Times Series Analysis (XV) – Forecasting (II)

Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University

31st Oct., 2017

Recap

Deterministic Trend Forecast

- The forecast $\hat{Y}_t(\ell) = E(Y_{t+\ell}|Y_1, \cdots, Y_t)$
  - $Y_t = \mu_t + X_t$, $X_t \sim \mathcal{N}(0, \gamma_0)$
  - $\hat{Y}_t(\ell) = \mu_{t+\ell}$
  - In case of linear trend, $\hat{Y}_t(\ell) = \beta_0 + \beta_1(t + \ell)$
  - In case of seasonal models, $\mu_t = \mu_{t+12}$, $\hat{Y}_t(\ell) = \hat{Y}_t(\ell + 12)$.
- The forecast error $e_t(\ell)$
  - $E(e_t(\ell)) = 0$ and $Var(e_t(\ell)) = \gamma_0$.

Recap

ARIMA Forecasting: AR(1): $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$

- Forecast
  - $\hat{Y}_t(1) = \mu + \phi(Y_t - \mu)$
  - $\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell - 1) - \mu]$ for $\ell \geq 1$ or $\hat{Y}_t(\ell) = \mu + \phi^{\ell}(Y_t - \mu)$, or
  - Since $|\phi| < 1$, $\hat{Y}_t(\ell) \approx \mu$.
- Forecast Error
  - one-step-ahead forecast error: $e_t(1) = e_{t+1}$, $Var(e_t(1)) = \sigma_e^2$
  - longer lead forecast error: $e_t(\ell) = e_{t+\ell} + \phi e_{t+\ell-1} + \cdots + \phi^{\ell-1} e_{t+1}$
  - $e_t(\ell) = e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \cdots + \Psi_{\ell-1} e_{t+1}$ holds for all ARIMA models.
  - $Var(e_t(\ell)) = \sigma_e^2(1 + \Psi_1^2 + \cdots + \Psi_{\ell-1}^2)$
  - For AR(1), $Var(e_t(\ell)) = \sigma_e^2 \left[\frac{1-\phi^{2\ell}}{1-\phi^2}\right] \approx \frac{\sigma_e^2}{1-\phi^2} = Var(Y_t) = \gamma_0$ for large ℓ.

Recap

MA(1): $Y_t = \mu + e_t - \theta e_{t-1}$

- $\hat{Y}_t(1) = \mu - \theta e_t$, $e_t(1) = e_{t+1}$
- $\hat{Y}_t(\ell) = \mu$ for $\ell > 1$ (cf. $\hat{Y}_t(\ell) \approx \mu$ for AR(1))
- $e_t(\ell) = e_{t+\ell} - \theta e_{t+\ell-1}$ for $\ell > 1$
- $Var(e_t(\ell)) = \begin{cases} \sigma_e^2, \ell = 1; \\ \sigma_e^2(1+\theta^2), \ell > 1 \end{cases}$

Random Walk with Drift: $Y_t = Y_{t-1} + \theta_0 + e_t$

- $\hat{Y}_t(1) = Y_t + \theta_0$ and $\hat{Y}_t(\ell) = Y_t + \theta_0 \ell$
- $e_t(1) = e_{t+1}$ and $e_t(\ell) = e_{t+1} + e_{t+2} + \cdots + e_{t+\ell}$, $Var(e_t(\ell)) = \ell \sigma_e^2$

Recap

ARMA(p, q):

- The forecast

$$\hat{Y}_t(\ell) = \phi_1 \hat{Y}_t(\ell-1) + \phi_2 \hat{Y}_t(\ell-2) + \cdots + \phi_p \hat{Y}_t(\ell-p) + \theta_0$$
$$- \theta_1 E(e_{t+\ell-1}|Y_1, Y_2, \cdots, Y_t) - \theta_2 E(e_{t+\ell-2}|Y_1, Y_2, \cdots, Y_t)$$
$$- \cdots - \theta_q E(e_{t+\ell-q}|Y_1, Y_2, \cdots, Y_t)$$

- For $\ell > q$,

$$\hat{Y}_t(\ell) = \phi_1 \hat{Y}_t(\ell-1) + \phi_2 \hat{Y}_t(\ell-2) + \cdots + \phi_p \hat{Y}_t(\ell-p) + \theta_0$$

- Write it in terms of deviations, $\rightarrow$ Yule-Walker Equation

$$\hat{Y}_t(\ell) - \mu = \phi_1[\hat{Y}_t(\ell-1) - \mu] + \phi_2[\hat{Y}_t(\ell-2) - \mu] + \cdots +$$
$$\phi_p[\hat{Y}_t(\ell-p) - \mu] \text{ for } \ \ell > q$$

Recap

- $\hat{Y}_t(\ell) - \mu$ can be expressed as a linear combination of exponentially decaying terms in ℓ (corresponding to the real roots) and damped sine wave terms (corresponding to the pairs of complex roots). That is $\hat{Y}_t(\ell) \approx \mu$ for large ℓ.
- Note that any ARIMA model can be written in truncated linear process form as $Y_{t+\ell} = C_t(\ell) + I_t(\ell)$ for $\ell > 1$, thus
- $e_t(\ell) = I_t(\ell) = e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \cdots + \Psi_{\ell-1} e_{t+1}$, and
- $E[e_t(\ell)] = 0$ for $\ell \geq 1$
- $Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \Psi_j^2$ for $\ell \geq 1$

As the random walk shows, forecasting for nonstationary ARIMA models is quite similar to forecasting for stationary ARMA models, but there are some striking differences.

Recall that an ARIMA$(p, 1, q)$ model can be written as a nonstationary ARMA$(p+1, q)$ model, We shall write this as

$$Y_t = \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \psi_3 Y_{t-3} + \cdots + \psi_p Y_{t-p} + \psi_{p+1} Y_{t-p-1} \\ + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

where the script coefficients ψ are directly related to the ϕ coefficients.

In particular,

$$\begin{aligned} \psi_1 &= 1+\phi_1, \psi_j = \phi_j - \phi_{j-1}, \qquad \text{for} \;\; j=1,2,\cdots,p \\ &\text{and} \\ \psi_{p+1} &= -\phi_p \end{aligned}$$

For a general order of differencing d, we would have $p+d$ of the ψ coefficients.

From this representation, we can immediately extend previous equations to cover the nonstationary cases by replacing p by $p+d$ and ψ_j by ϕ_j.

Nonstationary Models

As an example of the necessary calculations, consider the ARIMA(1,1,1) case. Here

$$Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + \theta_0 + e_t - \theta e_{t-1}$$

so that

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + \theta_0 + e_t - \theta e_{t-1}$$

Thus

$$\begin{aligned}
\hat{Y}_t(1) &= (1+\phi)Y_t - \phi Y_{t-1} + \theta_0 - \theta e_t \\
\hat{Y}_t(2) &= (1+\phi)\hat{Y}_t(1) - \phi Y_t + \theta_0 \\
&\text{and} \\
\hat{Y}_t(\ell) &= (1+\phi)\hat{Y}_t(\ell-1) - \phi \hat{Y}_t(\ell-2) + \theta_0
\end{aligned}$$

Nonstationary Models

For the general invertible ARIMA model, the truncated linear process representation and the calculations following these equations show that we can write

$$e_t(\ell) = e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \Psi_2 e_{t+\ell-2} + \cdots + \Psi_{\ell-1} e_{t+1} \qquad \text{for} \quad \ell \geq 1$$

and so

$$E(e_t(\ell)) = 0 \qquad Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \Psi_j^2 \qquad \text{for} \quad \ell \geq 1$$

However, for nonstationary series, the Ψ_j-weights do not decay to zero as j increases.

For example, for the random walk model, $\Psi_j = 1$ for all j; for the IMA(1,1) model, $\Psi_j = 1 - \theta$ for $j \geq 1$; for the IMA(2,2) case, $\Psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2 + 2)j$ for $j \geq 1$; and for the ARI(1,1) model, $\Psi_j = (1 - \phi^{j+1})/(1 - \phi)$ for $j \geq 1$.

Thus, for any nonstationary model, the equation shows that the forecast error variance will grow without bound as the lead time ℓ increases. This fact should not be too surprising since with nonstationary series the distant future is quite uncertain.

Prediction Limits

As in all statistical endeavors, in addition to forecasting or predicting the unknown $Y_{t+\ell}$, we would like to assess the precision of our predictions.

<u>Deterministic Trends</u>:

For the deterministic trend model with a white noise stochastic component $\{X_t\}$, we recall that

$$\hat{Y}_t(\ell) = \mu_{t+\ell} \qquad \text{and} \quad Var(e_t(\ell)) = Var(X_{t+\ell}) = \gamma_0$$

If the stochastic component is normally distributed, then the forecast error

$$e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell) = X_{t+\ell}$$

is also normally distributed.

Prediction Limits

Thus, for a given confidence level $1 - \alpha$, we could use a standard normal percentile, $z_{1-\alpha/2}$, to claim that

$$P\left[-z_{1-\alpha/2} < \frac{Y_{t+\ell} - \hat{Y}_t(\ell)}{\sqrt{Var(e_t(\ell))}} < z_{1-\alpha/2}\right] = 1 - \alpha$$

or equivalently,

$$P\left[\hat{Y}_t(\ell) - z_{1-\alpha/2}\sqrt{Var(e_t(\ell))} < Y_{t+\ell} \right.$$
$$\left. < \hat{Y}_t(\ell) + z_{1-\alpha/2}\sqrt{Var(e_t(\ell))}\right] = 1 - \alpha$$

Thus we may be $(1 - \alpha)100\%$ confident that the future observation $Y_{t+\ell}$ will be contained within the prediction limits

$$\hat{Y}_t(\ell) \pm z_{1-\alpha/2}\sqrt{Var(e_t(\ell))}$$

Prediction Limits

As a numerical example, consider the monthly average temperature series once more. We used the cosine model to predict the June 1976 average temperature as $68.3^{\circ}F$.The estimate of $\sqrt{Var(e_t(\ell))} = \sqrt{\gamma_0}$ for this model is $3.7^{\circ}F$. Thus 95% prediction limits for the average Jun 1976 temperature are

$$68.3 \pm 1.96(3.7) = 68.3 \pm 7.252 \text{ or } \ 61.05^{\circ}F \text{ to } \ 75.55^{\circ}F$$

Prediction Limits

- From standard regression analysis, since the forecast involves estimated regression parameters, the correct forecast error variance is given by $\gamma_0[1 + (1/n) + c_{n,\ell}]$, where $c_{n,\ell}$ is a certain function of the sample size n and the lead time ℓ.
- However, it may be shown that for the types of trends that we are considering (namely, cosines and polynomials in time) and for large sample sizes n, the $1/n$ and $c_{n,\ell}$ are both negligible relative to 1.
- For example, with a cosine trend of period 12 over $N = n/12$ years, we have that $c_{n\ell} = 2/n$; thus the correct forecast error variance is $\gamma_0[1 + (3/n)]$ rather than our approximate γ_0
- For the linear time trend model, it can be shown that $c_{n,\ell} = 3(n + 2\ell - 1)^2/[n(n^2 - 1)] \approx 3/n$ for moderate lead ℓ and large n. Thus, again our approximation seems justified.

Prediction Limits

ARIMA Models:

If the white noise terms $\{e_t\}$ in a general ARIMA series each arise independently from a normal distribution, then from previous equation, the forecast error $e_t(\ell)$ will also have a normal distribution, and the steps leading to

$$\hat{Y}_t(\ell) \pm z_{1-\alpha/2}\sqrt{Var(e_t(\ell))}$$

remain valid. However, in contrast to the deterministic trend model, recall that in the present case

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \Psi_j^2$$

Prediction Limits

In practice, σ_e^2 will be unknown and must be estimated from the observed time series. The necessary Ψ-weights are, of course, also unknown since they are certain functions of the unknown ϕ's and θ's. For large sample sizes, these estimations will have little effect on the actual prediction limits given above.

As a numerical example, consider the AR(1) model that we estimated for the industrial color property series. We use $\phi = 0.5705$, $\mu = 74.3293$, and $\sigma_e^2 = 24.8$.

For an AR(1) model, recall

$$Var(e_t(\ell)) = \sigma_e^2 \left[\frac{1 - \phi^{2\ell}}{1 - \phi^2}\right]$$

Prediction Limits

For a one-step-ahead prediction, we have

$$70.14793 \pm 1.96\sqrt{24.8} = 70.14793 \pm 9.760721 \text{ or } 60.39 \text{ to } 79.91$$

Two steps ahead, we obtain

$$71.86072 \pm 11.88343 \text{ or } 60.71 \text{ to } 83.18$$

Notice that this prediction interval is wider than the previous interval. Forecasting ten steps ahead leads to

$$74.173934 \pm 11.88451 \text{ or } 62.42 \text{ to } 86.19$$

By lead 10, both the forecast and the forecast limits have settled down to their long-lead values.

Deterministic Trends:

Figure: Forecasts and Limits for the Temperature Cosine Trend

This figure displays the last four years of the average monthly temperature time series together with forecasts and 95% forecast limits for two additional years. Since the model fits quite well with a relatively small error variance, the forecast limits are quite close to the fitted trend forecast.

ARIMA Models:

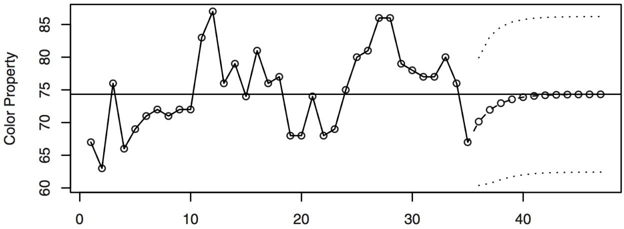

Figure: Forecasts and Forecast Limits for the AR(1) Model for Color

This figure displays this series together with forecasts out to lead time 12 with the upper and lower 95% prediction limits for those forecasts. In addition, a horizontal line at the estimate for the process mean is shown. Notice how the forecasts approach the mean exponentially as the lead time increases. Also note how the prediction limits increase in width.

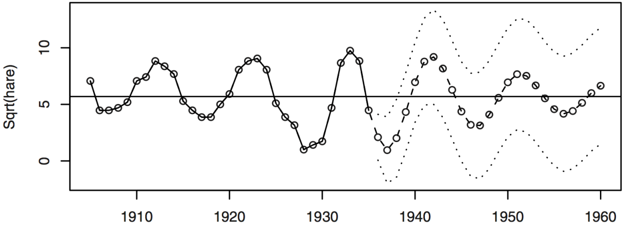

Figure: Forecasts from an AR(3) Model for Sqrt(Hare)

The Canadian hare abundance series was fitted by working with the square root of the abundance numbers and then fitting an AR(3) model. Notice how the forecasts mimic the approximate cycle in the actual series even when we forecast with a lead time out to 25 years in this figure.

Updating ARIMA Forecasts

Suppose we are forecasting a monthly time series. Our last observation is, say, for February, and we forecast for March, April, and May.

As time goes by, the actual value for March becomes available. With this new value in hand, we would like to update or revise (and, one hopes, improve) our forecasts for April and May.

Of course, we could compute new forecasts from scratch. However, there is a simpler way.

Updating ARIMA Forecasts

For a general forecast origin t and lead time $\ell + 1$, our original forecast is denoted $\hat{Y}_t(\ell + 1)$.

Once the observation at time $t + 1$ becomes available, we would like to update our forecast as $\hat{Y}_{t+1}(\ell)$. This yields

$$Y_{t+\ell+1} = C_t(\ell + 1) + e_{t+\ell+1} + \Psi_1 e_{t+\ell} + \Psi_2 e_{t+\ell-1} + \cdots + \Psi_\ell e_{t+1}$$

Since $C_t(\ell + 1)$ and e_{t+1} are functions of $Y_{t+1}, Y_t, \cdots$, whereas $e_{t+\ell+1}, e_{t+\ell}, \cdots, e_{t+2}$ are independent of $Y_{t+1}, Y_t, \cdots$,we quickly obtain the expression

$$\hat{Y}_{t+1}(\ell) = C_t(\ell + 1) + \Psi_\ell e_{t+1}$$

Updating ARIMA Forecasts

However, $\hat{Y}_t(\ell + 1) = C_t(\ell + 1)$, and, of course, $e_{t+1} = Y_{t+1} - \hat{Y}_t(1)$. Thus, we have the general **updating equation**

$$\hat{Y}_{t+1}(\ell) = \hat{Y}_t(\ell + 1) + \Psi_\ell[Y_{t+1} - \hat{Y}_t(1)]$$

Notice that $[Y_{t+1} - \hat{Y}_t(1)]$ is the actual forecast error at time $t + 1$ once Y_{t+1} has been observed.

Updating ARIMA Forecasts

As a numerical example, consider the the industrial color property time series. We fit an AR(1) model to forecast one step ahead as $\hat{Y}_{35}(1) = 70.14757$ and two steps ahead as $\hat{Y}_{35}(2) = 71.94342$.

If now the next color value becomes available as $Y_{t+1} = Y_{36} = 65$, then we update the forecast for time $t = 37$ as

$$\hat{Y}_{t+1}(1) = \hat{Y}_{36}(1) = 71.94342 + 0.5705(65 - 70.14757) = 69.00673$$

Forecast Weights and Exponentially Weighted Moving Averages

- For ARIMA models without moving average terms, it is clear how the forecasts are explicitly determined from the observed series $Y_t, Y_{t-1}, \cdots, Y_1$.
- However, for any model with $q > 0$, the noise terms appear in the forecasts, and the nature of the forecasts explicitly in terms of $Y_t, Y_{t-1}, \cdots, Y_1$ is hidden.
- To bring out this aspect of the forecasts, we return to the inverted form of any invertible ARIMA process, namely

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \cdots + e_t$$

Forecast Weights and Exponentially Weighted Moving Averages

Thus we can also write

$$Y_{t+1} = \pi_1 Y_t + \pi_2 Y_{t-1} + \pi_3 Y_{t-2} + \cdots + e_{t+1}$$

Taking conditional expectations of both sides, given $Y_t, Y_{t-1}, \cdots, Y_1$, we obtain

$$\hat{Y}_t(1) = \pi_1 Y_t + \pi_2 Y_{t-1} + \pi_3 Y_{t-2} + \cdots$$

(We are assuming the t is sufficiently large and/or that the π-weights die out sufficiently quickly so that $\pi_t, \pi_{t+1}, \cdots$ are all negligible.)

Forecast Weights and Exponentially Weighted Moving Averages

For any invertible ARIMA model, the π-weights can be calculated recursively from the expressions

$$\pi_j = \begin{cases} \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i} + \psi_j & \text{for } 1 \leq j \leq p+d \\ \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i} & \text{for } j > p+d \end{cases}$$

with initial value $\pi_0 = -1$. (Compare this with the ARMA model for the Ψ-weights.)

Forecast Weights and Exponentially Weighted Moving Averages

Consider in particular the nonstationary IMA(1,1) model, $Y_t = Y_{t-1} + e_t - \theta e_{t-1}$. Here $p = 0, d = 1, q = 1$, with $\phi_1 = 1$; thus

$$\begin{aligned} \pi_1 &= \theta\pi_0 + 1 = 1 - \theta \\ \pi_2 &= \theta\pi_1 = \theta(1 - \theta) \end{aligned}$$

and generally,

$$\pi_j = \theta\pi_{j-1} \text{ for } \; j > 1$$

Thus we have explicitly

$$\pi_j = (1 - \theta)\theta^{j-1}$$

so we have

$$\pi_j = (1 - \theta)\theta^{j-1} \text{ for } \; j \geq 1$$

Forecast Weights and Exponentially Weighted Moving Averages

That is, we can write

$$\hat{Y}_t(1) = (1-\theta)Y_t + (1-\theta)\theta Y_{t-1} + (1-\theta)\theta^2 Y_{t-2} + \cdots$$

In this case, the π-weights decrease exponentially, and futhermore,

$$\sum_{j=1}^{\infty} \pi_j = (1-\theta)\sum_{j=1}^{\infty} \theta^{j-1} = \frac{1-\theta}{1-\theta} = 1$$

Thus $\hat{Y}_t(1)$ is called the **exponentially weighted moving average (EWMA)**.

Forecast Weights and Exponentially Weighted Moving Averages

Simple algebra shows that we can also write

$$\hat{Y}_t(1) = (1-\theta)Y_t + \theta\hat{Y}_{t-1}(1)$$

and

$$\hat{Y}_t(1) = \hat{Y}_{t-1}(1) + (1-\theta)\left[Y_t - \hat{Y}_{t-1}(1)\right]$$

These equations show how to update forecasts from origin $t-1$ to origin t, and they express the result as a linear combination of the new observation and the old forecast or in terms of the old forecast and the last observed forecast error.

Forecast Weights and Exponentially Weighted Moving Averages

- The parameter $1 - \theta$ is called the **smoothing constant** in EWMA literature, and its selection (estimation) is often quite arbitrary.
- From the ARIMA model-building approach, we let the data indicate whether an IMA(1,1) model is appropriate for the series under consideration.
- If so, we then estimate θ in an efficient manner and compute an EWMA forecast that we are confident is the minimum mean square error forecast. A comprehensive treatment of exponential smoothing methods and their relationships with ARIMA models is given in Abraham and Ledolter (1983).

Forecasting Transformed Series

Differencing: Suppose we are interested in forecasting a series whose model involves a first difference to achieve stationarity. Two methods of forecasting can be considered:

- forecasting the original nonstationary series, for example by using the difference equation form of ARMA(p, q) with ϕ's replaced by ψ's throughout, or
- forecasting the stationary differenced series $W_t = Y_t - Y_{t-1}$ and then "undoing" the difference by summing to obtain the forecast in original terms.

We shall show that both methods lead to the same forecasts. This follows essentially because differencing is a linear operation and because conditional expectation of a linear combination is the same linear combination of the conditional expectations.

Forecasting Transformed Series

Consider in particular the IMA(1,1) model. Basing our work on the original nonstationary series, we forecast as

$$\hat{Y}_t(1) = Y_t - \theta e_t \text{ and } \hat{Y}_t(\ell) = \hat{Y}_t(\ell - 1) \text{ for } \ell > 1$$

Consider now the differenced stationary MA(1) series $W_t = Y_t - Y_{t-1}$. We would forecast $W_{t+\ell}$ as

$$\hat{W}_t(1) = -\theta e_t \text{ and } \hat{W}_t(\ell) = 0 \text{ for } \ell > 1$$

Forecasting Transformed Series

However, $\hat{W}_t(1) = \hat{Y}_t(1) - Y_t$; thus $\hat{W}_t(1) = -\theta e_t$ is equivalent to $\hat{Y}_t(1) = Y_t - \theta e_t$ as before. Similarly, $\hat{W}_t(\ell) = \hat{Y}_t(\ell) - \hat{Y}_t(\ell - 1)$.

The same result would apply to any model involving differences of any order and indeed to any type of linear transformation with constant coefficients. (Certain linear transformations other than differencing may be applicable to seasonal time series.

Log Transformations:
As we saw earlier, it is frequently appropriate to model the logarithms of the original series — a nonlinear transformation. Let Y_t denote the original series value and let $Z_t = \log(Y_t)$. It can be shown that we always have

$$E(Y_{t+\ell}|Y_t, Y_{t-1}, \cdots, Y_1) \geq \exp\left[E(Z_{t+\ell}|Z_t, Z_{t-1}, \cdots, Z_1)\right]$$

with equality holding only in trivial cases.

Forecasting Transformed Series

Thus, the naive forecast $\exp[\hat{Z}_t(\ell)]$ is not the minimum mean square error forecast of $Y_{t+\ell}$. To evaluate the minimum mean square error forecast in original terms, we shall find the following fact useful: If X has a normal distribution with mean μ and variance σ^2, then

$$E\left[\exp(X)\right] = \exp\left[\mu + \frac{\sigma^2}{2}\right]$$

In our application

$$\mu = E(Z_{t+\ell}|Z_t, Z_{t-1}, \cdots, Z_1)$$

and

$$\begin{aligned}
\sigma^2 &= Var(Z_{t+\ell}|Z_t, Z_{t-1}, \cdots, Z_1) \\
&= Var[e_t(\ell) + C_t(\ell)|Z_t, Z_{t-1}, \cdots, Z_1) \\
&= Var[e_t(\ell)|Z_t, Z_{t-1}, \cdots, Z_1) + Var[C_t(\ell)|Z_t, Z_{t-1}, \cdots, Z_1) \\
&= Var[e_t(\ell)|Z_t, Z_{t-1}, \cdots, Z_1) \\
&= Var[e_t(\ell)]
\end{aligned}$$

These follow from the truncated linear process (applied to Z_t) and the fact that $C_t(\ell)$ is a function of $Z_t, Z_{t-1}, \cdots$, whereas $e_t(\ell)$ is independent of $Z_t, Z_{t-1}, \cdots$. Thus the minimum mean square error forecast in the original series is given by

$$\exp\left\{\hat{Z}_t(\ell) + \frac{1}{2}Var[e_t(\ell)]\right\}$$

Forecasting Transformed Series

- Throughout our discussion of forecasting, we have assumed that minimum mean square forecast error is the criterion of choice. For normally distributed variables, this is an excellent criterion.
- However, if Z_t has a normal distribution, then $Y_t = \exp(Z_t)$ has a lognormal distribution, for which a different criterion may be desirable.
- In particular, since the log-normal distribution is asymmetric and has a long right tail, a criterion based on the mean absolute error may be more appropriate.
- For this criterion, the optimal forecast is the median of the distribution of $Z_{t+\ell}$ conditional on $Z_t, Z_{t-1}, \cdots, Z_1$.
- Since the log transformation preserves medians and since, for a normal distribution, the mean and median are identical, the naive forecast $\exp[\hat{Z}_t(\ell)]$ is the optimal forecast for $Y_{t+\ell}$ in the sense that it minimizes the mean absolute forecast error.

Summary of Forecasting with Certain ARIMA Models

IMA(1,1) with Constant Term: $Y_t = Y_{t-1} + \theta_0 + e_t - \theta e_{t-1}$

$$\begin{aligned}
\hat{Y}_t(\ell) &= \hat{Y}_t(\ell-1) + \theta_0 - \theta e_t = Y_t + \ell\theta_0 - \theta e_t \\
\hat{Y}_t(1) &= (1-\theta)Y_t + (1-\theta)\theta Y_{t-1} + \cdots (\text{the EWMA for } \theta_0 = 0) \\
e_t(\ell) &= e_{t+\ell} + (1-\theta)e_{t+\ell-1} + \cdots + (1-\theta)e_{t+1} \text{ for } \ \ell \geq 1 \\
Var(e_t(\ell)) &= \sigma_e^2[1 + (\ell-1)(1-\theta)^2] \\
\Psi_j &= 1 - \theta \text{ for } \ j > 0
\end{aligned}$$

Note that if $\theta_0 \neq 0$, the forecasts follow a straight line with slope θ_0, but if $\theta_0 = 0$, which is the usual case, then the forecast is the same for all lead times, namely

$$\hat{Y}_t(\ell) = Y_t - \theta e_t$$

Summary of Forecasting with Certain ARIMA Models

IMA(2,2): $Y_t = 2Y_{t-1} - Y_{t-2} + \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$

$$\begin{aligned}
\hat{Y}_t(1) &= 2Y_t - Y_{t-1} + \theta_0 - \theta_1 e_t - \theta_2 e_{t-1} \\
\hat{Y}_t(2) &= 2\hat{Y}_t(1) - Y_t + \theta_0 - \theta_2 e_t \\
\hat{Y}_t(\ell) &= 2\hat{Y}_t(\ell - 1) - \hat{Y}_t(\ell - 2) + \theta_0 \text{ for } \ell > 2
\end{aligned}$$

$$\hat{Y}_t(\ell) = A + B\ell + \frac{\theta_0}{2}\ell^2$$

where

$$\begin{aligned}
A &= 2\hat{Y}_t(1) - \hat{Y}_t(2) + \theta_0 \\
B &= \hat{Y}_t(2) - \hat{Y}_t(1) - \frac{3}{2}\theta_0
\end{aligned}$$

Summary of Forecasting with Certain ARIMA Models

For IMA(2,2),

- If $\theta_0 \neq 0$, the forecasts follow a quadratic curve in ℓ,
- but if $\theta_0 = 0$, the forecasts form a straight line with slope $\hat{Y}_t(2) - \hat{Y}_t(1)$ and will pass through the two initial forecasts $\hat{Y}_t(1)$ and $\hat{Y}_t(2)$.
- It can be shown that $Var(e_t(\ell))$ is a certain cubic function of ℓ; see Box, Jenkins, and Reinsel (1994, p. 156).
- We also have

$$\Psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2)j \text{ for } j > 0$$

- It can also be shown that forecasting the special case with $\theta_1 = 2\omega$ and $\theta_2 = -\omega^2$ is equivalent to so-called **double exponential smoothing** with smoothing constant $1 - \omega$; see Abraham and Ledolter (1983).

Summary

- Forecasting or predicting future as yet unobserved values is one of the main reasons for developing time series models.
- Methods discussed in this chapter are all based on minimizing the mean square forecasting error.
- When the model is simply deterministic trend plus zero mean white noise error, forecasting amounts to extrapolating the trend.
- However, if the model contains autocorrelation, the forecasts exploit the correlation to produce better forecasts than would otherwise be obtained. We showed how to do this with ARIMA models and investigated the computation and properties of the forecasts.

Summary

- In special cases, the computation and properties of the forecasts are especially interesting and we presented them separately.
- Prediction limits are especially important to assess the potential accuracy (or otherwise) of the forecasts.
- Finally, we addressed the problem of forecasting time series for which the models involve transformation of the original series.

Questions?