Lectures 13-14

Forecasting
Textbook Sections: FILL IN

General Setting

A typical goal of time series is to use the information available from time $t = 1, \cdots, n$ to obtain forecasts for time $t = n + h$, where $h = 1, 2, 3, \cdots$. Sometimes h is called the forecast horizon

Define $\hat{X}_{n+h}$ to be the **h-step ahead forecast**.

For any forecasting method, the quantity $\hat{\varepsilon}_t = X_t - \hat{X}_t$ is called an **innovation**.

We would also like to assess the variability of our forecast by constructing prediction intervals. To do this, we need to estimate the **variance of the (h-step) forecast error**,

$$\sigma^2(h) = Var(X_{n+h} - \hat{X}_{n+h}) = E(X_{n+h} - \hat{X}_{n+h})^2.$$

There is ample methodology to carry all this out for stationary, causal $ARMA(p, q)$ models. We will use R to do this in practice. The purpose of this lecture is to take a look at the way this is done in a few simple cases.

Forecasting With an AR Model

Forecasting with an $AR(p)$ process is particularly simple. We assume that we have observations of $X_1, \ldots, X_n$ from an $AR(p)$ sequence. The goal is to forecast x_{n+h} for some $h = 1, 2, 3, \cdots$. Since an autoregressive model is like a standard regression model, forecasting is rather straightforward. It is accomplished by simply plugging the previous observations into the model equation with the true or estimated coefficients.

Case 1. Suppose the true AR coefficients $\phi_1, \cdots, \phi_p$ are known. In this case, X_{n+h} is defined as

$$X_{n+h} = \phi_1 X_{n+h-1} + \cdots + \phi_p X_{n+h-p} + \varepsilon_{n+h},$$

and we can define the predictor as

$$\hat{X}_{n+h} = \phi_1 X_{n+h-1} + \cdots + \phi_p X_{n+h-p}.$$

The specific forecast values can be obtained by substituting the observed values $x_{n+h-1}, \cdots x_{n+h-p}$ into the expression.

Case 2. Now suppose the true AR coefficients $\phi_1, \cdots, \phi_p$ are not known. This is usually the case in practice. We must first fit a model to the data, and obtain the estimates $\hat{\phi}_1, \cdots, \hat{\phi}_p$. We will cover some technical details of this estimation in a few lectures. For now, let us just accept that we have the methodology to fit a stationary $AR(p)$ model to the data.

As before, X_{n+h} is defined as

$$X_{n+h} = \phi_1 X_{n+h-1} + \cdots + \phi_p X_{n+h-p} + \varepsilon_{n+h},$$

and we define the forecast using our estimated coefficients as

$$\hat{X}_{n+h} = \hat{\phi}_1 X_{n+h-1} + \cdots + \hat{\phi}_p X_{n+h-p}.$$

The specific forecast values can be obtained by substituting the observed values $x_{n+h-1}, \cdots x_{n+h-p}$ into the expression.

Note

Sometimes we need to rely on forecasts of previous time points to forecast farther into the future.

For example, suppose we fit an $AR(3)$ model, and want to obtain forecasts for $h = 1, \cdots, 7$. Since the order of the model is $p = 3$, each fitted value is computed from the previous three observed values. However, as we forecast into the future, we won't always have actual observations of the previous time points.

- Compute $\hat{X}_{n+1}$ from X_{n-2}, X_{n-1}, and X_n.
- Compute $\hat{X}_{n+2}$ from X_{n-1}, X_n, and $\hat{X}_{n+1}$.
- Compute $\hat{X}_{n+3}$ from X_n, $\hat{X}_{n+1}$, and $\hat{X}_{n+2}$.
- Compute $\hat{X}_{n+4}$ from $\hat{X}_{n+1}$, $\hat{X}_{n+2}$, and $\hat{X}_{n+3}$.

Forecasting With an MA Model

Forecasting with a moving average model is somewhat more complicated than forecasting with an autoregressive model. Fortunately, the computer packages do them, but it is important to know how the computer calculates the forecasts. For the purpose of discussion here, we will consider $MA(1)$ and $MA(2)$ models.

First consider a $MA(1)$ model $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$, where $\{\varepsilon_t\}$ is white noise. If μ and θ were known, our forecast based on $\{X_1, ..., X_{t-1}\}$ is $\hat{X}_t = \mu + \theta\varepsilon_{t-1}$ and the innovation is $X_t - \hat{X}_t = \varepsilon_t$.

For simplicity, let us assume that $\mu = 0$, $\theta = 0.5$ and we have a series of length $n = 5$.
Suppose the observed values of $\{X_1, ..., X_5\}$ are $\{1.5, 2.1, -1.9, -2.2, 0.4\}$.
We are interested in predicting $X_{n+1} = X_6$.
If we could know the value of ε_5, then the forecasted value of X_6 is $\hat{X}_6 = 0.5\varepsilon_5$.
Unfortunately, we do not know ε_5. So how do we guess it?
Note that $\varepsilon_5 = X_5 - 0.5\varepsilon_4$, and we observe $x_5 = 0.4$.
Therefore, we would know ε_5 if could obtain the value of ε_4.
The same argument will show that we could know ε_4 if we were able to obtain the value of ε_3.
A repetition of this argument will show that we could know $\varepsilon_1 = X_1 - 0.5\varepsilon_0$ if we knew ε_0.
So in order to predict X_6, in addition to the data $\{x_1, ..., x_5\}$, we also need to know ε_0.

Suppose we make a guess $\hat{\varepsilon}_0$ for ε_0 (the usual guess is $\hat{\varepsilon}_0 = 0$, though other complicated methods are also possible). Let us take $\hat{\varepsilon}_0 = 0$ here. Then we can use the expressions $\hat{X}_t = \theta\varepsilon_{t-1}$ and $\varepsilon_t = X_t - \theta\varepsilon_{t-1} = X_t - \hat{X}_t$ to compute the following:

$$\begin{aligned}
&\hat{X}_1 = \theta\hat{\varepsilon}_0 = 0, && \hat{\varepsilon}_1 = X_1 - \hat{X}_1 = 1.5 - 0 = 1.5,\\
&\hat{X}_2 = \theta\hat{\varepsilon}_1 = (0.5)(1.5) = 0.75, && \hat{\varepsilon}_2 = X_2 - \hat{X}_2 = 2.1 - 0.75 = 1.35,\\
&\hat{X}_3 = \theta\hat{\varepsilon}_2 = (0.5)(1.35) = 0.675, && \hat{\varepsilon}_3 = X_3 - \hat{X}_3 = -1.9 - 0.675 = -2.575,\\
&\hat{X}_4 = \theta\hat{\varepsilon}_3 = (0.5)(-2.575) = -1.2875, && \hat{\varepsilon}_4 = X_4 - \hat{X}_4 = -2.2 - (-1.2875) = -0.9125,\\
&\hat{X}_5 = \theta\hat{\varepsilon}_4 = (0.5)(-0.9125) = -0.4563, && \hat{\varepsilon}_5 = X_5 - \hat{X}_5 = 0.4 - (-0.4563) = 0.8563.
\end{aligned}$$

So our predicted value of X_6 is

$$\hat{X}_6 = \theta\hat{\varepsilon}_5 = (0.5)(0.8563) = 0.4282.$$

How does one predict X_{n+2}, X_{n+3} etc.?
Note that the predicted value of X_{n+2} is $\theta\varepsilon_{n+1}$. Now, an estimate of ε_{n+1} is $X_{n+1} - \hat{X}_{n+1}$. Since we do not have X_{n+1}, we have to substitute it by $\hat{X}_{n+1}$. So $\hat{\varepsilon}_{n+1} = \hat{X}_{n+1} - \hat{X}_{n+1} = 0$. So the predicted value of X_{n+2} is $\theta\hat{\varepsilon}_{n+1} = 0$. As a matter of fact, $\hat{X}_{n+2}, \hat{X}_{n+3}, ...$ are all zeros.

If $\mu \neq 0$, then the estimate of ε_t at each stage will be $\hat{\varepsilon}_t = X_t - \mu - \theta\hat{\varepsilon}_{t-1}$. The argument is basically the same as before by taking $X_t - \mu$ instead of X_t. So one can first create the centered values $X_t - \mu$ and then apply the arguments as before with the centered values $X_t - \mu$.

For the MA(2) case, a similar argument will show that in order to guess X_{n+1} on the basis of the data $\{X_1, ..., X_n\}$, we will have to guess ε_{-1} and ε_0 (which are often taken to be equal to zero), then obtain estimates of $\varepsilon_1, ..., \varepsilon_n$. Then the guess for X_{n+1} is $\hat{X}_{n+1} = \mu + \theta_1\hat{\varepsilon}_n + \theta_2\hat{\varepsilon}_{n-1}$. We can now estimate ε_{n+1} as $\hat{\varepsilon}_{n+1} = X_{n+1} - \hat{X}_{n+1}$. As in the MA(1) case, we substitute X_{n+1} by $\hat{X}_{n+1}$, thus leading to $\hat{\varepsilon}_{n+1} = 0$. So the predicted value of X_{n+2} is $\hat{X}_{n+2} = \mu + \theta_1\hat{\varepsilon}_{n+1} + \theta_2\hat{\varepsilon}_n = \mu + \theta_2\hat{\varepsilon}_n$. It is easy to check that subsequent predicted values $\hat{X}_{n+3}, \hat{X}_{n+4}, ...$ are all zeros.

Remark 1: For the MA(1) case, the guess for X_{n+1} is $\hat{X}_{n+1} = \mu + \theta\hat{\varepsilon}_n$, where $\hat{\varepsilon}_n$ is obtained the way as described before. The question how sensitive is the value of $\hat{\varepsilon}_n$ on the initial guess. The answer is "not much" if n is large and $-1 < \theta < 1$. An $MA(1)$ model with $-1 < \theta < 1$, is called an "invertible". The issue of "invertibility" will be discussed soon. If $\theta = 1$ or -1, then the moving average model is called "non-invertible". In such a case, it is still possible to get a prediction of X_{n+1}, but is considerably more complicated. For instance, if $\theta = -1$, then the best predictor of X_{n+1} is of the form

$$\hat{X}_{n+1} = \mu + \sum_{j=0}^{n-1} w_{n,j}(X_{n-j} - \mu), where\ w_{n,j} = 1 - (j+1)/(n+1).$$

Forecasting With an ARMA Model

The method for forecasting ARMA models has the same type of difficulty associated with that for MA models. Fortunately, computer packages carry them out. But let us understand the issues that come up and they mirror the MA case.

Let us assume that we have an $ARMA(1,1)$ model $X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$ and we want to predict X_{n+1} given observed values of $X_1, ..., X_n$. Note the best predictor for X_{n+1} is $0.8X_n + 0.5\varepsilon_n$. The value X_n is known, but not the value of ε_n. As in$MA(1)$ case, we can argue that in order to estimate the value of ε_n we need an estimate of ε_{n-1} which in turn requires an estimate of ε_{n-2} an so on. Ultimately, as in the $MA(1)$ case, we can see that if we can guess the value of ε_1 then we can make a guess of the value of ε_n. The method is, first guess a value $\hat{\varepsilon}_1$ of ε_1 (often $\hat{\varepsilon}_1$ is taken to be equal to zero), then successively estimate $\varepsilon_2, \varepsilon_3, ..$ etc. as

$$
\begin{aligned}
&\hat{X}_2 = 0.8X_1 + 0.5\hat{\varepsilon}_1, \hat{\varepsilon}_2 = X_2 - \hat{X}_2,\\
&\hat{X}_3 = 0.8X_2 + 0.5\hat{\varepsilon}_2, \hat{\varepsilon}_3 = X_3 - \hat{X}_3,\\
&\quad\vdots\\
&\hat{X}_n = 0.8X_{n-1} + 0.5\hat{\varepsilon}_{n-1}, \hat{\varepsilon}_n = X_n - \hat{X}_n.
\end{aligned}
$$

So the predicted value of X_{n+1} is $\hat{X}_{n+1} = 0.8X_n + 0.5\hat{\varepsilon}_n$. Now $\hat{\varepsilon}_{n+1} = X_{n+1} - \hat{X}_{n+1}$. Since X_{n+1} is unknown, we estimate X_{n+1} by $\hat{X}_{n+1}$ leading to $\hat{\varepsilon}_{n+1} = 0$. Thus predicted value of X_{n+2} is $\hat{X}_{n+2} = 0.8\hat{X}_{n+1} + 0.5\hat{\varepsilon}_{n+1} = 0.8\hat{X}_{n+1}$. Similarly, $\hat{X}_{n+3} = 0.8\hat{X}_{n+2}$ and so on.

Usefulness of Invertibility

Recall that a random sequence $\{X_t\}$ is called **invertible** (or possesses the property of **invertibility**) if it can be written as an $AR(\infty)$ sequence

$$\sum_{j=0}^{\infty} \pi_j (X_{t-j} - \mu) = \varepsilon_t, \text{ with } \pi_0 = 1,$$

where $\{\varepsilon_t\}$ is white noise.

If we know a sequence $\{X_t\}$, and we know the weights π_j, then computing forecasts is straightforward as in the $AR(p)$ case. For notational simplicity, let us assume that $\mu = 0$. Then we can write this sequence as

$$X_t = \varepsilon_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \cdots.$$

So if we have the data $\{X_1, ..., X_n\}$, then the forecasted value of of X_{n+1} is

$$\hat{X}_{n+1} = -\pi_1 X_n - \pi_2 X_{n-1} - \pi_3 X_{n-2} - \cdots.$$

Similarly, our forecast of X_{n+2} is

$$\hat{X}_{n+2} = -\pi_1 X_{n+1} - \pi_2 X_n - \pi_3 X_{n-1} - \cdots.$$

Since x_{n+1} is unknown, then we can substitute it by $\hat{X}_{n+1}$, thus leading to the forecasted value of X_{n+2} as

$$\hat{X}_{n+2} = -\pi_1 \hat{X}_{n+1} - \pi_2 X_n - \pi_3 X_{n-1} - \cdots.$$

This method can now be replicated to forecast X_{n+3}, X_{n+4}, etc.

Typically μ is not equal to zero, but obtaining forecasts is still not difficult with the known π_j values. For instance, the forecasting formula for X_{n+1} in the presence of μ is given by

$$\hat{X}_{n+1} - \mu = -\pi_1 (X_n - \mu) - \pi_2 (X_{n-1} - \mu) - \pi_3 (X_{n-2} - \mu) - \cdots.$$

Prediction Intervals and Usefulness of Causality

We have talked about forecasting a series, but have not addressed the issue of prediction limits (or prediction intervals).

If the observed series is $\{X_1, ..., X_n\}$, then h step ahead forecast is denoted by $\hat{X}_{n+h}$. The forecast error is $X_{n+h} - \hat{X}_{n+h}$ which is not known since X_{n+h} is unknown.

For all the cases we consider in this course, the mean of the forecast error is equal to zero (or close to zero when the parameters of the model being used for forecasting are estimated). The variance of the forecast error is denoted by

$$\sigma^2(h) = Var(X_{n+h} - \hat{X}_{n+h}) = E(X_{n+h} - \hat{X}_{n+h})^2.$$

If an estimate $\sigma^2(1)$ of the variance of the error for forecasting $\hat{X}_{n+1}$ is available, we can give a 95% prediction interval for X_{n+1} as $\hat{X}_{n+1} \pm 1.96\sigma(1)$. We can do the same for X_{n+2} as $\hat{X}_{n+2} \pm 1.96\sigma(2)$, where $\sigma^2(2)$ is the variance of the error for forecasting X_{n+2}, and so on.

However getting the values $\sigma^2(1), \sigma^2(2)$ etc. is not easy except in the $MA(q)$ case. This is where the causal representation is useful. Recall that a random sequence $\{X_t\}$ is called **causal** (or possesses the property of **causality**) if it can be written as an $MA(\infty)$ sequence

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \text{ with } \psi_0 = 1,$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2 and $\psi_1, \psi_2, ...$ are constants satisfying the constraint $\sum |\psi_j| < \infty$.

We want to predict X_{n+1}, X_{n+2},... along with the prediction intervals from the observed series $\{X_1, ..., X_n\}$. Suppose that we want to predict X_{n+1} given the data $\{X_1, ..., X_n\}$. Since

$$X_{n+1} - \mu = \varepsilon_{n+1} + \psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \psi_3\varepsilon_{n-2} + \cdots.$$

Then the best predictor of X_{n+1} is

$$\hat{X}_{n+1} - \mu = \psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \psi_3\varepsilon_{n-2} + \cdots. \quad (1)$$

Hence the variance of the forecast error is

$$\sigma^2(1) = E(X_{n+1} - \hat{X}_{n+1})^2 = E\varepsilon_{n+1}^2 = \sigma^2.$$

Note that X_{n+2} has the representation

$$X_{n+2} - \mu = \varepsilon_{n+2} + \psi_1\varepsilon_{n+1} + \psi_2\varepsilon_n + \psi_3\varepsilon_{n-1} + \cdots,$$

and hence the best linear predictor of X_{n+2} based on the past up to time n is

$$\hat{X}_{n+2} - \mu = \psi_2\varepsilon_n + \psi_3\varepsilon_{n-1} + \cdots, and$$
$$X_{n+2} - \hat{X}_{n+2} = \varepsilon_{n+2} + \psi_1\varepsilon_{n+1}.$$

So the variance of the forecast error $X_{n+2} - \hat{X}_{n+2}$ is

$$\sigma^2(2) = E(X_{n+2} - \hat{X}_{n+2})^2 = E(\varepsilon_{n+2} + \psi_1\varepsilon_{n+1})^2 = \sigma^2 + \psi_1^2\sigma^2 = (1 + \psi_1^2)\sigma^2.$$

A similar argument will show that the variances for prediction error for X_{n+3}, X_{n+4} are given by

$$\sigma^2(3) = E(X_{n+3} - \hat{X}_{n+3})^2 = (1 + \psi_1^2 + \psi_2^2)\sigma^2,$$
$$\sigma^2(4) = E(X_{n+4} - \hat{X}_{n+4})^2 = (1 + \psi_1^2 + \psi_2^2 + \psi_3^2)\sigma^2.$$