Times Series Analysis – Forecasting (I)

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

27th Oct., 2017

- One of the primary objectives of building a model for a time series is to be able to forecast the values for that series at future times.
- Of equal importance is the assessment of the precision of those forecasts
- We shall consider the calculation of forecasts and their properties for both deterministic trend models and ARIMA models.
- Forecasts for models that combine deterministic trends with ARIMA stochastic components are considered also.

Minimum Mean Square Error Forecasting

Based on the available history of the series up to time t, namely $Y_1, Y_2, \cdots, Y_{t-1}, Y_t$, we would like to forecast the value of Y_{t+l} that will occur ℓ time units into the future.

We call time t the **forecast origin** and ℓ the **lead time** for the forecast, and denote the forecast itself as $\hat{Y}_t(\ell)$.

According to the Minimum Mean Square Error Prediction theory, we know

$$\hat{Y}_t(\ell) = E(Y_{t+\ell}|Y_1, Y_2, \cdots, Y_t)$$

Deterministic Trend

Consider the deterministic trend model:

$$Y_t = \mu_t + X_t$$

where the stochastic component, X_t, has a mean of zero.

We shall assume that $\{X_t\}$ is in fact white noise with variance γ_0, then we have:

Deterministic Trend

$$
\begin{aligned}
\hat{Y}_t(\ell) &= E(\mu_{t+\ell} + X_{t+\ell} | Y_1, Y_2, \cdots, Y_t) \\
&= E(\mu_{t+\ell} | Y_1, Y_2, \cdots, Y_t) + E(X_{t+\ell} | Y_1, Y_2, \cdots, Y_t) \\
&= \mu_{t+\ell} + E(X_{t+\ell})
\end{aligned}
$$

or

$$\hat{Y}_t(\ell) = \mu_{t+\ell}$$

since for $\ell \geq 1$, $X_{t+\ell}$ is independent of $Y_1, Y_2, \cdots, Y_{t-1}, Y_t$ and has expected value zero.

Thus, in this simple case, forecasting amounts to extrapolating the deterministic time trend into the future.

Deterministic Trend

For the linear trend case, $\mu_t = \beta_0 + \beta_1 t$, the forecast is

$$\hat{Y}_t(\ell) = \beta_0 + \beta_1(t + \ell)$$

this model assumes that the same linear time trend persists into the future, and the forecast reflects that assumption.

Note that it is the lack of statistical dependence between $Y_{t+\ell}$ and $Y_1, Y_2, \cdots, Y_{t-1}, Y_t$ that prevents us from improving on $\mu_{t+\ell}$ as a forecast.

For seasonal models, where, say $\mu_t = \mu_{t+12}$, our forecast is $\hat{Y}_t(\ell) = \mu_{t+12+\ell} = \hat{Y}_t(\ell + 12)$. Thus the forecast will be also periodic.

Deterministic Trend

The **forecast error**, $e_t(\ell)$, is given by

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \hat{Y}_t(\ell) \\ &= \mu_{t+\ell} + X_{t+\ell} - \mu_{t+\ell} = X_{t+\ell} \end{aligned}$$

so that

$$E(e_t(\ell)) = E(X_{t+\ell}) = 0$$

That is, the forecasts are **unbiased**. Also

$$Var(e_t(\ell)) = Var(X_{t+\ell}) = \gamma_0$$

is the **forecast error variance** for all lead times ℓ.

The cosine trend model for the average monthly temperature series was estimated

$$\hat{\mu}_t = 46.2660 + (-26.7079)\cos(2\pi t) + (-2.1697)\sin(2\pi t)$$

Here time is measured in years with a starting value of January 1964, frequency $f = 1$ per year, and the final observed value is for December 1975. To forecast the June 1976 temperature value, we use $t = 1976.41667$ as the time value and obtain

$$\hat{\mu}_t = 46.2660 + (-26.7079)\cos(2\pi(1976.41667)) + (-2.1697)\sin(2\pi(1976.41667)) = 68.3^\circ F$$

ARIMA Forecasting

- For ARIMA models, the forecasts can be expressed in several different ways.
- Each expression contributes to our understanding of the overall forecasting procedure with respect to computing, updating, assessing precision, or long-term forecasting behavior.

AR(1):

Consider the simple AR(1) process with a nonzero mean that satisfies

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

Consider the problem of forecasting one time unit into the future. Replacing t by $t+1$, we have

$$Y_{t+1} - \mu = \phi(Y_t - \mu) + e_{t+1}$$

Given $Y_1, Y_2, \cdots, Y_t$, we take the conditional expectations of both sides and obtain

$$\hat{Y}_t(1) - \mu = \phi[E(Y_t | Y_1, Y_2, \cdots, Y_t) - \mu] + E(e_{t+1} | Y_1, Y_2, \cdots, Y_t)$$

ARIMA Forecasting

From the properties of conditional expectation, we have

$$E(Y_t | Y_1, Y_2, \cdots, Y_t) = Y_t$$

Also, since e_{t+1} is independent of $Y_1, Y_2, \cdots, Y_t$, we obtain

$$E(e_{t+1} | Y_1, Y_2, \cdots, Y_t) = E(e_{t+1}) = 0$$

Thus, we have

$$\hat{Y}_t(1) = \mu + \phi(Y_t - \mu)$$

In words, a proportion ϕ of the current deviation from the process mean is added to the process mean to forecast the next process value.

ARIMA Forecasting

Now consider a general lead time ℓ. Replacing t by $t+\ell$ and taking the conditional expectations of both sides produces

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell-1) - \mu] \qquad \text{for } \ell \geq 1$$

since $E(Y_{t+\ell-1}|Y_1, Y_2, \cdots, Y_t) = \hat{Y}_t(\ell-1)$ and, for $\ell \geq 1$, $e_{t+\ell}$ is independent of $Y_1, Y_1, \cdots, Y_t$.

ARIMA Forecasting

The above equation is recursive in the lead time ℓ. It shows how the forecast for any lead time ℓ can be built up from the forecasts for shorter lead times by starting with the initial forecast $\hat{Y}_t(\ell)$.

The forecast $\hat{Y}_t(2)$ is then obtained by $\hat{Y}_t(2) = \mu + \phi[\hat{Y}_t(1) - \mu]$, then $\hat{Y}_t(3)$ from $\hat{Y}_t(2)$, and so on until the desired $\hat{Y}_t(\ell)$ is found.

The above equation and its generalizations for other ARIMA models are most convenient for actually computing the forecasts. It is sometimes called the **difference equation form** of the forecasts.

ARIMA Forecasting

However, this equation can also be solved to yield an explicit expression for the forecast in terms of the observed history of the series. Iterating backward on ℓ, we have

$$\begin{aligned}\hat{Y}_t(\ell) &= \phi[\hat{Y}_t(\ell-1)-\mu]+\mu \\ &= \phi\left\{\phi[\hat{Y}_t(\ell-2)-\mu]\right\}+\mu \\ &\vdots \\ &= \phi^{\ell-1}[\hat{Y}_t(1)-\mu]+\mu\end{aligned}$$

so

$$\hat{Y}_t(\ell)=\mu+\phi^{\ell}(Y_t-\mu)$$

The current deviation from the mean is discounted by a factor ϕ^{ℓ}, whose magnitude decreases with increasing lead time. The discounted deviation is then added to the process mean to produce the lead ℓ forecast.

ARIMA Forecasting

As a numerical example, consider the AR(1) model that we have fitted to the industrial color property time series.

Coefficients:	ar1	intercept[†]
	0.5705	74.3293
s.e.	0.1435	1.9151

sigma^2 estimated as 24.8: log-likelihood = –106.07, AIC = 216.15

[†] Remember that the intercept here is the estimate of the process mean μ—not θ_0.

Figure: MLE of an AR(1) Model for Color.

For illustration purposes, we assume that the estimates $\phi = 0.5705$ and $\mu = 74.3293$ are true values. The final forecasts may then be rounded.

ARIMA Forecasting

The last observed value of the color property is 67, so we would forecast one time period ahead as

$$\begin{aligned} \hat{Y}_t(1) &= \hat{\mu} + \hat{\phi}[Y_t - \hat{\mu}] \\ &= 74.3293 + (0.5705)(67 - 74.3293) = 70.14793 \end{aligned}$$

For lead time 2, we have

$$\hat{Y}_t(2) = \hat{\mu} + \hat{\phi}[\hat{Y}_t(1) - \mu] = 71.94383$$

Alternatively, we have

$$\hat{Y}_t(2) = \hat{\mu} + \hat{\phi}^2[Y_t - \mu]$$

At lead 5, we have

$$\hat{Y}_t(2) = \hat{\mu} + \hat{\phi}^5[Y_t - \mu] = 73.88636$$

and $\hat{Y}_t(10) = 74.30253$ which is nearly $\hat{\mu} = 74.3293$

In general, since $|\phi| < 1$, we have simply

$$\hat{Y}_t(\ell) \approx \mu \qquad \text{for large} \quad t$$

Later we shall see that this equation holds *for all stationary* ARMA models.

Consider now the **one-step-ahead forecast error**, $e_t(1)$. We have

$$\begin{aligned} e_t(1) &= Y_{t+1} - \hat{Y}_t(1) \\ &= [\phi(Y_t - \mu) + \mu + e_{t+1}] - [\phi(Y_t - \mu) + \mu] \end{aligned}$$

or $e_t(1) = e_{t+1}$.

The white noise process $\{e_t\}$ can now be reinterpreted as a sequence of one-step-ahead forecast errors. We shall see that this equation persists for completely general ARIMA models.

Note also that $e_t(1) = e_{t+1}$ implies that the forecast error $e_t(1)$ is independent of the history of the process $Y_1, Y_2, \cdots, Y_t$ up to time t. If this were not so, the dependence could be exploited to improve our forecast.

It also implies that our one-step-ahead forecast error variance is given by

$$Var(e_t(1)) = \sigma_e^2$$

ARIMA Forecasting

To investigate the properties of the forecast errors for longer leads, it is convenient to express the AR(1) model in general linear process, or MA(∞), form.

Recall that

$$Y_t = \mu + e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots +$$

Together with $\hat{Y}_t(\ell) = \mu + \phi^\ell (Y_t - \mu)$, we have

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \mu - \phi^\ell (Y_t - \mu) \\ &= e_{t+\ell} + \phi e_{t+\ell-1} + \cdots + \phi^{\ell-1} e_{t+1} + \phi^\ell e_t + \cdots \\ &\quad -\phi^\ell (e_t + \phi e_{t-1} + \cdots) \end{aligned}$$

so that

$$e_t(\ell) = e_{t+\ell} + \phi e_{t+\ell-1} + \cdots + \phi^{\ell-1} e_{t+1}$$

which can also be written as

$$e_t(\ell) = e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \Psi_2 e_{t+\ell-2} + \cdots + \Psi_{\ell-1} e_{t+1}$$

We will show that this equation holds for all ARIMA models.

Note that $E(e_t(\ell)) = 0$; thus, the forecast are **unbiased**. Furthermore, from this equation, we have

$$Var(e_t(\ell)) = \sigma_e^2(1 + \Psi_1^2 + \cdots + \Psi_{\ell-1}^2)$$

We see that the forecast error variance increases as the lead ℓ increases. Contrast this with the result for deterministic trend models.

For the AR(1) case,

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

which we obtain by summing a finite geometric series. For long lead times, we have

$$Var(e_t(\ell)) \approx \frac{\sigma_e^2}{1-\phi^2}$$

for large ℓ. Notice that for AR(1), $\gamma_0 = \frac{\sigma_e^2}{1-\phi^2}$

$$Var(e_t(\ell)) \approx Var(Y_t) = \gamma_0 \qquad \text{for large } \ell.$$

This equation will be shown to be valid for *all stationary* ARMA process.

MA(1):

To illustrate how to solve the problems that arise in forecasting moving average or mixed models, consider the MA(1) case with nonzero mean:

$$Y_t = \mu + e_t - \theta e_{t-1}$$

Again replacing t by $t + 1$ and taking conditional expectations of both sides, we have

$$\hat{Y}_t(1) = \mu - \theta E(e_t | Y_1, Y_2, \cdots, Y_t)$$

For an invertible model, we have

$$Y_t = (-\theta_1 Y_{t-1} - \theta^2 Y_{t-2} - \cdots) + e_t$$

which indicates e_t is a function of $Y_1, \cdots, Y_t$, so

$$E(e_t | Y_1, Y_2, \cdots, Y_t) = e_t$$

In fact, an approximation is involved in this equation since we are conditioning only on $Y_1, Y_2, \cdots, Y_t$ and not on the infinite history of the process.

However, if, as in practice, t is large and the model is invertible, the error in the approximation will be very small. If the model is not invertible — for example, if we have overdifferenced the data — then this equation is not even approximately valid.

ARIMA Forecasting

Previous equations also give us the one-step-ahead forecast for an invertible MA(1) expressed as

$$\hat{Y}_t(1) = \mu - \theta e_t$$

The computation of e_t will be a by-product of estimating the parameters in the model.

The one-step-ahead forecast error is:

$$\begin{aligned} e_t(1) &= Y_{t_1} - \hat{Y}_t(1) \\ &= (\mu + e_{t+1} - \theta e_t) - (\mu - \theta e_t) = e_{t+1} \end{aligned}$$

ARIMA Forecasting

For longer lead time, we have

$$\hat{Y}_t(\ell) = \mu + E(e_{t+\ell}|Y_1, Y_2, \cdots, Y_t) - \theta E(e_{t+\ell-1}|Y_1, Y_2, \cdots, Y_t)$$

But for $\ell > 1$, both $e_{t+\ell}$ and $e_{t+\ell-1}$ are independent of $Y_1, \cdots, Y_t$. Consequently, these conditional expected values are the unconditional expected values, namely zero, and we have

$$\hat{Y}_t(\ell) = \mu \qquad \text{for} \quad \ell > 1.$$

Note that for large ℓ, the forecast for AR(1) model is approximately to μ

The Random Walk with Drift:

To illustrate forecasting with nonstationary ARIMA series, consider the random walk with drift defined by

$$Y_t = Y_{t-1} + \theta_0 + e_t$$

Here

$$\hat{Y}_t(1) = E(Y_t|Y_1, Y_2, \cdots, Y_t) + \theta_0 + E(e_{t+1}|Y_1, Y_2, \cdots, Y_t)$$

so that

$$\hat{Y}_t(\ell) = Y_t + \theta_0$$

ARIMA Forecasting

Similarly, the difference equation form for the lead ℓ forecast is

$$\hat{Y}_t(\ell) = \hat{Y}_t(\ell - 1) + \theta_0 \qquad \text{for} \quad l \geq 1$$

and iterating backward on ℓ yields the explicit expression

$$\hat{Y}_t(\ell) = Y_t + \theta_0 \ell \qquad \text{for} \quad \ell \geq 1.$$

Different from the forecast for AR(1) model $\hat{Y}_t(\ell) \approx \mu$ for large ℓ. If $\theta \neq 0$, the forecast does not converge for long leads but rather follows a straight line with slope θ_0 for all ℓ.

ARIMA Forecasting

- Note that the presence or absence of the constant term θ_0 significantly alters the nature of the forecast.
- For this reason, constant terms should not be included in nonstationary ARIMA models unless the evidence is clear that the mean of the differenced series is significantly different from zero.
- The following equation

$$Var(\bar{Y}) = \frac{\gamma_0}{n} + \frac{2}{n} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \gamma_k \right) = \frac{1}{n} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n} \right) \gamma_k$$

for the variance of the sample mean will help assess this significance.

ARIMA Forecasting

as we have seen in the AR(1) and MA(1) cases, the one-step-ahead forecast error is

$$e_t(1) = Y_{t+1} - \hat{Y}_t(1) = e_{t+1}$$

Also

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

which agrees with previous equation for AR model since in this model $\Psi_j = 1$ for all j.

So we have

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

In contrast to the stationary case, here $Var(e_t(\ell))$ grows without limit as the forecast lead time ℓ increases. We shall see that this property is characteristic of the forecast error variance for *all nonstationary* ARIMA processes.

ARMA(p, q):

For the general stationary ARMA(p, q) model, the difference equation form for computing forecasts is given by:

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

where

$$E(e_{t+j}|Y_1, Y_2, \cdots, Y_t) = \begin{cases} 0 & \text{for } \; j > 0 \\ e_{t+j} & \text{for } \; j \le 0 \end{cases}$$

We note that $\hat{Y}_t(j)$ is a true forecast for $j > 0$,but for $j \le 0$, $\hat{Y}_t(j) = Y_{t+j}$.

ARIMA Forecasting

Previous equation involves some minor approximation. For an invertible model, we know that using the π-weights, e_t can be expressed as a linear combination of the infinite sequence $Y_t, Y_{t-1}, Y_{t-2}, \cdots$. However, the π-weights die out exponentially fast, and the approximation assumes that π_j is negligible for $j > t - q$.

ARIMA Forecasting

Consider an ARMA(1,1) model, we have

$$\hat{Y}_t(1) = \phi Y_t + \theta_0 - \theta e_t$$

with $\hat{Y}_t(2) = \phi \hat{Y}_t(1) + \theta_0$

more generally,

$$\hat{Y}_t(\ell) = \phi \hat{Y}_t(\ell - 1) + \theta_0 \qquad \text{for} \quad \ell \geq 2$$

using previous equation to get the recursion started.

ARIMA Forecasting

Previous equations can be written in terms of the process mean and then solved by iteration to get the alternative explicit expression

$$\hat{Y}_t(\ell) = \mu + \phi^{\ell}(Y_t - \mu) - \phi^{\ell-1} e_t \qquad \text{for} \quad \ell \geq 1$$

The noise terms $e_{t-(q-1)}, \cdots, e_{t-1}, e_t$ appear directly in the computation of the forecasts for leads $\ell = 1, 2, \cdots, q$. However, for $\ell > q$, the AR portion of the difference equation takes over and we have

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

ARIMA Forecasting

Thus the general nature of the forecast for long lead times will be determined by the autoregressive parameters $\phi_1, \phi_2, \cdots, \phi_p$ (and the constant term, θ_0, which is related to the mean of the process).

Recall that $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$, we can write previous equation in terms of deviations from μ as

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

ARIMA Forecasting

- As a function of lead time ℓ, $\hat{Y}_t(\ell) - \mu$ follows the same Yule-Walker recursion as the autocorrelation function ρ_k of the process. Recall that for an ARMA model, the autocorrelation function satisfies

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \text{ for } \quad k > q.$$

- Thus, the roots of the characteristic equation will determine the general behavior of $\hat{Y}_t(\ell) - \mu$ for large lead times.
- In particular, $\hat{Y}_t(\ell) - \mu$ 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).

ARIMA Forecasting

- Thus, for any stationary ARMA model, $\hat{Y}_t(\ell) - \mu$ decays to zero as ℓ increases, and the long-term forecast is simply the process mean μ as given in $\hat{Y}_t(\ell) \approx \mu$
- This agrees with common sense since for stationary ARMA models the dependence dies out as the time span between observations increases, and this dependence is the only reason we can improve on the "naive" forecast of using μ alone.

ARIMA Forecasting

To argue the validity of

$$Var(e_t(\ell)) = \sigma_e^2(1 + \Psi_1^2 + \cdots + \Psi_{\ell-1}^2)$$

in the present generality, we need to consider a new representation for ARIMA processes. Appendix G shows that any ARIMA model can be written in **truncated linear process form** as

$$Y_{t+\ell} = C_t(\ell) + I_t(\ell) \text{ for } \ell > 1$$

where for our present purposes, we need only know that $C_t(\ell)$ is a certain function of $Y_t, Y_{t-1}, \cdots$ and

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

ARIMA Forecasting

Furthermore, for invertible models with t reasonably large, $C_t(\ell)$ is a certain function of the finite history $Y_t, Y_{t-1}, \cdots, Y_1$. Thus we have

$$\hat{Y}_t(\ell) = E(C_t(\ell)|Y_1, \cdots, Y_t) + E(I_t(\ell)|Y_1, \cdots, Y_t) = C_t(\ell)$$

Finally,

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \hat{Y}_t(\ell) \\ &= [C_t(\ell) + I_t(\ell)] - C_t(\ell) = I_t(\ell) \\ &= e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \Psi_{\ell-1} e_{t+1} \end{aligned}$$

Thus for a general invertible ARIMA process,

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

Recall that for a general linear process $Y_t = e_t + \Psi_1 e_{t-1} + \cdots$, we have

$$Cov(Y_t, Y_{t-k}) = \sigma_e^2 \sum_{i=0}^{\infty} \Psi_i \Psi_{i+k}$$

We see that for long lead times in stationary ARMA models, we have

$$Var(e_t(\ell)) \approx \sigma_e^2 \sum_{j=0}^{\infty} \Psi_j^2 \quad \text{or} \quad Var(e_t(\ell)) \approx \gamma_0 \quad \text{for large} \quad \ell$$

Questions?