Lecture 7: Forecasting

Bus 41910, Time Series Analysis, Mr. R. Tsay

Forecasting is one of the main objectives of time series analysis. Typically, we use the criterion of **minimum mean squared errors** to produce point forecasts. Like other statistical forecasts, there are two main sources of uncertainty involved. The first uncertainty is concerned with "future" variables and the second is the uncertainty about the model used. An example of the second uncertainty is that the parameters of the model used are estimates, not "true" values. For simplicity, we shall first focus on "conditional forecasts" which is conditioned on the model used. In other words, we shall begin with methods that ignore the second source of uncertainty. Later we shall discuss several methods that take into consideration the model uncertainty.

A. Forecast of a general linear model: Recall that there are three representations for a general linear time series Z_t:

- MA representation:
$$Z_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \quad \text{with} \quad \psi_0 = 1.$$

- AR representation:
$$Z_t = d + \sum_{i=1}^{\infty} \pi_i Z_{t-i} + a_t,$$
where $d = \pi(1)\mu$ is a constant.

- ARMA (or ARIMA) representation:
$$\phi(B)(1-B)^d Z_t = c + \theta(B)a_t.$$

Here we do not separate seasonal from non-seasonal models in the discussion of ARIMA representation. The argument in effect applies to all the ARMA models. Also, for simplicity, we shall let $c = d = \mu = 0$.

These three representations are useful in different aspects of forecasting. Suppose that the origin of forecast is T and we are interested in forecasting $Z_{T+\ell}$ for $\ell > 0$. Such a forecast is called an ℓ-step ahead forecast. Mathematically, we like to find a forecast, say $\hat{Z}$, which satisfies
$$E(Z_{T+\ell} - \hat{Z})^2 = \min_f E(Z_{T+\ell} - f)^2,$$
where $\hat{Z}$ and f are functions of $Z_T, Z_{T-1}, \cdots$ and the model. Obviously, the forecast turns out to be the conditional expectation of $Z_{T+\ell}$ given $Z_T, Z_{T-1}, \cdots$ and the model. Denote the ℓ-step ahead forecast by $\hat{Z}_T(\ell)$. Then,
$$\hat{Z}_T(\ell) = E(Z_{T+\ell}|Z_T, Z_{T-1}, \cdots, \text{model}).$$

The ℓ-step ahead forecast error is

$$e_T(\ell) = Z_{T+\ell} - \hat{Z}_T(\ell)$$

and the variance of the forecast error is

$$V[e_T(\ell)] = V[Z_{T+\ell} - \hat{Z}_T(\ell)].$$

From the MA representation, we have (assuming $\mu = 0$)

$$\hat{Z}_T(\ell) = \sum_{i=0}^{\infty} \psi_{\ell+i} a_{T-i}$$

and

$$e_T(\ell) = \sum_{i=0}^{\ell-1} \psi_i a_{t+\ell-i} \quad \text{and} \quad V[e_T(\ell)] = (\sum_{i=0}^{\ell-1} \psi_i^2)\sigma_a^2.$$

Alternatively, from the AR representation, we obtain

$$\hat{Z}_T(1) = \sum_{i=0}^{\infty} \pi_{1+i} Z_{T-i} = \sum_{i=1}^{\infty} \pi_i Z_{T+1-i} \quad \text{and} \quad e_T(1) = a_{T+1}.$$

For $\ell > 1$, consider

$$Z_{T+\ell} = \pi_1 Z_{T+\ell-1} + \pi_2 Z_{T+\ell-2} + \pi_3 Z_{T+\ell-3} + \cdots + a_{T+\ell}.$$

Taking conditional expectation, we obtain

$$Z_T(\ell) = \pi_1 \hat{Z}_T(\ell-1) + \cdots + \pi_{\ell-1}\hat{Z}_T(1) + \pi_\ell Z_T + \pi_{\ell+1} Z_{T-1} + \cdots.$$

This equation can be used repeatedly to obtain a general formula for forecasting. For instance, for $\ell = 2$,

$$\begin{aligned}
\hat{Z}_T(2) &= \pi_1 \hat{Z}_T(1) + \pi_2 Z_T + \pi_3 Z_{T-1} + \cdots \\
&= \pi_1(\pi_1 Z_T + \pi_2 Z_{T-1} + \cdots) + \pi_2 Z_T + \pi_3 Z_{T-1} + \cdots \\
&= (\pi_1^2 + \pi_2) Z_T + (\pi_1\pi_2 + \pi_3) Z_{T-1} + (\pi_1\pi_3 + \pi_4) Z_{T-2} + \cdots \\
&= \sum_{i=0}^{\infty} \pi_{1+i}^{(2)} Z_{T-i}
\end{aligned}$$

where $\pi_j^{(2)} = \pi_1\pi_j + \pi_{j+1}$ for $j > 0$. In general, we have

$$\hat{Z}_T(\ell) = \sum_{i=0}^{\infty} \pi_{1+i}^{(\ell)} Z_{T-i} = (\pi_1^{(\ell)} + \pi_2^{(\ell)} B + \cdots) Z_T = (\pi_1^{(\ell)} B^\ell + \pi_2^{(\ell)} B^{\ell+1} + \cdots) Z_{T+\ell}$$

where $\pi_j^{(\ell)}$ is a function of π-weights and ψ-weights of Z_t.

To derive the general formula for $\hat{Z}_T(\ell)$, recall that its forecast error is

$$e_T(\ell) = \sum_{i=0}^{\ell-1} \psi_i a_{T+\ell-i} = (\psi_0 + \psi_1 B + \cdots + \psi_{\ell-1} B^{\ell-1}) a_{T+\ell}.$$

From the AR representation, we have

$$\pi(B) Z_{T+\ell} = a_{T+\ell}.$$

Multiplying the above equation by $(\psi_0 + \psi_1 B + \cdots + \psi_{\ell-1} B^{\ell-1})$, we obtain

$$(\psi_0 + \psi_1 B + \cdots + \psi_{\ell-1} B^{\ell-1}) \pi(B) Z_{T+\ell} = (\psi_0 + \psi_1 B + \cdots + \psi_{\ell-1} B^{\ell-1}) a_{T+\ell}.$$

Since $\psi(B)\pi(B) = 1$ for all B, it is easily seen that the coefficients of B^i in the left hand side are zero for $i = 1, 2, \cdots, \ell - 1$. Therefore, in conjunction with the fact that the right hand side is the ℓ-step ahead forecast error, the above equation in effect is in the form

$$Z_{T+\ell} - \hat{Z}_T(\ell) = \text{forecast error}.$$

Consequently, we have

$$(\psi_0 + \psi_1 B + \cdots + \psi_{\ell-1} B^{\ell-1})(1 - \pi_1 B - \pi_2 B^2 - \cdots) = 1 - \pi_1^{(\ell)} B^\ell - \pi_2^{(\ell)} B^{\ell+1} + \cdots.$$

By equating coefficients, we obtain that, for $j > 0$,

$$\pi_j^{(\ell)} = \sum_{i=0}^{\ell-1} \pi_{j+i} \psi_{\ell-i-1}$$

In particular, for $j = 1$, we have

$$\pi_1^{(\ell)} = \psi_\ell.$$

It is easy to show that $\pi_j^{(\ell)} = \pi_{j+1}^{(\ell-1)} + \psi_{\ell-1}\pi_j$ where, of course, $\pi_j^{(1)} = \pi_j$.

Eventual Forecasting Function: In many applications, we are interested in obtaining ℓ-step ahead forecasts for $\ell = 1, \cdots, k$. In this situation, the ARMA representation is useful. For convenience, we extend the forecasting notation by defining

$$\hat{Z}_T(\ell) = Z_{T+\ell} \quad \text{for} \quad \ell \le 0.$$

Also, define

$$\hat{a}_T(\ell) = \begin{cases} 0 & \text{for } \ell > 0 \\ a_{T+\ell} & \text{for } \ell \le 0 \end{cases}$$

Then, taking conditional expectation of the equation

$$Z_{T+\ell} - \phi_1 Z_{T+\ell-1} - \cdots - \phi_p Z_{T+\ell-p} = a_{T+\ell} - \theta_1 a_{T+\ell-1} - \cdots - \theta_q a_{T+\ell-q}$$

we obtain

$$\hat{Z}_T(\ell) - \phi_1 \hat{Z}_T(\ell-1) - \cdots - \phi_p \hat{Z}_T(\ell-p) = \hat{a}_T(\ell) - \theta_1 \hat{a}_T(\ell-1) - \cdots - \theta_q \hat{a}_T(\ell-q).$$

This is a recursive formula which can be used to compute the forecasts. In particular, for $\ell > q$, we have

$$\hat{Z}_T(\ell) - \phi_1 \hat{Z}_T(\ell-1) - \cdots - \phi_p \hat{Z}_T(\ell-p) = 0.$$

By letting the backshift operator "B" operates on ℓ, we have

$$\phi(B)\hat{Z}_T(\ell) = 0 \quad \text{for} \quad \ell > q.$$

This is called the "eventual forecasting function" of Z_t. It describes the forecasting pattern of Z_t. Once again, the forecasts satisfy the AR difference equation for $\ell > q$. Similar results hold for the ARIMA models.

Updating Formula: In some applications, we may wish to update the forecast in light of newly available information. For instance, we may wish to update $\hat{Z}_T(2)$ when Z_{T+1} becomes available. A simple example is that the US government may revise its October's forecast of December unemployment rate when the November unemployment rate is available. The general situation is as follows:
Original forecast: $\hat{Z}_T(\ell)$.
Updated forecast: $\hat{Z}_{T+1}(\ell-1)$.
What is the relation between these two forecasts?
To answer this question, write the model as

$$Z_{T+\ell} = \hat{Z}_T(\ell) + a_{T+\ell} + \psi_1 a_{T+\ell-1} + \cdots + \psi_{\ell-1} a_{T+1}.$$

On the other hand, we can also write it as

$$Z_{T+\ell} = \hat{Z}_{T+1}(\ell-1) + a_{T+\ell} + \psi_1 a_{T+\ell-1} + \cdots + \psi_{\ell-2} a_{T+2}.$$

Therefore, an updating formula is

$$\hat{Z}_{T+1}(\ell-1) = \hat{Z}_T(\ell) + \psi_{\ell-1} a_{T+1}.$$

Thus, $\psi_{\ell-1}$ is the only quantity needed to revise the forecast when a_{T+1} is available. This is natural in light of the fact that $\psi_{\ell-1}$ is the effect of the innovation a_{T+1} on $Z_{T+\ell}$.

Remark: For models with some deterministic component such as a time trend, one can use the above method to forecast the stochastic components, then adjust the forecasts by adding the values of the deterministic component. Of course, only stochastic component has forecast error, conditional on the given model.

B. Some simple models: We next consider the forecasts of some simple ARIMA models. Results of these simple models are informative.

a. AR(1) Model. $Z_t - \phi Z_{t-1} = a_t$. Here the π-weights are $\pi_1 = \phi$ and $\pi_j = 0$ for $j > 1$, and the ψ-weights are $\psi_i = \phi^i$. From the model and the above result, we have

$$\hat{Z}_T(\ell) = \phi^\ell Z_T, \quad e_T(\ell) = a_{T+\ell} + \phi a_{T+\ell-1} + \cdots + \phi^{\ell-1} a_{T+1}, \quad V[e_T(\ell)] = (1 + \phi^2 + \cdots + \phi^{2(\ell-1)})\sigma_a^2.$$

Clearly, the forecast $\hat{Z}_T(\ell) \to 0$ as $\ell \to \infty$. In general, $\hat{Z}_T(\ell)$ goes to the mean of Z_t as ℓ goes to infinity. Since $|\phi| < 1$, the variance of forecast error converges to $\frac{1}{1-\phi^2}\sigma_a^2$, which is the variance of Z_t. In sum, for stationary AR(1) series, the serial correlation decays exponentially to zero, implying that the current value Z_T has essentially no information about the remote future observation $Z_{T+\ell}$ for large ℓ. Therefore, the long-term forecast is the marginal distribution of Z_t. Of course, the short-term forecast is different from the marginal distribution of Z_t.

b. MA(1) Model: $Z_t = a_t - \theta a_{t-1}$ with $|\theta| < 1$. It is easy to see that for this simple model

$$\hat{Z}_T(\ell) = \begin{cases} -\theta a_T & \text{for } \ell = 1 \\ 0 & \text{for } \ell > 1 \end{cases} \qquad e_T(\ell) = \begin{cases} a_{T+1} & \text{for } \ell = 1 \\ a_{T+\ell} - \theta a_{T+\ell-1} & \text{for } \ell > 1 \end{cases}$$

$$V[e_T(\ell)] = \begin{cases} \sigma_a^2 & \text{for } \ell = 1 \\ (1+\theta^2)\sigma_a^2 & \text{for } \ell > 1. \end{cases}$$

Thus, only the 1-step ahead forecast is different from the marginal distribution of Z_t. This is obvious as the memory of an MA(1) process vanishes after 1 time period.

c. ARMA(2,1) Model: $Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} = a_t - \theta a_{t-1}$. For simplicity, I shall only give the forecasts of this example. You can easily obtain the forecast errors and their variance via the ψ-weights of Z_t. The forecasts are

$$\hat{Z}_T(1) = \phi_1 Z_T + \phi_2 Z_{T-1} - \theta a_T, \quad \hat{Z}_T(2) = \phi_1 \hat{Z}_T(1) + \phi_2 Z_T$$

$$\hat{Z}_T(\ell) = \phi_1 \hat{Z}_T(\ell-1) + \phi_2 \hat{Z}_T(\ell-2), \quad \text{for} \quad \ell \geq 3.$$

From the stationarity, the forecasts go to the mean of Z_t exponentially. For the short-term forecasts, the pattern depends on the roots of $(1 - \phi_1 B - \phi_2 B^2) = 0$. For instance, the forecasts will have a damped sine-cosine pattern for complex roots.

d. Random Walk: $Z_t = Z_{t-1} + a_t$. Here the ψ-weights are $\psi_i = 1$ for all i. Therefore,

$$\hat{Z}_T(\ell) = Z_t, \quad e_T(\ell) = a_{T+\ell} + \cdots + a_{T+1}, \quad V[e_T(\ell)] = \ell\sigma_a^2.$$

Thus, the forecasts form a *horizontal line* with value Z_T. The variance of forecast error diverges to infinity as ℓ increases. This makes sense as no one would trust the long-term forecasts of a random walk.

e. ARIMA(0,1,1) Model: $Z_t = Z_{t-1} + a_t - \theta a_{t-1}$. Recall that the ψ-weights of this exponential smoothing model are $\psi_i = (1-\theta)$ for all i. Therefore,

$$\hat{Z}_T(\ell) = \begin{cases} Z_T - \theta a_T & \text{for } \ell = 1 \\ \hat{Z}_T(1) & \text{for } \ell > 1 \end{cases} \qquad e_T(\ell) = \begin{cases} a_{T+1} & \text{for } \ell = 1 \\ a_{T+\ell} + (1-\theta)(a_{T+\ell-1} + \cdots + a_{T+1}) & \text{for } \ell > 1 \end{cases}$$

$$V[e_T(1)] = \sigma_a^2 \quad V[e_T(\ell)] = [1 + (\ell-1)(1-\theta)^2]\sigma_a^2, \quad \text{for} \quad \ell > 1.$$

f. ARIMA(0,2,2) Model: $Z_t = 2Z_{t-1} - Z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$. Again, for simplicity, I shall only give the results of forecasts.

$$\hat{Z}_T(\ell) = \begin{cases} 2Z_T - Z_{T-1} - \theta_1 a_T - \theta_2 a_{T-1} & \text{for } \ell = 1 \\ 2\hat{Z}_T(1) - Z_T - \theta_2 a_T & \text{for } \ell = 2 \\ 2\hat{Z}_T(\ell-1) - \hat{Z}_T(\ell-2) & \text{for } \ell > 2. \end{cases}$$

g. Seasonal Model: $Z_t = \Phi Z_{t-4} + a_t$. Here the ψ-weights are $\psi_{4i} = \Phi^i$ and $\psi_i = 0$, otherwise. The forecasts are

$$\hat{Z}_T(\ell) = \begin{cases} \Phi^{v+1} Z_{T-3} & \text{if } \ell = 4v+1 \\ \Phi^{v+1} Z_{T-2} & \text{if } \ell = 4v+2 \\ \Phi^{v+1} Z_{T-1} & \text{if } \ell = 4v+3 \\ \Phi^{v} Z_{T} & \text{if } \ell = 4v \end{cases}$$

which has a damped seasonal pattern. The forecast errors and their variances are easy to compute. (Exercise!)

C. Simple Exponential Smoothing

Consider the 1-step ahead prediction at origin T. Given data $\{Z_T, Z_{T-1}, \cdots, Z_1\}$, what is the (linear) prediction of Z_{T+1}? Common sense implies that in general the most recent observations should be more relevant in such a prediction. Mathematically, we can formulate the idea as follows: (i) Suppose that the weight for Z_T is ω, i.e. initial weight, and (ii) the weight is discounted by a constant rate δ, i.e. the weight for Z_{T-1} is $\delta\omega$, that for Z_{T-2} is $\delta^2\omega$, etc., where $0 < \delta < 1$. Thus, the prediction is

$$\tilde{Z}_T(1) = \omega Z_T + \delta\omega Z_{T-1} + \delta^2\omega Z_{T-2} + \cdots + \delta^{T-1}\omega Z_1.$$

However, any decent prediction *should* not change the scale of the measurement meaning that the weights should sum to 1, i.e.

$$\omega + \delta\omega + \delta^2\omega + \cdots + \delta^{T-1}\omega = 1.$$

In other words,

$$\omega(1 + \delta + \cdots + \delta^{T-1}) = \omega\frac{1-\delta^T}{1-\delta} = 1.$$

Therefore, $\omega = (1-\delta)/(1-\delta^T)$. Consequently, $\omega \to 1-\delta$ as $T \to \infty$.
For simplicity, we assume that T is sufficiently large so that $\omega = 1-\delta$. In this case, the 1-step prediction is

$$\tilde{Z}_T(1) = (1-\delta)\sum_{i=0}^{\infty} \delta^i Z_{T-i}, \tag{1}$$

which depends only on a single parameter δ, the discount rate.
<u>Updating</u>
Assume that δ is known. The 1-step ahead prediction at time $T+1$ is

$$\begin{aligned}
\tilde{Z}_{T+1}(1) &= (1-\delta)[Z_{T+1} + \delta Z_T + \delta^2 Z_{T-1} + \cdots] \\
&= (1-\delta)Z_{T+1} + \delta[Z_T + \delta Z_{T-1} + \delta^2 Z_{T-2} + \cdots] \\
&= (1-\delta)Z_{T+1} + \delta \tilde{Z}_T(1). \qquad (2)
\end{aligned}$$

This result says that given the "old" prediction $Z_T(1)$ and the new observation Z_{T+1}, the new prediction is $(1-\delta)Z_{T+1} + \delta\tilde{Z}_T(1)$, which is a weighted average of the old prediction and the new data point. The original forecast is discounted by δ.
<u>Estimation</u>
The prediction error at the forecast origin T is $e_T(1) = Z_{T+1} - \tilde{Z}_T(1)$, which is a function of δ. Consider an estimation period for T from t_0 to t_1. The parameter δ can be obtained by minimizing the sum of squared errors of prediction, i.e.

$$\hat{\delta} = \arg\min_{0<\delta<1} \sum_{T=t_0}^{t_1} e_T^2(1).$$

In this class, we use maximum likelihood estimate via ARIMA models.
<u>Relation to ARIMA models</u>
The time series Z_t can be written as

$$Z_t = \tilde{Z}_{t-1}(1) + a_t, \tag{3}$$

where $\tilde{Z}_{t-1}(1)$ is the forecast of simple exponential smoothing at origin $t-1$ and a_t is the forecast error. Similarly,

$$Z_{t-1} = \tilde{Z}_{t-2}(1) + a_{t-1}. \tag{4}$$

Using the updating formula, $\tilde{Z}_{t-1}(1) = (1-\delta)Z_{t-1} + \delta\tilde{Z}_{t-2}(1)$, we have $\tilde{Z}_{t-1}(1) - \delta\tilde{Z}_{t-2}(1) = (1-\delta)Z_{t-1}$. Multiplying Eq.(4) by δ, subtracting it from Eq. (3), and using the above identity, we have

$$Z_{t-1} - \delta Z_{t-1} = (1-\delta)Z_{t-1} + a_t - \delta z_{t-1}.$$

Consequently,

$$Z_t - Z_{t-1} = a_t - \delta a_{t-1},$$

which is an ARIMA(0,1,1) model with $\theta = \delta$. Thus, simple exponential smoothing model is an ARIMA(0,1,1) model with the constraint that $0 < \theta < 1$.

D. Combining Forecasts

Again, consider the 1-step ahead forecast of Z_{T+1} at the origin T. Suppose that there are m forecasting methods available and they produce *unbiased* forecasts $Z_{T,j}(1)$ for $j = 1, \cdots, m$. By unbiased forecast we mean that the expectation of the associated forecast error is zero. Empirical experience shows that a linear combination of these m forecasts often performs better in the mean squared error sense than the individual forecast. For instance, the simple average $Z_T(1) = \frac{1}{m}\sum_{j=1}^{M} Z_{T,j}(1)$ is often used in practice; see the book by Granger and Newbold for reference. Other combined forecasts include median forecast or weighted averages. However, there is no single combined forecast that systematically outperforms the others.

Methods of combining forecasts

Consider the case of two unbiased forecasts $Z_{T,1}(1)$ and $Z_{T,2}(1)$. Then,

$$Z_{T+1} = Z_{T,1}(1) + a_{1,T+1} = Z_{T,2}(1) + a_{2,T+1},$$

where $a_{i,T+1}$ denotes the forecast error of $Z_{T,i}(1)$ and $E(a_{i,T+1}) = 0$. Assume that the variance of $a_{i,T+1}$ is σ_i^2. If the two forecasts are independent, then we can derive the optimal combined forecast as follows. Let $Z_T(1) = \alpha Z_{T,1}(1) + (1-\alpha)Z_{T,2}(1)$ be an arbitrary combined forecast. The weight α can be determined by minimizing the variance of forecast error $\text{Var}[\alpha a_{1,T+1} + (1-\alpha)a_{2,T+1}] = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2$. Therefore, $\alpha = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. Consequently, we have

$$\begin{aligned}
Z_T(1) &= \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} Z_{T,1}(1) + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} Z_{T,2}(1) \\
&= \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} Z_{T,1}(1) + \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} Z_{T,2}(1) \\
&= \frac{p_1}{p_1+p_2} Z_{T,1}(1) + \frac{p_2}{p_1+p_2} Z_{T,2}(1),
\end{aligned}$$

where $p_i = \sigma_i^{-2}$ is called the precision of the forecast $Z_{T,i}(1)$. In other words, the weight is determined by the relative precision of the individual forecast. This is appealing. In general, if all forecasts are uncorrelated with precision p_i, then the optimal forecast is

$$Z_T(1) = \sum_{i=1}^{m} \frac{p_i}{\sum_{j=1}^{m} p_j} Z_{T,i}(1).$$

In practice, the forecasts are often correlated because they are typically based on similar information. The combination then becomes much more involved. A practice procedure is to use multiple linear regression. Again, consider the case of two forecast methods. Suppose that there is a *forecasting* period available for T from t_0 to t_1. Let $Z_{T,i}(1)$ be the 1-step ahead forecast of method i at forecast origin T. Consider the multiple linear regression

$$Z_{T+1} = \beta_0 + \beta_1 Z_{T,1}(1) + \beta_2 Z_{T,2}(1), \quad T = t_0, \cdots, t_1.$$

Estimate the above multiple linear regression by the ordinary least squares method. The estimate $\hat{\beta}_i$ for $i = 0, 1, 2$ are then used to combine forecast for origin $T > t_1$.
The constant term β_0 is used to handle any bias in the individual forecasts. Obviously, adopting this regression approach means that we assume that the weights continue to hold for $T > t_1$. This may not be true in real application. Consequently, optimal combined forecast may not exist. The approach applies to more than two individual forecasts.

E. Forecast Evaluation

It is hard in general to evaluate the performance of different forecasting methods. However, mean squared error of 1-step ahead out-of-sample forecasts is often used as an evaluation criterion. This criterion also has its own share of weakness, however. For instance, the choice of the evaluation period may affect the performance. Also, the criterion may select different forecasting methods for different forecast horizons. Furthermore, the criterion is sensitive to outliers in the evaluation period.
Finally, in application, the uncertainty in the model used and the uncertainty in parameter estimates of a specified model are also important. For a given forecast model, parameter uncertainty can be handled by using the predictive distribution of a forecast. Model uncertainty is much harder to handle.