sigma}_{t+1|t} &= \boldsymbol{T}_t \boldsymbol{\Sigma}_{t|t} \boldsymbol{T}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'.
\end{aligned}$$

Steady State

If the state-space model is time-invariant, that is, all system matrices are time-invariant, then the matrices $\boldsymbol{\Sigma}_{t|t-1}$ converge to a constant matrix $\boldsymbol{\Sigma}_*$, which is a solution of the matrix equation

$$\boldsymbol{\Sigma}_* = \boldsymbol{T} \boldsymbol{\Sigma}_* \boldsymbol{T}' - \boldsymbol{T} \boldsymbol{\Sigma}_* \boldsymbol{Z} \boldsymbol{V}^{-1} \boldsymbol{Z} \boldsymbol{\Sigma}_* \boldsymbol{T}' + \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R}',$$

where $\boldsymbol{V} = \boldsymbol{Z} \boldsymbol{\Sigma}_* \boldsymbol{Z}' + \boldsymbol{H}$. The solution that is reached after convergence to $\boldsymbol{\Sigma}_*$ is referred to as the *steady-state solution* of the Kalman filter. Once the steady state is reached, $\boldsymbol{V}_t$, $\boldsymbol{K}_t$, and $\boldsymbol{\Sigma}_{t+1|t}$ are all constant. This can lead to considerable saving in computing time.

11.4.2 State Estimation Error and Forecast Error

Define the state prediction error as

$$\boldsymbol{x}_t = \boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}.$$

From the definition, the covariance matrix of $\boldsymbol{x}_t$ is $\mathrm{Var}(\boldsymbol{x}_t | F_{t-1}) = \mathrm{Var}(\boldsymbol{s}_t | F_{t-1}) = \boldsymbol{\Sigma}_{t|t-1}$. Following Section 11.1, we investigate properties of $\boldsymbol{x}_t$. First, from Eq. (11.55),

$$\boldsymbol{v}_t = \boldsymbol{Z}_t (\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{e}_t = \boldsymbol{Z}_t \boldsymbol{x}_t + \boldsymbol{e}_t.$$

Second, from Eqs. (11.62) and (11.24), and the prior equation, we have

$$\begin{aligned}
\boldsymbol{x}_{t+1} &= \boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|t} \\
&= \boldsymbol{T}_t (\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1}) + \boldsymbol{R}_t \boldsymbol{\eta}_t - \boldsymbol{K}_t \boldsymbol{v}_t \\
&= \boldsymbol{T}_t \boldsymbol{x}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t - \boldsymbol{K}_t (\boldsymbol{Z}_t \boldsymbol{x}_t + \boldsymbol{e}_t) \\
&= \boldsymbol{L}_t \boldsymbol{x}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t - \boldsymbol{K}_t \boldsymbol{e}_t,
\end{aligned}$$

where, as before, $\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t$. Consequently, we obtain a state-space form for $\boldsymbol{v}_t$ as

$$\boldsymbol{v}_t = \boldsymbol{Z}_t \boldsymbol{x}_t + \boldsymbol{e}_t, \quad \boldsymbol{x}_{t+1} = \boldsymbol{L}_t \boldsymbol{x}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t - \boldsymbol{K}_t \boldsymbol{e}_t, \tag{11.63}$$

with $\boldsymbol{x}_1 = \boldsymbol{s}_1 - \boldsymbol{s}_{1|0}$ for $t = 1, \ldots, T$.

Finally, similar to the local trend model in Section 11.1, we can show that the 1-step ahead forecast errors $\{\boldsymbol{v}_t\}$ are independent of each other and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ is independent of F_{t-1}.

11.4.3 State Smoothing

State smoothing focuses on the conditional distribution of $\boldsymbol{s}_t$ given F_T. Notice that (a) F_{t-1} and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ are independent and (b) $\boldsymbol{v}_t$ are serially independent. We can apply Theorem 11.1 to the joint distribution of $\boldsymbol{s}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1} and obtain

$$\begin{aligned} \boldsymbol{s}_{t|T} &= E(\boldsymbol{s}_t|F_T) = E(\boldsymbol{s}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) \\ &= E(\boldsymbol{s}_t|F_{t-1}) + \sum_{j=t}^{T} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)[\text{Var}(\boldsymbol{v}_t)]^{-1}\boldsymbol{v}_t \\ &= \boldsymbol{s}_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)\boldsymbol{V}_t^{-1}\boldsymbol{v}_t, \end{aligned} \tag{11.64}$$

where the covariance matrices are conditional on F_{t-1}. The covariance matrices $\text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)$ for $j = t, \ldots, T$ can by derived as follows. By Eq. (11.63),

$$\begin{aligned} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j) &= E(\boldsymbol{s}_t\boldsymbol{v}_j') \\ &= E[\boldsymbol{s}_t(\boldsymbol{Z}_j\boldsymbol{x}_j + \boldsymbol{e}_j)'] = E(\boldsymbol{s}_t\boldsymbol{x}_j')\boldsymbol{Z}_j', \quad j = t, \ldots, T. \end{aligned} \tag{11.65}$$

Furthermore,

$$\begin{aligned} E(\boldsymbol{s}_t\boldsymbol{x}_t') &= E[\boldsymbol{s}_t(\boldsymbol{s}_t - \boldsymbol{s}_{t|t-1})'] = \text{Var}(\boldsymbol{s}_t) = \boldsymbol{\Sigma}_{t|t-1}, \\ E(\boldsymbol{s}_t\boldsymbol{x}_{t+1}') &= E[\boldsymbol{s}_t(\boldsymbol{L}_t\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{e}_t)'] = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t', \\ E(\boldsymbol{s}_t\boldsymbol{x}_{t+2}') &= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{L}_{t+1}', \\ &\vdots = \vdots \\ E(\boldsymbol{s}_t\boldsymbol{x}_T') &= \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}'. \end{aligned} \tag{11.66}$$

Plugging the prior two equations into Eq. (11.64), we have

$$\begin{aligned} \boldsymbol{s}_{T|T} &= \boldsymbol{s}_{T|T-1} + \boldsymbol{\Sigma}_{T|T-1}\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T, \\ \boldsymbol{s}_{T-1|T} &= \boldsymbol{s}_{T-1|T-2} + \boldsymbol{\Sigma}_{T|T-1}\boldsymbol{Z}_{T-1}'\boldsymbol{V}_{T-1}^{-1}\boldsymbol{v}_{T-1} + \boldsymbol{\Sigma}_{T|T-1}\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T, \\ \boldsymbol{s}_{t|t} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} \\ &\quad + \cdots + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T, \end{aligned}$$

for $t = T-2, T-3, \ldots, 1$, where it is understood that $\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}' = \boldsymbol{I}_m$ when $t = T$. These smoothed state vectors can be expressed as

$$\boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{q}_{t-1}, \tag{11.67}$$

where $\boldsymbol{q}_{T-1} = \boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T$, $\boldsymbol{q}_{T-2} = \boldsymbol{Z}_{T-1}'\boldsymbol{V}_{T-1}^{-1}\boldsymbol{v}_{T-1} + \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T$, and

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} + \cdots + \boldsymbol{L}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T,$$

for $t = T-2, T-3, \ldots, 1$. The quantity $\boldsymbol{q}_{t-1}$ is a weighted sum of the 1-step ahead forecast errors $\boldsymbol{v}_j$ occurring after time $t-1$. From the definition in the prior equation, $\boldsymbol{q}_t$ can be computed recursively backward as

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \quad t = T, \ldots, 1, \tag{11.68}$$

with $\boldsymbol{q}_T = \boldsymbol{0}$. Putting the equations together, we have a backward recursion for the smoothed state vectors as

$$\boldsymbol{q}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \quad \boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{q}_{t-1}, \quad t = T, \ldots, 1, \tag{11.69}$$

starting with $\boldsymbol{q}_T = \boldsymbol{0}$, where $\boldsymbol{s}_{t|t-1}$, $\boldsymbol{\Sigma}_{t|t-1}$, $\boldsymbol{L}_t$, and $\boldsymbol{V}_t$ are available from the Kalman filter. This algorithm is referred to as the *fixed interval smoother* in the literature; see de Jong (1989) and the references therein.

Covariance Matrix of Smoothed State Vector

Next, we derive the covariance matrices of the smoothed state vectors. Applying Theorem 11.1(4) to the conditional joint distribution of $\boldsymbol{s}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1}, we have

$$\boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|t-1} - \sum_{j=t}^{T} \text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)[\text{Var}(\boldsymbol{v}_j)]^{-1}[\text{Cov}(\boldsymbol{s}_t, \boldsymbol{v}_j)]'.$$

Using the covariance matrices in Eqs. (11.65) and (11.66), we further obtain

$$\begin{aligned} \boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t\boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1}\boldsymbol{L}_t\boldsymbol{\Sigma}_{t|t-1} \\ &\quad - \cdots - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1} \cdots \boldsymbol{L}_t\boldsymbol{\Sigma}_{t|t-1} \\ &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{M}_{t-1}\boldsymbol{\Sigma}_{t|t-1}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1}\boldsymbol{L}_t \\ &\quad + \cdots + \boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1} \cdots \boldsymbol{L}_t. \end{aligned}$$

Again, $\boldsymbol{L}_t' \cdots \boldsymbol{L}_{T-1} = \boldsymbol{I}_m$ when $t = T$. From its definition, the $\boldsymbol{M}_{t-1}$ matrix satisfies

$$\boldsymbol{M}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \quad t = T, \ldots, 1, \tag{11.70}$$

with the starting value $\boldsymbol{M}_T = \boldsymbol{0}$. Collecting the results, we obtain a backward recursion to compute $\boldsymbol{\Sigma}_{t|T}$ as

$$\boldsymbol{M}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t, + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \quad \boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{M}_{t-1}\boldsymbol{\Sigma}_{t|t-1}, \quad (11.71)$$

for $t = T, \ldots, 1$ with $\boldsymbol{M}_T = \boldsymbol{0}$. Note that, like that of the local trend model in Section 11.1, $\boldsymbol{M}_t = \mathrm{Var}(\boldsymbol{q}_t)$.

Combining the two backward recursions of smoothed state vectors, we have

$$\begin{aligned}
\boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \\
\boldsymbol{s}_{t|T} &= \boldsymbol{s}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{q}_{t-1}, \\
\boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{M}_{t-1}\boldsymbol{\Sigma}_{t|t-1}, \quad t = T, \ldots, 1,
\end{aligned} \quad (11.72)$$

with $\boldsymbol{q}_T = \boldsymbol{0}$ and $\boldsymbol{M}_T = \boldsymbol{0}$.

Suppose that the state-space model in Eqs. (11.24) and (11.25) is known. Application of the Kalman filter and state smoothing can proceed in two steps. First, the Kalman filter in Eq. (11.62) is used for $t = 1, \ldots, T$ and the quantities $\boldsymbol{v}_t, \boldsymbol{V}_t, \boldsymbol{K}_t$, $\boldsymbol{s}_{t|t-1}$, and $\boldsymbol{\Sigma}_{t|t-1}$ are stored. Second, the state smoothing algorithm in Eq. (11.72) is applied for $t = T, T-1, \ldots, 1$ to obtain $\boldsymbol{s}_{t|T}$ and $\boldsymbol{\Sigma}_{t|T}$.

11.4.4 Disturbance Smoothing

Let $\boldsymbol{e}_{t|T} = E(\boldsymbol{e}_t|F_T)$ and $\boldsymbol{\eta}_{t|T} = E(\boldsymbol{\eta}_t|F_T)$ be the smoothed disturbances of the observation and transition equation, respectively. These *smoothed disturbances* are useful in many applications, for example, in model checking. In this subsection, we study recursive algorithms to compute smoothed disturbances and their covariance matrices. Again, applying Theorem 11.1 to the conditional joint distribution of $\boldsymbol{e}_t$ and $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_T\}$ given F_{t-1}, we obtain

$$\boldsymbol{e}_{t|T} = E(\boldsymbol{e}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) = \sum_{j=t}^{T} E(\boldsymbol{e}_t\boldsymbol{v}_j')\boldsymbol{V}_j^{-1}\boldsymbol{v}_j, \quad (11.73)$$

where $E(\boldsymbol{e}_t|F_{t-1}) = \boldsymbol{0}$ is used. Using Eq. (11.63),

$$E(\boldsymbol{e}_t\boldsymbol{v}_j') = E(\boldsymbol{e}_t\boldsymbol{x}_j')\boldsymbol{Z}_j' + E(\boldsymbol{e}_t\boldsymbol{e}_j').$$

Since $E(\boldsymbol{e}_t\boldsymbol{x}_t') = \boldsymbol{0}$, we have

$$E(\boldsymbol{e}_t\boldsymbol{v}_j') = \begin{cases} \boldsymbol{H}_t, & \text{if } j = t, \\ E(\boldsymbol{e}_t\boldsymbol{x}_j')\boldsymbol{Z}_j', & \text{for } j = t+1, \ldots, T. \end{cases} \quad (11.74)$$

Using Eq. (11.63) repeatedly and the independence between $\{\boldsymbol{e}_t\}$ and $\{\boldsymbol{\eta}_t\}$, we obtain

$$\begin{aligned}
E(\boldsymbol{e}_t\boldsymbol{x}_{t+1}') &= -\boldsymbol{H}_t\boldsymbol{K}_t', \\
E(\boldsymbol{e}_t\boldsymbol{x}_{t+2}') &= -\boldsymbol{H}_t\boldsymbol{K}_t'\boldsymbol{L}_{t+1}', \\
&\vdots = \vdots \\
E(\boldsymbol{e}_t\boldsymbol{x}_T') &= -\boldsymbol{H}_t\boldsymbol{K}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}',
\end{aligned} \quad (11.75)$$

where it is understood that $\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}' = \boldsymbol{I}_m$ if $t = T-1$. Based on Eqs. (11.74) and (11.75),

$$\begin{aligned}
\boldsymbol{e}_{t|T} &= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} - \cdots - \boldsymbol{K}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{v}_T) \\
&= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{q}_t) \\
&= \boldsymbol{H}_t\boldsymbol{o}_t, \quad t = T, \ldots, 1,
\end{aligned} \quad (11.76)$$

where $\boldsymbol{q}_t$ is defined in Eq. (11.67) and $\boldsymbol{o}_t = \boldsymbol{V}_t^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{q}_t$. We refer to $\boldsymbol{o}_t$ as the *smoothing measurement error*.

The smoothed disturbance $\boldsymbol{\eta}_{t|T}$ can be derived analogously and we have

$$\boldsymbol{\eta}_{t|T} = \sum_{j=t}^{T} E(\boldsymbol{\eta}_t\boldsymbol{v}_j')\boldsymbol{V}_j^{-1}\boldsymbol{v}_j. \quad (11.77)$$

The state-space form in Eq. (11.67) gives

$$E(\boldsymbol{\eta}_t\boldsymbol{v}_j') = \begin{cases} \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{Z}_{t+1}', & \text{if } j = t+1, \\ E(\boldsymbol{\eta}_t\boldsymbol{x}_j')\boldsymbol{Z}_j', & \text{if } j = t+2, \ldots, T, \end{cases}$$

where

$$\begin{aligned}
E(\boldsymbol{\eta}_t\boldsymbol{x}_{t+2}') &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{L}_{t+1}', \\
E(\boldsymbol{\eta}_t\boldsymbol{x}_{t+3}') &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{L}_{t+1}'\boldsymbol{L}_{t+2}', \\
&\vdots = \vdots \\
E(\boldsymbol{\eta}_t\boldsymbol{x}_T') &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}',
\end{aligned}$$

for $t = 1, \ldots, T$. Consequently, Eq. (11.77) implies

$$\begin{aligned}
\boldsymbol{\eta}_{t|T} &= \boldsymbol{Q}_t\boldsymbol{R}_t'(\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{v}_{t+1} + \boldsymbol{L}_{t+1}'\boldsymbol{Z}_{t+2}'\boldsymbol{V}_{t+2}^{-1}\boldsymbol{v}_{t+2} \\
&\quad + \cdots + \boldsymbol{L}_{t+1}' \cdots \boldsymbol{L}_{T-1}'\boldsymbol{Z}_T\boldsymbol{V}_T^{-1}\boldsymbol{v}_T) \\
&= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{q}_t, \quad t = T, \ldots, 1,
\end{aligned} \quad (11.78)$$

where $\boldsymbol{q}_t$ is defined earlier in Eq. (11.68).

Koopman (1993) uses the smoothed disturbance $\boldsymbol{\eta}_{t|T}$ to derive a new recursion for computing $\boldsymbol{s}_{t|T}$. From the transition equation in Eq. (11.24),

$$\boldsymbol{s}_{t+1|T} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|T} + \boldsymbol{R}_t \boldsymbol{\eta}_{t|T}.$$

Using Eq. (11.78), we have

$$\boldsymbol{s}_{t+1|T} = \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{s}_{t|T} + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' \boldsymbol{q}_t, \quad t = 1, \ldots, T, \tag{11.79}$$

where the initial value is $\boldsymbol{s}_{1|T} = \boldsymbol{s}_{1|0} + \boldsymbol{\Sigma}_{1|0}\boldsymbol{q}_0$ with $\boldsymbol{q}_0$ obtained from the recursion in Eq. (11.68).

Covariance Matrices of Smoothed Disturbances

The covariance matrix of the smoothed disturbance can also be obtained using Theorem 11.1. Specifically,

$$\begin{aligned}\text{Var}(\boldsymbol{e}_t|F_T) &= \text{Var}(\boldsymbol{e}_t|F_{t-1}, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T) \\ &= \text{Var}(\boldsymbol{e}_t|F_{t-1}) - \sum_{j=t}^{T} \text{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j)\boldsymbol{V}_j^{-1}[\text{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j)]'.\end{aligned}$$

Note that $\text{Cov}(\boldsymbol{e}_t, \boldsymbol{v}_j) = E(\boldsymbol{e}_t \boldsymbol{v}_j')$, which is given in Eq. (11.74). Thus, we have

$$\begin{aligned}\text{Var}(\boldsymbol{e}_t|F_T) &= \boldsymbol{H}_t - \boldsymbol{H}_t(\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1}\boldsymbol{K}_t \\ &\quad + \boldsymbol{K}_t'\boldsymbol{L}_{t+1}'\boldsymbol{Z}_{t+2}'\boldsymbol{V}_{t+2}^{-1}\boldsymbol{Z}_{t+2}\boldsymbol{L}_{t+1}\boldsymbol{K}_t \\ &\quad + \cdots + \boldsymbol{K}_t'\boldsymbol{L}_{t+1}'\cdots\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1}\cdots\boldsymbol{L}_{t+1}\boldsymbol{K}_t)\boldsymbol{H}_t \\ &= \boldsymbol{H}_t - \boldsymbol{H}_t(\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{M}_t\boldsymbol{K}_t)\boldsymbol{H}_t \\ &= \boldsymbol{H}_t - \boldsymbol{H}_t\boldsymbol{N}_t\boldsymbol{H}_t,\end{aligned}$$

where $\boldsymbol{N}_t = \boldsymbol{V}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{M}_t\boldsymbol{K}_t$, where $\boldsymbol{M}_t$ is given in Eq. (11.70). Similarly,

$$\text{Var}(\boldsymbol{\eta}_t|F_T) = \text{Var}(\boldsymbol{\eta}_t) - \sum_{j=t}^{T} \text{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_t)\boldsymbol{V}_t^{-1}[\text{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_t)]^{-1},$$

where $\text{Cov}(\boldsymbol{\eta}_t, \boldsymbol{v}_j) = E(\boldsymbol{\eta}_t \boldsymbol{v}_j')$, which is given before when we derived the formula for $\boldsymbol{\eta}_{t|T}$. Consequently,

$$\begin{aligned}\text{Var}(\boldsymbol{\eta}_t|F_T) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'(\boldsymbol{Z}_{t+1}'\boldsymbol{V}_{t+1}^{-1}\boldsymbol{Z}_{t+1} + \boldsymbol{L}_{t+1}'\boldsymbol{Z}_{t+2}'\boldsymbol{V}_{t+2}^{-1}\boldsymbol{Z}_{t+2}\boldsymbol{L}_{t+1} \\ &\quad + \cdots + \boldsymbol{L}_{t+1}'\cdots\boldsymbol{L}_{T-1}'\boldsymbol{Z}_T'\boldsymbol{V}_T^{-1}\boldsymbol{Z}_T\boldsymbol{L}_{T-1}\cdots\boldsymbol{L}_{t+1})\boldsymbol{R}_t\boldsymbol{Q}_t \\ &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{M}_t\boldsymbol{R}_t\boldsymbol{Q}_t.\end{aligned}$$

In summary, the disturbance smoothing algorithm is as follows:

$$\begin{aligned}\boldsymbol{e}_{t|T} &= \boldsymbol{H}_t(\boldsymbol{V}_t^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{q}_t), \\ \boldsymbol{\eta}_{t|T} &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{q}_t, \\ \boldsymbol{q}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{q}_t, \\ \text{Var}(\boldsymbol{e}_t|F_T) &= \boldsymbol{H}_t - \boldsymbol{H}_t(\boldsymbol{V}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{M}_t\boldsymbol{K}_t)\boldsymbol{H}_t, \\ \text{Var}(\boldsymbol{\eta}_t|F_T) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{M}_t\boldsymbol{R}_t\boldsymbol{Q}_t, \\ \boldsymbol{M}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{M}_t\boldsymbol{L}_t, \quad t = T, \ldots, 1,\end{aligned} \tag{11.80}$$

where $\boldsymbol{q}_T = \boldsymbol{0}$ and $\boldsymbol{M}_T = \boldsymbol{0}$.

11.5 MISSING VALUES

For the general state-space model in Eqs. (11.24) and (11.25), we consider two cases of missing values. First, suppose that similar to the local trend model in Section 11.1 the observations $\boldsymbol{y}_t$ at $t = \ell + 1, \ldots, \ell + h$ are missing. In this case, there is no new information available at these time points and we set

$$\boldsymbol{v}_t = \boldsymbol{0}, \quad \boldsymbol{K}_t = \boldsymbol{0}, \quad \text{for} \quad t = \ell + 1, \ldots, \ell + h.$$

The Kalman filter in Eq. (11.62) can then proceed as usual. That is,

$$\boldsymbol{s}_{t+1|t} = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{s}_{t|t-1}, \quad \boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t',$$

for $t = \ell + 1, \ldots, \ell + h$. Similarly, the smoothed state vectors can be computed as usual via Eq. (11.72) with

$$\boldsymbol{q}_{t-1} = \boldsymbol{T}_t'\boldsymbol{q}_t, \quad \boldsymbol{M}_{t-1} = \boldsymbol{T}_t'\boldsymbol{M}_t\boldsymbol{T}_t,$$

for $t = \ell + 1, \ldots, \ell + h$.

In the second case, some components of $\boldsymbol{y}_t$ are missing. Let $\boldsymbol{y}_t^* = \boldsymbol{J}\boldsymbol{y}_t$ be the vector of observed data at time t, where $\boldsymbol{J}$ is an indicator matrix identifying the observed data. More specifically, rows of $\boldsymbol{J}$ are a subset of the rows of the $k \times k$ identity matrix. In this case, the observation equation (11.25) of the model can be transformed as

$$\boldsymbol{y}_t^* = \boldsymbol{c}_t^* + \boldsymbol{Z}_t^*\boldsymbol{s}_t + \boldsymbol{e}_t^*,$$

where $\boldsymbol{c}_t^* = \boldsymbol{J}\boldsymbol{c}_t$, $\boldsymbol{Z}_t^* = \boldsymbol{J}\boldsymbol{Z}_t$, and $\boldsymbol{e}_t^* = \boldsymbol{J}\boldsymbol{e}_t$ with covariance matrix $\text{Var}(\boldsymbol{e}_t^*) = \boldsymbol{H}_t^* = \boldsymbol{J}\boldsymbol{H}_t\boldsymbol{J}'$. The Kalman filter and state-smoothing recursion continue to apply except that the modified observation equation is used at time t. Consequently, the ease in handling missing values is a nice feature of the state-space model.

11.6 FORECASTING

Suppose that the forecast origin is t and we are interested in predicting $\boldsymbol{y}_{t+j}$ for $j = 1, \ldots, h$, where $h > 0$. Also, we adopt the minimum mean squared error forecasts. Similar to the ARMA models, the j-step ahead forecast $\boldsymbol{y}_t(j)$ turns out to be the expected value of $\boldsymbol{y}_{t+j}$ given F_t and the model. That is, $\boldsymbol{y}_t(j) = E(\boldsymbol{y}_{t+j}|F_t)$. In what follows, we show that these forecasts and the covariance matrices of the associated forecast errors can be obtained via the Kalman filter in Eq. (11.62) by treating $\{\boldsymbol{y}_{t+1}, \ldots, \boldsymbol{y}_{t+h}\}$ as missing values, that is, case one of Section 11.5.

Consider the 1-step ahead forecast. From Eq. (11.25),

$$\boldsymbol{y}_t(1) = E(\boldsymbol{y}_{t+1}|F_t) = \boldsymbol{c}_{t+1} + \boldsymbol{Z}_{t+1}\boldsymbol{s}_{t+1|t},$$

where $\boldsymbol{s}_{t+1|t}$ is available via the Kalman filter at the forecast origin t. The associated forecast error is

$$\boldsymbol{e}_t(1) = \boldsymbol{y}_{t+1} - \boldsymbol{y}_t(1) = \boldsymbol{Z}_{t+1}(\boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|t}) + \boldsymbol{e}_{t+1}.$$

Therefore, the covariance matrix of the 1-step ahead forecast error is

$$\text{Var}[\boldsymbol{e}_t(1)] = \boldsymbol{Z}_{t+1}\boldsymbol{\Sigma}_{t+1|t}\boldsymbol{Z}_{t+1}' + \boldsymbol{H}_{t+1}.$$

This is precisely the covariance matrix $\boldsymbol{V}_{t+1}$ of the Kalman filter in Eq. (11.62). Thus, we have showed the case for $h = 1$.

Now, for $h > 1$, we consider 1-step to h-step ahead forecasts sequentially. From Eq. (11.25), the j-step ahead forecast is

$$\boldsymbol{y}_t(j) = \boldsymbol{c}_{t+j} + \boldsymbol{Z}_{t+j}\boldsymbol{s}_{t+j|t}, \tag{11.81}$$

and the associated forecast error is

$$\boldsymbol{e}_t(j) = \boldsymbol{Z}_{t+j}(\boldsymbol{s}_{t+j} - \boldsymbol{s}_{t+j|t}) + \boldsymbol{e}_{t+j}.$$

Recall that $\boldsymbol{s}_{t+j|t}$ and $\boldsymbol{\Sigma}_{t+j|t}$ are, respectively, the conditional mean and covariance matrix of $\boldsymbol{s}_{t+j}$ given F_t. The prior equation says that

$$\text{Var}[\boldsymbol{e}_t(j)] = \boldsymbol{Z}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\boldsymbol{Z}_{t+j}' + \boldsymbol{H}_{t+j}. \tag{11.82}$$

Furthermore, from Eq. (11.24),

$$\boldsymbol{s}_{t+j+1|t} = \boldsymbol{d}_{t+j} + \boldsymbol{T}_{t+j}\boldsymbol{s}_{t+j|t},$$

which in turn implies that

$$\boldsymbol{s}_{t+j+1} - \boldsymbol{s}_{t+j+1|t} = \boldsymbol{T}_{t+j}(\boldsymbol{s}_{t+j} - \boldsymbol{s}_{t+j|t}) + \boldsymbol{R}_{t+j}\boldsymbol{\eta}_{t+j}.$$

Consequently,

$$\boldsymbol{\Sigma}_{t+j+1|t} = \boldsymbol{T}_{t+j}\boldsymbol{\Sigma}_{t+j|t}\boldsymbol{T}_{t+j}' + \boldsymbol{R}_{t+j}\boldsymbol{Q}_{t+j}\boldsymbol{R}_{t+j}'. \tag{11.83}$$

Note that $\text{Var}[\boldsymbol{e}_t(j)] = \boldsymbol{V}_{t+j}$ and Eqs. (11.81)–(11.83) are the recursion of the Kalman filter in Eq. (11.62) for $t + j$ with $j = 1, \ldots, h$ when $\boldsymbol{v}_{t+j} = \boldsymbol{0}$ and $\boldsymbol{K}_{t+j} = \boldsymbol{0}$. Thus, the forecast $\boldsymbol{y}_t(j)$ and the covariance matrix of its forecast error $\boldsymbol{e}_t(j)$ can be obtained via the Kalman filter with missing values.

Finally, the prediction error series $\{\boldsymbol{v}_t\}$ can be used to evaluate the likelihood function for estimation and the standardized prediction errors $\boldsymbol{D}_t^{-1/2}\boldsymbol{v}_t$ can be used for model checking, where $\boldsymbol{D}_t = \text{diag}\{\boldsymbol{V}_t(1, 1), \ldots, \boldsymbol{V}_t(k, k)\}$ with $\boldsymbol{V}_t(i, i)$ being the (i, i)th element of $\boldsymbol{V}_t$.

11.7 APPLICATION

In this section, we consider some applications of the state-space model in finance and business. Our objectives are to highlight the applicability of the model and to demonstrate the practical implementation of the analysis in S-Plus with *SsfPack*.

Example 11.2. Consider the CAPM for the monthly simple excess returns of General Motors (GM) stock from January 1990 to December 2003; see Chapter 9. We use the simple excess returns of the S&P 500 composite index as the market returns. Our illustration starts with a simple market model

$$r_t = \alpha + \beta r_{M,t} + e_t, \quad e_t \sim N(0, \sigma_e^2) \tag{11.84}$$

for $t = 1, \ldots, 168$. This is a fixed-coefficient model and can easily be estimated by the ordinary least squares (OLS) method. Denote the GM stock return and the market return by `gm` and `sp`, respectively. The result is given below.

```
> fit=OLS(gm~sp)
> summary(fit)
Call:
OLS(formula = gm ~ sp)
Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 0.0020 0.0063     0.3151  0.7531
         sp 1.0457 0.1453     7.1964  0.0000

Regression Diagnostics:
        R-Squared 0.238
Adjusted R-Squared 0.233
Durbin-Watson Stat 2.029

Residual Diagnostics:
              Stat P-Value
```

```
Jarque-Bera  2.537  0.281
  Ljung-Box 24.207  0.337

Residual standard error: 0.0813
```

Thus, the fitted model is

$$r_t = 0.02 + 1.0457 r_{M,t} + e_t, \quad \hat{\sigma}_e = 0.0813.$$

Based on the residual diagnostics, the model appears to be adequate for the GM stock returns with adjusted $R^2 = 23.3\%$.

As shown in Section 11.3, model (11.84) is a special case of the state-space model. We then estimate the model using *SsfPack*. The result is as follows:

```
> reg.m=function(parm,mX=NULL){
+ parm=exp(parm)    % log(sigma.e) used to ensure positiveness.
+ ssf.reg=GetSsfReg(mX)
+ ssf.reg$mOmega[3,3]=parm[1]
+ CheckSsf(ssf.reg)
+ }
> c.start=c(0.1)
> reg.fit=SsfFit(c.start,gm,"reg.m",mX=X.mtx)
RELATIVE FUNCTION CONVERGENCE
> sqrt(exp(reg.fit$parameters))
[1] 0.08129934
>
% Next, perform smoothing
> ssf.reg$mOmega[3,3]=exp(reg.fit$parameters)
> reg.s=SsfMomentEst(gm,ssf.reg,task="STSMO")
> reg.s$state.moment[10,]     % use 10th row to avoid impact
          state.1  state.2    % of the starting value.
[10,] 0.001985928 1.045712
% Next, obtain standard errors of estimates
> sqrt(reg.s$state.variance[10,])
     state.1   state.2
 0.006301927 0.1453096
```

As expected, the result is in total agreement with that of the OLS method.

Finally, we entertain the time-varying CAPM of Section 11.3.1. The estimation result, including time plot of the smoothed response variable, is given below. The command `SsfCondDens` is used to compute the smoothed estimates of the state vector and observation without variance estimation.

```
> tv.capm = function(parm,mX=NULL){ % Setup the model
+ parm=exp(parm)  %parameterize in log for positiveness.
+ Phi.t = rbind(diag(2),rep(0,2))
+ Omega=diag(parm)
```



```
+ JPhi=matrix(-1,3,2)
+ JPhi[3,1]=1
+ JPhi[3,2]=2
+ Sigma=-Phi.t
+ ssf.tv=list(mPhi=Phi.t,
+ mOmega=Omega,
+ mJPhi=JPhi,
+ mSigma=Sigma,
+ mX=mX)
+ CheckSsf(ssf.tv)
+ }
> tv.start=c(0,0,0)    %starting values
> tv.mle=SsfFit(tv.start,gm,"tv.capm",mX=X.mtx) %estimation
> sigma.mle=sqrt(exp(tv.mle$parameters))
> sigma.mle
  1.168806e-05 0.0007428207  0.08129916
% Smoothing
> smoEst.tv=SsfCondDens(gm,tv.capm(tv.mle$parameters,
+ mX=X.mtx),task="STSMO")
> names(smoEst.tv)
[1] "state"    "response" "task"
> par(mfcol=c(2,2))  %plotting
> plot(gm,type='l',ylab='excess return')
> title(main='(a) Monthly simple excess returns')
> plot(smoEst.tv$response,type='l',ylab='rtn')
> title(main='(b) Expected returns')
> plot(smoEst.tv$state[,1],type='l',ylab='value')
> title(main='(c) Alpha(t)')
> plot(smoEst.tv$state[,2],type='l',ylab='value')
> title(main='(d) Beta(t)')
```

Note that estimates of σ_η and σ_ε are 1.17×10^{-5} and 0.74×10^{-3}, respectively. These estimates are close to zero, indicating that α_t and β_t of the time-varying market model are essentially constant for the GM stock returns. This is in agreement with the fact that the fixed-coefficient market model fits the data well. Figure 11.5 shows some plots for the time-varying CAPM fit. Part (a) is the monthly simple excess returns of GM stock from January 1990 to December 2003. Part (b) is the expected returns of GM stock, that is, $r_{t|T}$, where $T = 168$ is the sample size. Parts (c) and (d) are the time plots of the estimates of α_t and β_t. Given the tightness in the vertical scale, these two time plots confirm the assertion that a fixed-coefficient market model is adequate for the monthly GM stock return.

Example 11.3. In this example we reanalyze the series of quarterly earnings per share of Johnson and Johnson from 1960 to 1980 using the unobserved component model; see Chapter 2 for details of the data. The model considered is

$$y_t = \mu_t + \gamma_t + e_t, \quad e_t \sim N(0, \sigma_e^2), \tag{11.85}$$

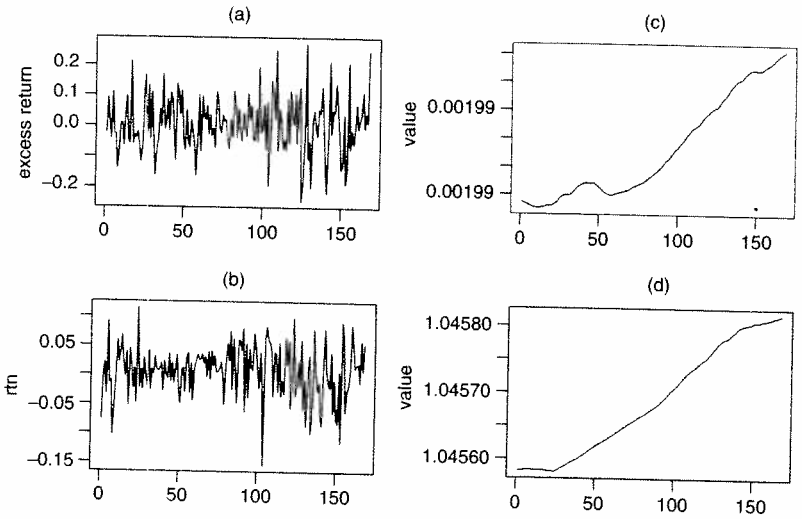

Figure 11.5. Time plots of some statistics for a time-varying CAPM applied to the monthly simple excess returns of General Motors stock. The S&P 500 composite index return is used as the market return: (a) monthly simple excess return, (b) expected returns $\boldsymbol{r}_{t|T}$, (c) α_t estimate, and (d) β_t estimate.

where y_t is the logarithm of the observed earnings per share, μ_t is the local trend component satisfying

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2),$$

and γ_t is the seasonal component that satisfies

$$(1 + B + B^2 + B^3)\gamma_t = \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2),$$

that is, $\gamma_t = -\sum_{j=1}^{3} \gamma_{t-j} + \omega_t$. This model has three parameters σ_e, σ_η, and σ_ω and is a simple unobserved component model. It can be put in a state-space form as

$$\begin{bmatrix} \mu_{t+1} \\ \gamma_{t+1} \\ \gamma_t \\ \gamma_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t \\ \omega_t \end{bmatrix},$$

where the covariance matrix of $(\eta_t, \omega_t)'$ is $\text{diag}\{\sigma_\eta^2, \sigma_\omega^2\}$, and $y_t = [1, 1, 0, 0]s_t + e_t$; see Section 11.3. This is a special case of the structural time series in *SsfPack* and can easily be specified using the command `GetSsfStsm`. Performing the maximum likelihood estimation, we obtain $(\hat{\sigma}_e, \hat{\sigma}_\eta, \hat{\sigma}_\omega) = (0.00143, 0.2696, 0.1712)$.


```
> jnj=scan(file='q-jnj.txt')
> y=log(jnj)
% Estimation
> jnj.m=function(parm){
+ parm=exp(parm)
+ jnj.sea=GetSsfStsm(irregular=parm[1],level=parm[2],
+ seasonalDummy=c(parm[3],4))
+ CheckSsf(jnj.sea)
+ }
>
> c.start=c(0,0,0)  % Starting values
> jnj.est=SsfFit(c.start,y,"jnj.m")
> names(jnj.est)
 [1] "parameters" "objective" "message" "grad.norm"
 [5] "iterations" "f.evals"   "g.evals" "hessian"
 [9] "scale"      "aux"       "call"
> jnjest=sqrt(exp(jnj.est$parameters))
> jnjest
[1] 0.001429867 0.269622976 0.171221806  % Estimates
% Next, specify the model with estimates
> jnj.ssf=GetSsfStsm(irregular=jnjest[1],level=jnjest[2],
+ seasonalDummy=c(jnjest[3],4))
> CheckSsf(jnj.ssf)
$mPhi:
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0   -1   -1   -1
[3,]    0    1    0    0
[4,]    0    0    1    0
[5,]    1    1    0    0
$mOmega:
          [,1]     [,2] [,3] [,4]      [,5]
[1,] 0.07270 0.00000    0    0         0
[2,] 0.00000 0.02932    0    0         0
[3,] 0.00000 0.00000    0    0         0
[4,] 0.00000 0.00000    0    0         0
[5,] 0.00000 0.00000    0    0 2.044e-06
$mSigma:
     [,1] [,2] [,3] [,4]
[1,]   -1    0    0    0
[2,]    0   -1    0    0
[3,]    0    0   -1    0
[4,]    0    0    0   -1
[5,]    0    0    0    0
$mDelta:
     [,1]
[1,]    0
[2,]    0
[3,]    0
```

```
 [4,]    0
 [5,]    0
$mJPhi:
 [1] 0
$mJOmega:
 [1] 0
$mJDelta:
 [1] 0
$mX:
 [1] 0
$cT:
 [1] 0
$cX:
 [1] 0
$cY:
 [1] 1
$cSt:
 [1] 4
attr(, "class"):
 [1] "ssf"   %below: smoothed components
> jnj.smo=SsfMomentEst(y,jnj.ssf,task="STSMO")
> up1=jnj.smo$state.moment[,1] +
+  2*sqrt(jnj.smo$state.variance[,1])
> lw1=jnj.smo$state.moment[,1] -
+  2*sqrt(jnj.smo$state.variance[,1])
> max(up1)  %obtain range for plotting
[1] 3.067664
> min(lw1)
[1] -1.063997
> up=jnj.smo$state.moment[,2] +
+  2*sqrt(jnj.smo$state.variance[,2])
> lw=jnj.smo$state.moment[,2] -
+  2*sqrt(jnj.smo$state.variance[,2])
> max(up)
[1] 0.5909587
> min(lw)
[1] -0.6157968
> par(mfcol=c(2,1)) %plotting
> plot(tdx,jnj.smo$state.moment[,1],type='l',xlab='year',
+ ylab='value',ylim=c(-1.1,3.1))
> lines(tdx,up1,lty=2)
> lines(tdx,lw1,lty=2)
> title(main='(a) Trend component')
> plot(tdx,jnj.smo$state.moment[,2],type='l',xlab='year',
+ ylab='value',ylim=c(-.62,0.6))
> lines(tdx,up,lty=2)
> lines(tdx,lw,lty=2)
> title(main='(b) Seasonal component')
% Filtering and smoothing
```



```
> jnj.fil=KalmanFil(y,jnj.ssf,task="STFIL")
> jnj.smo=KalmanSmo(jnj.fil,jnj.ssf)
> plot(tdx,jnj.fil$mOut[,1],type='l',xlab='year',
+  ylab='resi')
> title(main='(a) 1-Step forecast error')
> plot(tdx,jnj.smo$response.residuals[2:85],type='l',
+ xlab='year',ylab='resi')
> title(main='(b) Smoothing residual')
```

Figure 11.6 shows the smoothed estimates of the trend and seasonal components, that is, $\mu_{t|T}$ and $\gamma_{t|T}$ with $T = 84$, of the data. Of particular interest is that the seasonal pattern seems to evolve over time. Also shown are 95% pointwise confidence regions of the unobserved components. Figure 11.7 shows the residual plots, where part (a) gives the 1-step ahead forecast errors computed by Kalman filter and part (b) is the smoothed response residuals of the fitted model. Thus, state-space modeling provides an alternative approach for analyzing seasonal time series. It should be noted that the estimated components in Figure 11.6 are not unique. They depend on the model specified and constraints used. In fact, there are infinitely many ways to decompose an observed time series into unobserved components. For instance, one can use a different specification for the seasonal component,

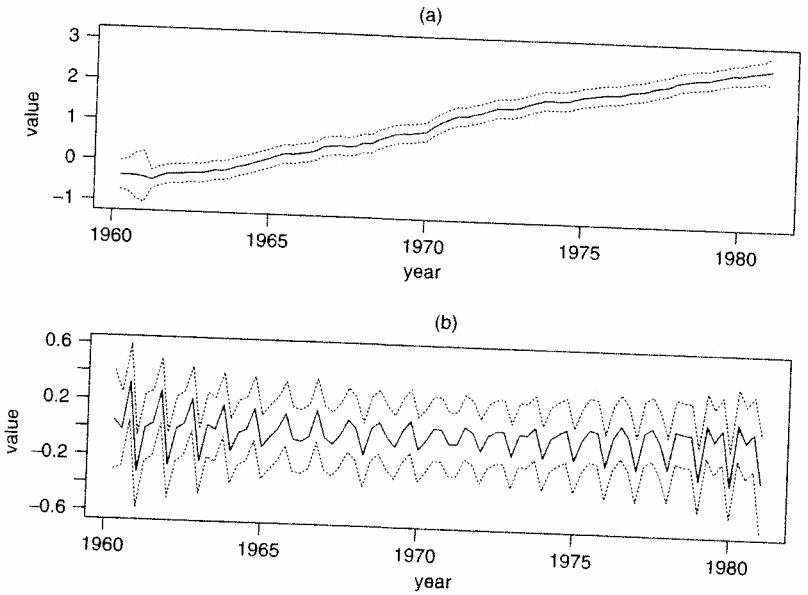

Figure 11.6. Smoothed components of fitting model (11.85) to the logarithm of quarterly earnings per share of Johnson and Johnson Company from 1960 to 1980: (a) trend component and (b) seasonal component. Dotted lines indicate pointwise 95% confidence regions.

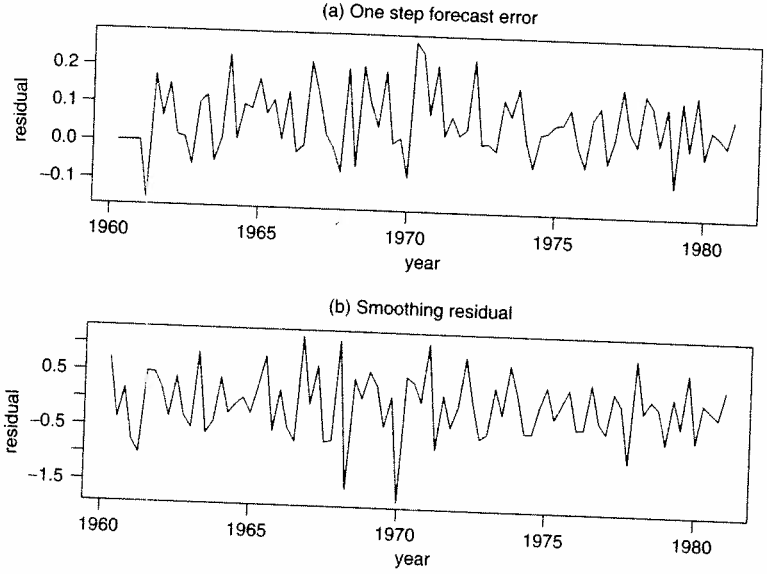

Figure 11.7. Residual series of fitting model (11.85) to the logarithm of quarterly earnings per share of Johnson and Johnson Company from 1960 to 1980: (a) 1-step ahead forecast error v_t and (b) smoothed residuals of response variable.

for example, `seasonalTrig` in *SsfPack*, to obtain another decomposition for the earnings series of Johnson and Johnson. Thus, care must be exercised in interpreting the estimated components. However, for forecasting purposes, the choice of decomposition does not matter provided that the chosen one is a valid decomposition.

EXERCISES

11.1. Consider the ARMA(1,1) model $y_t - 0.8y_{t-1} = a_t + 0.4a_{t-1}$ with $a_t \sim N(0, 0.49)$. Convert the model into a state-space form using (a) Akaike's method, (b) Harvey's approach, and (c) Aoki's approach.

11.2. The file `aa-rv-20m.txt` contains the realized daily volatility series of Alcoa stock returns from January 2, 2003 to May 7, 2004; see the example in Section 11.1. The volatility series is constructed using 20-minute intradaily log returns.

(a) Fit an ARIMA(0,1,1) model to the log volatility series and write down the model.

(b) Estimate the local trend model in Eqs. (11.1) and (11.2) for the log volatility series. What are the estimates of σ_e and σ_η? Obtain time plots for the filtered and smoothed state variables with pointwise 95% confidence interval.

11.3. Consider the monthly simple excess returns of Pfizer stock and the S&P 500 composite index from January 1990 to December 2003. The excess returns are in `m-pfesp-ex9003.txt` with Pfizer stock returns in the first column.

(a) Fit a fixed-coefficient market model to the Pfizer stock return. Write down the fitted model.

(b) Fit a time-varying CAPM to the Pfizer stock return. What are the estimated standard errors of the innovations to the α_t and β_t series? Obtain time plots of the smoothed estimates of α_t and β_t.

11.4. Consider the AR(3) model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + a_t, \quad a_t \sim N(0, \sigma_a^2),$$

and suppose that the observed data are

$$y_t = x_t + e_t, \quad e_t \sim N(0, \sigma_e^2),$$

where $\{e_t\}$ and $\{a_t\}$ are independent and the initial values of x_j with $j \leq 0$ are independent of e_t and a_t for $t > 0$.

(a) Convert the model into a state-space form.

(b) If $E(e_t) = c$, which is not zero, what is the corresponding state-space form for the system?

11.5. The file `m-ppiaco.txt` contains year, month, day, and U.S. Producer Price Index (PPI) from January 1947 to August 2004. The index is for all commodities and not seasonally adjusted. Let $z_t = \ln(Z_t) - \ln(Z_{t-1})$, where Z_t is the observed monthly PPI. It turns out that an AR(3) model is adequate for y_t if the minor seasonal dependence is ignored. Let y_t be the sample-mean corrected series of z_t.

(a) Fit an AR(3) model to y_t and write down the fitted model.

(b) Suppose that y_t has independent measurement errors so that $y_t = x_t + e_t$, where x_t is an AR(3) process and $\text{Var}(e_t) = \sigma_e^2$. Use a state-space form to estimate parameters, including the innovational variances to the state and σ_e^2. Write down the fitted model and obtain a time plot of the smoothed estimate of x_t. Also, show the time plot of filtered response residuals of the fitted state-space model.

REFERENCES

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control* **13**: 162–173.

Aoki, M. (1987). *State Space Modeling of Time Series*. Springer-Verlag, New York.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ.

Chan, N. H. (2002). *Time Series: Applications to Finance*. Wiley, Hoboken, NJ.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association* **84**: 1085–1088.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Harvey, A. C. (1993). *Time Series Models*, 2nd edition. Harvester Wheatsheaf, Hemel Hempstead, UK.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA Series D* **82**: 35–45.

Kim, C. J. and Nelson, C. R. (1999). *State Space Models with Regime Switching*. Academic Press, New York.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Springer-Verlag, New York.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**: 117–126.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1999). Statistical algorithms for models in state-space form using *SsfPack* 2.2. *Econometrics Journal* **2**: 113–166. Also, http://www.ssfpack.com/.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. Springer-Verlag, New York.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd edition. Springer-Verlag, New York.

CHAPTER 12

Markov Chain Monte Carlo Methods with Applications

Advances in computing facilities and computational methods have dramatically increased our ability to solve complicated problems. The advances also extend the applicability of many existing econometric and statistical methods. Examples of such achievements in statistics include the Markov chain Monte Carlo (MCMC) method and data augmentation. These techniques enable us to make some statistical inference that was not feasible just a few years ago. In this chapter, we introduce the ideas of MCMC methods and data augmentation that are widely applicable in finance. In particular, we discuss Bayesian inference via Gibbs sampling and demonstrate various applications of MCMC methods. Rapid developments in the MCMC methodology make it impossible to cover all the new methods available in the literature. Interested readers are referred to some recent books on Bayesian and empirical Bayesian statistics (e.g., Carlin and Louis, 2000; Gelman, Carlin, Stern, and Rubin, 2003).

For applications, we focus on issues related to financial econometrics. The demonstrations shown in this chapter represent only a small fraction of all possible applications of the techniques in finance. As a matter of fact, it is fair to say that Bayesian inference and the MCMC methods discussed here are applicable to most, if not all, of the studies in financial econometrics.

We begin the chapter by reviewing the concept of a *Markov process*. Consider a stochastic process $\{X_t\}$, where each X_t assumes a value in the space Θ. The process $\{X_t\}$ is a Markov process if it has the property that, given the value of X_t, the values of X_h, $h > t$, do not depend on the values X_s, $s < t$. In other words, $\{X_t\}$ is a Markov process if its conditional distribution function satisfies

$$P(X_h|X_s, s \leq t) = P(X_h|X_t), \quad h > t.$$

If $\{X_t\}$ is a discrete-time stochastic process, then the prior property becomes

$$P(X_h|X_t, X_{t-1}, \ldots) = P(X_h|X_t), \quad h > t.$$

Analysis of Financial Time Series, Second Edition By Ruey S. Tsay

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

A complete list of the titles in this series appears at the end of this volume.

Analysis of Financial Time Series

Second Edition

RUEY S. TSAY

University of Chicago
Graduate School of Business

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION