financial data has been discussed in great detail in the literature. We simply used this data to illustrate that in time series model fitting we can end up with fundamentally different models that will fit the data equally well. At this point, process knowledge can provide the needed guidance in picking the "right" model.

It should be noted that, in this example, we tried to keep the models simple for illustration purposes. Indeed, a more thorough analysis would (and should) pay close attention to the changing variance issue. In fact, this is a very common concern particularly when dealing with financial data. For that, we once again refer the reader to Section 7.3.

5.8 FORECASTING ARIMA PROCESSES

Once an appropriate time series model has been fit, it may be used to generate forecasts of future observations. If we denote the current time by T, the forecast for $y_{T+\tau}$ is called the τ-period-ahead forecast and denoted by $\hat{y}_{T+\tau}(T)$. The standard criterion to use in obtaining the best forecast is the mean squared error for which the expected value of the squared forecast errors, $E[(y_{T+\tau} - \hat{y}_{T+\tau}(T))^2] = E[e_T(\tau)^2]$, is minimized. It can be shown that the best forecast in the mean square sense is the conditional expectation of $y_{T+\tau}$ given current and previous observations, that is, y_T, $y_{T-1}, \ldots$:

$$\hat{y}_{T+\tau}(T) = E\left[y_{T+\tau} \middle| y_T, y_{T-1}, \ldots\right] \tag{5.84}$$

Consider, for example, an ARIMA(p, d, q) process at time $T + \tau$ (i.e., τ period in the future):

$$\boxed{y_{T+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{T+\tau-i} + \varepsilon_{T+\tau} - \sum_{i=1}^{q} \theta_i \varepsilon_{T+\tau-i}} \tag{5.85}$$

Further consider its infinite MA representation,

$$\boxed{y_{T+\tau} = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{T+\tau-i}} \tag{5.86}$$

We can partition Eq. (5.86) as

$$\boxed{y_{T+\tau} = \mu + \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i}} \tag{5.87}$$

Future Past & Present

In this partition, we can clearly see that the $\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}$ component involves the future errors, whereas the $\sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i}$ component involves the present and past errors. From the relationship between the current and past observations and the corresponding random shocks as well as the fact that the random shocks are assumed to have mean zero and to be independent, we can show that the best forecast in the mean square sense is

Conditional

$$\hat{y}_{T+\tau}(T) = E\left[y_{T+\tau} \middle| y_T, y_{T-1}, \ldots\right] = \mu + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.88}$$

since

$$E\left[\varepsilon_{T+\tau-i} \middle| y_T, y_{T-1}, \ldots\right] = \begin{cases} 0 & \text{if} \quad i < \tau \\ \varepsilon_{T+\tau-i} & \text{if} \quad i \geq \tau \end{cases}$$

Subsequently, the forecast error is calculated from

$$e_T(\tau) = y_{T+\tau} - \hat{y}_{T+\tau}(T) = \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} \tag{5.89}$$

Since the forecast error in Eq. (5.89) is a linear combination of random shocks, we have

$$E\left[e_T(\tau)\right] = 0 \tag{5.90}$$

$$\begin{aligned} \operatorname{Var}\left[e_T(\tau)\right] &= \operatorname{Var}\left[\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}\right] = \sum_{i=0}^{\tau-1} \psi_i^2 \operatorname{Var}(\varepsilon_{T+\tau-i}) \\ &= \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2 \\ &= \sigma^2(\tau), \quad \tau = 1, 2, \ldots \end{aligned} \tag{5.91}$$

$V(ax)$
$= a^2 V(x)$

It should be noted that the variance of the forecast error gets bigger with increasing forecast lead times τ. This intuitively makes sense as we should expect more uncertainty in our forecasts further into the future. Moreover, if the random shocks are assumed to be normally distributed, $N(0, \sigma^2)$, then the forecast errors will also be normally distributed with $N(0, \sigma^2(\tau))$. We can then obtain the $100(1-\alpha)$ percent prediction intervals for the future observations from

$$P\left(\hat{y}_{T+\tau}(T) - z_{\alpha/2}\sigma(\tau) < y_{T+\tau} < \hat{y}_{T+\tau}(T) + z_{\alpha/2}\sigma(\tau)\right) = 1-\alpha \tag{5.92}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution, $N(0, 1)$. Hence the $100(1-\alpha)$ percent prediction interval for $y_{T+\tau}$ is

$$\boxed{\hat{y}_{T+\tau}(T) \pm z_{\alpha/2}\sigma(\tau)} \tag{5.93}$$

There are two issues with the forecast equation in (5.88). First, it involves infinitely many terms in the past. However, in practice, we will only have a finite amount of data. For a sufficiently large data set, this can be overlooked. Second, Eq. (5.88) requires knowledge of the magnitude of random shocks in the past, which is unrealistic. A solution to this problem is to "estimate" the past random shocks through one-step-ahead forecasts. For the ARIMA model we can calculate

$$\hat{\varepsilon}_t = y_t - \left[\delta + \sum_{i=1}^{p+d} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i \hat{\varepsilon}_{t-i}\right] \tag{5.94}$$

recursively by setting the initial values of the random shocks to zero for $t < p + d + 1$. For more accurate results, these initial values together with the y_t for $t \leq 0$ can also be obtained using back-forecasting. For further details, see Box, Jenkins, and Reinsel (2008).

p d q

As an illustration consider forecasting the ARIMA(1, 1, 1) process

$$(1-\phi B)(1-B)\, y_{T+\tau} = (1-\theta B)\, \varepsilon_{T+\tau} \tag{5.95}$$

We will consider two of the most commonly used approaches:

1. As discussed earlier, this approach involves the infinite MA representation of the model in Eq. (5.95), also known as the **random shock** form of the model:

$$\begin{aligned} y_{T+\tau} &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{T+\tau-i} \\ &= \psi_0 \varepsilon_{T+\tau} + \psi_1 \varepsilon_{T+\tau-1} + \psi_2 \varepsilon_{T+\tau-2} + \cdots \end{aligned} \tag{5.96}$$

Hence the τ-step-ahead forecast can be calculated from

$$\hat{y}_{T+\tau}(T) = \psi_\tau \varepsilon_T + \psi_{\tau+1}\varepsilon_{T-1} + \cdots \tag{5.97}$$

The weights ψ_i can be calculated from

$$\left(\psi_0 + \psi_1 B + \cdots\right)(1-\phi B)(1-B) = (1-\theta B) \tag{5.98}$$

and the random shocks can be estimated using the one-step-ahead forecast error; for example, ε_T can be replaced by $e_{T-1}(1) = y_T - \hat{y}_T(T-1)$.

2. Another approach that is often employed in practice is to use **difference equations** as given by

$$y_{T+\tau} = (1+\phi)\,y_{T+\tau-1} - \phi y_{T+\tau-2} + \varepsilon_{T+\tau} - \theta\varepsilon_{T+\tau-1} \tag{5.99}$$

For $\tau = 1$, the best forecast in the mean squared error sense is

$$\hat{y}_{T+1}(T) = E\left[y_{T+1} \,\middle|\, y_T, y_{T-1}, \ldots\right] = (1+\phi)\,y_T - \phi y_{T-1} - \theta e_T(1) \tag{5.100}$$

We can further show that for lead times $\tau > 2$, the forecast is

$$\hat{y}_{T+\tau}(T) = (1-\phi)\,\hat{y}_{T+\tau-1}(T) - \phi\hat{y}_{T+\tau-2}(T) \tag{5.101}$$

Prediction intervals for forecasts of future observations at time period $T+\tau$ are found using equation 5.87. However, in using Equation 5.87 the ψ weights must be found in order to compute the variance (or standard deviation) of the τ-step ahead forecast error. The ψ weights for the general ARIMA(p, d, q) model may be obtained by equating like powers of B in the expansion of

$$\begin{aligned}(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)&(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1-B)^d \\ &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\end{aligned}$$

and solving for the ψ weights. We now illustrate this with three examples.

Example 5.3 The ARMA(1, 1) Model For the ARMA(1, 1) model the product of the required polynomials is

B^0

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1-\phi B) = (1-\theta B)$$

B^0

Equating like power of B we find that

$$
\begin{aligned}
&B^0\colon \psi_0 = 1\\
&B^1\colon \psi_1 - \phi = -\theta, \text{ or } \psi_1 = \phi - \theta\\
&B^2\colon \psi_2 - \phi\psi_1 = 0, \text{ or } \psi_2 = \phi(\phi - \theta)
\end{aligned}
$$

In general, we can show for the ARMA(1,1) model that $\psi_j = \phi^{j-1}(\phi - \theta)$.

Example 5.4 The AR(2) Model For the AR(2) model the product of the required polynomials is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \phi_2 B^2) = 1$$

B^0 B^0

Equating like power of B, we find that

$$
\begin{aligned}
&B^0\colon \psi_0 = 1\\
&B^1\colon \psi_1 - \phi_1 = 0, \text{ or } \psi_1 = \phi_1\\
&B^2\colon \psi_2 - \phi_1\psi_1 - \phi_2 = 0, \text{ or } \psi_2 = \phi_1\psi_1 + \phi_2
\end{aligned}
$$

In general, we can show for the AR(2) model that $\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2}$.

Example 5.5 The ARIMA(0, 1, 1) or IMA(1,1) Model Now consider a nonstationary model, the IMA(1, 1) model. The product of the required polynomials for this model is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - B) = (1 - \theta B)$$

It is straightforward to show that the ψ weights for this model are

$$
\begin{aligned}
\psi_0 &= 1\\
\psi_1 &= 1 - \theta\\
\psi_j &= \psi_{j-1},\ j = 2, 3, \ldots
\end{aligned}
$$

Notice that the prediction intervals will increase in length rapidly as the forecast lead time increases. This is typical of nonstationary ARIMA models. It implies that these models may not be very effective in forecasting more than a few periods ahead.