Forecasting with ARIMA Models:

Part II

Video 34

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Video 34 Learning Objectives

By the end of this video, we should be able to:

- Understand when noise terms do or don't appear in an ARIMA forecast equation
- Obtain a forecast for the original process from the forecast for a transformed version of the process, for certain special cases
- Calculate the prediction limits for a forecast from an ARIMA model

Forecasting for ARMA(p,q): Review

Suppose we are forecasting a future value $Y_{t+\ell}$ ($\ell > 0$), using our time series dataset $Y_1, Y_2, \dots, Y_t$.

The minimum MSE forecast of $Y_{t+\ell}$ is:

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

where

$$\hat{Y}_t(j) = \begin{cases} \text{the actual forecast } \hat{Y}_t(j) & \text{if } j > 0 \\ \text{the observed value } Y_{t+j} & \text{if } j \le 0 \end{cases}$$

and

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

Forecasting for ARMA(p,q): Models with MA Terms

Therefore, for $\ell \leq q$, the forecasting equation will have some noise term(s).

We learned about one possible way of estimating these noise terms in Video 30:

Example: For an MA(1) model, $Y_t - \mu = e_t - \theta e_{t-1}$:

1. Note that the process can be re-written as

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

2. Obtain $e_1, e_2, \ldots, e_t$ recursively, using the assumption that $e_0 = 0$:

$$\begin{aligned} e_1 &= Y_1 - \mu + \theta e_0 = Y_1 - \mu \\ e_2 &= Y_2 - \mu + \theta e_1 \\ &\vdots \\ e_t &= Y_{t-1} - \mu + \theta e_{t-1} \end{aligned}$$

However, for lead times $\ell > q$, the AR part of the model takes over, and there are no more noise terms.

Forecasting Transformed Series: Differences

In Video 33, we discussed the approach to forecasting for an ARIMA process:

1. Write out the difference equation form for the non-stationary process $\{Y_t\}$
2. Use the same equations for $\hat{Y}_t(\ell)$ and $Var(e_t(\ell))$ that you would use for a stationary ARMA process

Forecasting Transformed Series: Differences

In Video 33, we discussed the approach to forecasting for an ARIMA process:

1. Write out the difference equation form for the non-stationary process $\{Y_t\}$
2. Use the same equations for $\hat{Y}_t(\ell)$ and $Var(e_t(\ell))$ that you would use for a stationary ARMA process

Alternative approach:

1. Obtain the forecast on the stationary ARMA process $\{W_t\}$ instead $\implies \hat{W}_t(\ell)$
2. Since $\{W_t\} = \{\nabla^d Y_t\}$, "undo" the difference by summing, in order to obtain the forecast $\hat{Y}_t(\ell)$
   - For example, for $d = 1$:

$$\hat{Y}_t(\ell) = \hat{W}_t(\ell) + \hat{Y}_t(\ell - 1)$$
$$\hat{Y}_t(1) = \hat{W}_t(1) + Y_t$$

§ 9.8 gives an example of how these two approaches are equivalent.

Forecasting Transformed Series: Other Transformations

What about other transformations (e.g., logs)?

Suppose $\{Y_t\}$ is non-stationary, but $W_t = \log(Y_t)$ is a stationary ARMA process.

Forecasting Transformed Series: Other Transformations

What about other transformations (e.g., logs)?

Suppose $\{Y_t\}$ is non-stationary, but $W_t = \log(Y_t)$ is a stationary ARMA process.

We obtain the forecast $\hat{Y}_t(\ell)$ as follows:

1. Obtain the forecast on the stationary ARMA process: $\hat{W}_t(\ell)$
2. Since $W_t = \log(Y_t)$, a "naive" forecast for $Y_{t+\ell}$ would be $\exp[\hat{W}_t(\ell)]$.

   However, since the logarithm is a non-linear transformation, this will *not* be the minimum MSE forecast for $Y_{t+\ell}$!

   Instead, we should use (pg. 210):

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

   where $e_t(\ell) = W_{t+\ell} - \hat{W}_t(\ell)$, and its variance is obtained using the usual equations.

Prediction Limits

We have learned how to obtain the forecast of a future value $Y_{t+\ell}$.

We would now like to assess the precision of the forecast.

Prediction Limits

We have learned how to obtain the forecast of a future value $Y_{t+\ell}$.

We would now like to assess the precision of the forecast.

Recall: The *forecast error*, $e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell)$, can be shown to be equal to

$$e_t(\ell) = \sum_{j=0}^{\ell-1} \psi_j e_{t+\ell-j}$$

where $\psi_0 = 1$, and $\psi_1, \psi_2, \ldots, \psi_{\ell-1}$ are the first $\ell - 1$ coefficients in the general linear process formulation of the ARMA model.

If the white noise terms $\{e_t\}$ are assumed to be iid normal, then $e_t(\ell)$ is also normal.

Therefore: $e_t(\ell) \sim \mathcal{N}(0, Var(e_t(\ell)))$.

Prediction Limits (cont'd)

Therefore:

$$\Pr\left[-z_{\alpha/2} < \frac{e_t(\ell) - 0}{\sqrt{Var(e_t(\ell))}} < z_{\alpha/2}\right] = 1 - \alpha$$

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

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

where $z_{\alpha/2}$ is the $(1-\alpha/2)^{\text{th}}$ percentile of the standard normal distribution.
(For example, for a probability of $1 - \alpha = 0.95$, we use $z_{\alpha/2} = z_{0.025} = 1.96$.)

Prediction Limits (cont'd)

Therefore:

$$\Pr\left[-z_{\alpha/2} < \frac{e_t(\ell) - 0}{\sqrt{Var(e_t(\ell))}} < z_{\alpha/2}\right] = 1 - \alpha$$

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

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

where $z_{\alpha/2}$ is the $(1-\alpha/2)^{\text{th}}$ percentile of the standard normal distribution. (For example, for a probability of $1 - \alpha = 0.95$, we use $z_{\alpha/2} = z_{0.025} = 1.96$.)

So, the **$100(1-\alpha)\%$ prediction limits** for $Y_{t+\ell}$ are:

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

Recall that, for each of our examples in Video 33, we practiced obtaining $Var(e_t(\ell))$! Also, $Var(e_t(\ell)) \approx \gamma_0$ for large ℓ, for any stationary ARMA process.

Example 1: Color Dataset

Example: We fit an AR(1) model to the color dataset:

```
data(color)
color.ar1.model <- arima(color, order=c(1,0,0))
color.ar1.model
```

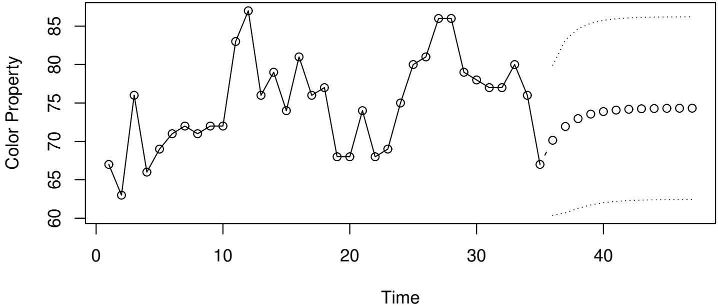
$\hat{\phi} = 0.57$	$\hat{\mu} = 74.33$	$\hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}) = 31.92$

$$\implies Y_t = 31.92 + 0.57 Y_{t-1} + e_t$$

Example 1: Color Dataset (cont'd)

Forecasting future values:

```
plot(color.ar1.model, n.ahead=12, type='b', xlab='Time',
     ylab='Color Property')
```



Example 1: Color Dataset (cont'd)

There are $t = 35$ observations in the dataset, and the final observation is $Y_t = 67$.

For example:

Forecast of Y_{40}:

$$\hat{Y}_{35}(5) = \hat{\mu} + (\hat{\phi})^{\ell}(Y_{35} - \hat{\mu}) = 74.3293 + (0.5705)^5(67 - 74.3293) = 73.88$$

Forecast error variance for Y_{40}:

$$Var(e_{35}(5)) = \hat{\sigma}_e^2 \left(\frac{1 - (\hat{\phi})^{2\ell}}{1 - (\hat{\phi})^2} \right) = (25.0259) \left(\frac{1 - (0.5705)^{10}}{1 - (0.5705)^2} \right) = 38.29$$

since $\hat{\sigma}_e^2 = (1 - \hat{\phi} r_1) s^2 = (1 - 0.5705 \times 0.5282)(37.1042) = 25.03$.

Example 1: Color Dataset (cont'd)

$\therefore$ 95% prediction limits for Y_{40} are:

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

$$73.8862 \pm 1.96 \times \sqrt{38.2926}$$

$$73.8862 \pm 12.1287$$

$$\implies (61.76,\ 86.01)$$

Example 2: Hare Dataset

Example: We fit an AR(3) model to the square root of the hare dataset (see pg. 137):

```
data(hare)
hare.ar3.model <- arima(sqrt(hare), order=c(3,0,0))
hare.ar3.model
```

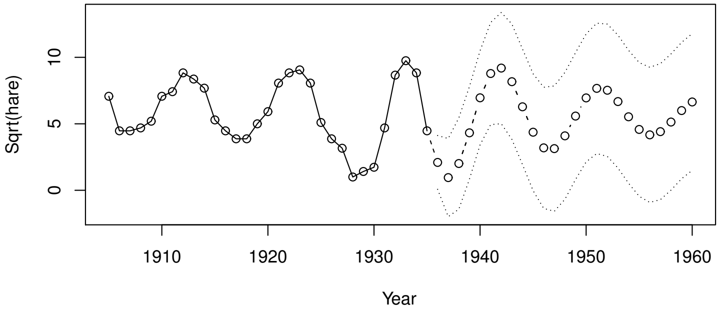
$\hat{\phi}_1 = 1.05$	$\hat{\phi}_2 = -0.23$	$\hat{\mu} = 5.69$	$\hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \hat{\phi}_2) = -6.28$

$$\implies \sqrt{Y_t} = -6.28 + 1.05\sqrt{Y_{t-1}} - 0.23\sqrt{Y_{t-2}} + e_t$$

Example 2: Hare Dataset (cont'd)

Forecasting future values:

```
plot(hare.ar3.model, n.ahead=25, type='b', xlab='Year', ylab='Sqrt(hare)')
```



Note: Obtaining forecasts of $Y_{t+\ell}$ would require a more complicated transformation than just $(\hat{W}_t(\ell))^2$, since it is a non-linear transformation.

Final Comments

That's all for now!

In this video, we've learned about some more important topics in ARIMA forecasting.

We've discussed the effect of noise terms in the forecasts, and how to consider certain types of transformations.

We've also shown how to obtain prediction limits as a way of assessing the precision of the forecasts.

Next Week in STAT 485/685: Review for the final exam.

Thank you!

References:

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

[2] Chan, K. S., & Ripley, B. (2020). TSA: Time Series Analysis. R package version 1.2.1.