Handout 10

[Chapters 2.5.2 and 3.3 in Brockwell and Davis].

Some methods for modeling time series in the presence of trend and/or sesasonality.

We have employed the following methods for modeling a time series $\{Y_t\}$ when either trend or seasanality (or both are present):

a) series with trend: $Y_t = m_t + X_t$, where $\{m_t\}$ is the trend and $\{X_t\}$ is ARMA,

b) series with trend and seasonality: $Y_t = m_t + s_t + X_t$, where $\{m_t\}$ is the trend, $\{s_t\}$ are the seasonals and $\{X_t\}$ is ARMA.

In some cases, we may have to transform the series via a Box-Cox transformation first and then apply either (a) or (b) as necessary.

There is another class of methods which is quite popular. In order to see what they are, let us look at an example. The temperature series has a trend, but the plot of the first difference of the series $\{X_t = Y_t - Y_{t-1}\}$ against time looks stationary. So, instead of modeling $\{Y_t - m_t\}$ as stationary, we may model the first difference of the observed series by ARMA. This scheme is called ARIMA modeling (ARIMA stands for "autoregressive integrated moving average"):

c) series with trend: model the first or the second difference of the series as an ARMA sequence.

There is another class of models called the **seasonal ARIMA models** which are often used for series with both trend and seasonal components. We **will not** discuss seasonal ARIMA in this note.

ARIMA models

Recall the temperature series $\{Y_t\}$ (1850-2012) in Handouts 1, 4 and 6. We discussed the method of modeling it as $Y_t = m_t + X_t$, where m_t is the trend and X_t is stationary. There is another method that is often used to model this. Look at plot of the first difference $X_t = \nabla Y_t = Y_t - Y_{t-1}$ of the series and now the trend seems to be gone. As a matter of fact, the first difference X_t looks stationary. Moreover, the ACF plot for the differenced series indicates that only autocorrelations of lag 1 and 3 are present. PACF plot of the differenced series also indicates that only partial autocorrelations of lags 1 and 3 are present. We have fitted an $ARMA(1,1)$ model for the differenced series and ACF of the the residuals indicate that this model is reasonable. In this case the fitted model for $\{X_t = \nabla Y_t\}$ is $ARMA(1,1)$. We have used the astsa package and gave the command

> sarima(x,p=1,d=0,q=1).

This leads to the following output plus a graph given at the end of the handout:

Coefficient:			
	ar1	ma1	xmean
	0.4016	−0.7878	0.0050
se	0.1171	0.0754	0.0026

sigma^2 estimated as 0.00862: log likelihood = 154.99, aic = -301.97

$degrees_of_freedom

[1] 159

$ttable

	Estimate	se	t.value	p.value
ar1	0.4016	0.1171	3.4296	0.0008
ma1	−0.7878	0.0754	−10.4465	0.0000
xmean	0.0050	0.0026	1.8853	0.0612

$AIC

[1] -3.716608

$AICc

[1] -3.70269

$BIC

[1] -4.65943

We now want to predict Y_{n+1}, Y_{n+2} and Y_{n+3}. Note that $Y_n = 0.403$. In the astsa package we now give the command

> sarima.for(x,n.ahead=3,p=1,d=0,q=1)

and we get the out the following output plus a graph (not given here)

$$\hat{X}_{n+1} = 0.0121, \hat{X}_{n+2} = 0.0078, \hat{X}_{n+3} = 0.0061.$$

Hence the forecasted valyes of Y_{n+1}, Y_{n+2} and Y_{n+3} are

$$\begin{aligned}
\hat{Y}_{n+1} &= Y_n + \hat{X}_{n+1} = 0.403 + 0.0121 = 0.4151, \\
\hat{Y}_{n+2} &= \hat{Y}_{n+1} + \hat{X}_{n+2} = 0.4151 + 0.0078 = 0.4229, \\
\hat{Y}_{n+3} &= \hat{Y}_{n+2} + \hat{X}_{n+3} = 0.4229 + 0.0061 = 0.4291.
\end{aligned}$$

We can also get the forecasts by using the following R command in the astsa package:

> sarima.for(y,n.ahead=3,p=1,d=1,q=1).

ARIMA(p,d,q) Models

In some cases it may turn out that the first difference may not enough to get to stationarity. In such a case we may try and see if the second order difference is stationary, where the second order difference is $\nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$. In general one may try and check if the d^{th} order difference of the observed is stationary. In practice, $d = 1$ (the first order difference) is usually enough. In some cases, the second order difference may be needed. It is very rare to find examples where third order (or higher) differencing is needed to attain stationarity.

If the d^{th} order difference of the sequence $\{Y_t\}$ follows an $ARMA(p, q)$ model, the we say that the series $\{Y_t\}$ follows an $ARIMA(p, d, q)$ model. [ARIMA stands for "integrated autoregressive moving average"]. If we model the first difference of the temperature series by $ARMA(1, 1)$, we can say that the temperature series can be modeled by $ARIMA(1, 1, 1)$.

Notations:

(a) An $AR(p)$ model is often written as $ARIMA(p, 0, 0)$

(b) An $MA(q)$ model is often written as $ARIMA(0, 0, q)$.

(c) An $ARMA(p,q)$ is often written as $ARIMA(p,0,q)$.

(d) If the d^{th} order difference of a series follows an $AR(p)$ model, then we say that the series itself follows an $ARIMA(p,d,0)$ model.

(e) If the d^{th} order difference of a series follows an $MA(q)$ model, then we say that the series itself follows an $ARIMA(0,d,q)$ model.

Forecasting with MA models

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.

For an AR(1) model $X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$, $t = 1, \ldots, n$, we have already discussed the method for forecasting $X_{n+1}, X_{n+2}, \ldots$, when μ and ϕ are known (or are estimated). The forcasts are

$$\hat{X}_{n+1} = \mu + \phi(X_n - \mu),$$
$$\hat{X}_{n+2} = \mu + \phi(\hat{X}_{n+1} - \mu).$$

Note that the forcasted value of X_{n+1} depends only on the recent past X_n, and the forecasted value of X_{n+2} also depends on X_n (why?).

Similarly, for $AR(2)$, the forecasted value of X_{n+1} depends only on two recent pasts X_n and X_{n-1}. Similarly, the forecasted value of X_{n+2} depends only on X_n and X_{n-1}.

However, for an MA(1) model, $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}, t = 1, \ldots, n$, the forecasted value of X_{n+1} depends on the whole past $X_1, \ldots, X_n$ plus a term involving ε_0. As a matter of fact, any MA(1) model can be written as a long AR model plus a term involving ε_0. If $-1 < \theta < 1$, then the term involving ε_0 is negligible. The same is true for MA(2) models. An MA(2) model can be written as a long AR model plus two terms involving ε_{-1} and ε_0. Under some conditions on θ_1, θ_2 (to be described in Handout 11), the terms involving ε_{-1} and ε_0. are negligible if t is not small. This phenomenon is true for any MA(q) model or any ARMA(p,q) under appropriate conditions on $\theta_1, \ldots, \theta_q$. We know how to do forecast for an AR model. So in principle once an ARMA model is converted into an approximate AR model, we can carry out the forecasts.

Converting an MA(1) model into an approximate AR model

Consider an MA(1) model with $\theta = 0.5$. Then we can rewrite the model as an AR model in which the autoregressive coefficients do not vanish

$$X_t - \mu = -(-\theta)(X_{t-1} - \mu) - (-\theta)^2(X_{t-2} - \mu) - (-\theta)^3(X_{t-3} - \mu)$$
$$- \cdots - (-\theta)^{t-1}(X_1 - \mu) - (-\theta)^t\varepsilon_0 + \varepsilon_t.$$

If $-1 < \theta < 1$, then the dependence of X_t on ε_0 and the initial observations X_1, X_2 etc. are negligible if t is not small. For instance if $t = 20$, then $(-0.5)^{20} = 9.5 \times 10^{-7}$. However if $|\theta| \geq 1$, then the dependence will persist. The method of forecasting described here works only if $-1 < \theta < 1$. **As long as $-1 < \theta < 1$, we can regard MA(1) models as being approximately a long AR model whose autoregressive coefficients can be written in terms of θ**

If we have observed $X_t, t = 1, \ldots, n$, $\theta = 0.5$, then taking $t = n + 1$ we have

$$X_{n+1} - \mu = -(-0.5)(X_n - \mu) - (-0.5)^2(X_{n-1} - \mu) - (-0.5)^3(X_{n-2} - \mu) \\ - \cdots - (-0.5)^n(X_1 - \mu) - (-0.5)^{n+1}\varepsilon_0 + \varepsilon_{n+1},$$

If we could somehow know ε_0, then the forecasted value of X_{n+1} would be

$$\hat{X}_{n+1} - \mu \approx -(-0.5)(X_n - \mu) - (-0.5)^2(X_{n-1} - \mu) - (-0.5)^3(X_{n-2} - \mu) \\ - \cdots - (-0.5)^n(X_1 - \mu) - (-0.5)^{n+1}\varepsilon_0.$$

So if we know ε_0 or guess it, we can forecast X_{n+1}. If $n = 20$, $(-0.5)^{20} = 9.5 \times 10^{-7}$ and $(-0.5)^{n+1} = 4.8 \times 10^{-7}$. **If n is not small, the guess of ε_0 has neligible effect on the value of $\hat{X}_{n+1}$. For this reason, often one takes $\varepsilon_0 = 0$.**
Also note that even though $\hat{X}_{n+1}$ depends on $X_n, \ldots X_1$, but the dependence decreases rapidly as we go back farther in time. As a matter of fact, dependence of $\hat{X}_{n+1}$ on observations at the beginning of the series such as X_1, X_2 are almost negligible.

Foreasting an MA(q) series

Under appropriate conditions on $\theta_1, \ldots, \theta_q$, mathematical results tell us we can rewrite an MA(q) series as a long AR series plus terms involving q initial ε's, whose effects of X_t are neligible when t is not small.. Thus if we could know the expressions of the autoregressive coefficients in terms of $\theta_1, \ldots, \theta_q$, and the observed series is $\{X_1, \ldots, X_n\}$, we can make forecasts of X_{n+1}, X_{n+2} etc. with guesses of initial $\varepsilon's$ (often taken to be zeros). However, except for the MA(1) case, autoregressive coefficients have no simple expressions in terms of $\theta_1, \ldots, \theta_q$. Fortunately, there is a simple way to carry out the forecasts without having to explicity know the coefficients of the approximate long AR model. The method is described below.

A simple iterative method for forecasting MA(1) series

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 predicted value on the basis of the past $\{X_1, ..., X_{t-1}\}$ is $\hat{X}_t = \mu + \theta\varepsilon_{t-1}$ and the 'innovation' is $X_t - \hat{X}_t = \varepsilon_t$. [For any time series (not necessarily $MA(q)$ or $AR(p)$ or $ARMA(p, q)$), the innovation is defined to be equal to $X_t - \hat{X}_t$, i.e., the observed minus the predicted. In the last handout we denoted the theoretical forecasted value of X_t as $X_t^{(f)}$. According to that notation, $X_t^{(f)}$ is equal to $\mu + \theta\varepsilon_{t-1}$ in the $MA(1)$ case. Ideally, we should write $\hat{X}_t$ as an estimate of $X_t^{(f)}$, i.e, $\hat{X}_t$ should be equal to $\mu + \theta\hat{\varepsilon}_{t-1}$, where $\hat{\varepsilon}_{t-1}$ is an estimate of ε_{t-1}. However, in order to keep the notations simple we will not worry about this subtle distinction in this handout.]

For simplicity, let us assume that $\mu = 0, \theta = 0.5$ and we have a series of length $n = 5$. The values of the observed series $\{X_1, ..., X_5\}$ are $\{1.5, 2.1, -1.9, -2.2, 0.4\}$. We are interested in predicting $X_{n+1} = X_6$. Note that if 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? Since $\varepsilon_5 = X_5 - 0.5\varepsilon_4$ and we know the value of X_5 (which equals 0.4), we would know ε_5 if we are able to obtain the value of ε_4. The same argument will show that we can know ε_4 if we are able to obtain the value of ε_3. A repetition of this argument will show that we can know $\varepsilon_1 = X_1 - 0.5\varepsilon_0$ if we are able to guess 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 $\hat{\varepsilon}_0$ of 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 our estimates of $\varepsilon_1, .., \varepsilon_5$ are given below (note $\hat{X}_t = \theta\varepsilon_{t-1}$ and $\hat{\varepsilon}_t = X_t - \theta\varepsilon_{t-1} = X_t - \hat{X}_t$). Note that each step consists of two substeps: obtain $\hat{X}_t$ and $\hat{\varepsilon}_t$.

$$
\begin{aligned}
(1a)\ \hat{X}_1 &= \theta\hat{\varepsilon}_0 = 0,\\
(1b)\ \hat{\varepsilon}_1 &= X_1 - \hat{X}_1 = 1.5 - 0 = 1.5,\\
(2a)\ \hat{X}_2 &= \theta\hat{\varepsilon}_1 = (0.5)(1.5) = 0.75,\\
(2b)\ \hat{\varepsilon}_2 &= X_2 - \hat{X}_2 = 2.1 - 0.75 = 1.35,\\
(3a)\ \hat{X}_3 &= \theta\hat{\varepsilon}_2 = (0.5)(1.35) = 0.675,\\
(3b)\ \hat{\varepsilon}_3 &= X_3 - \hat{X}_3 = -1.9 - 0.675 = -2.575,\\
(4a)\ \hat{X}_4 &= \theta\hat{\varepsilon}_3 = (0.5)(-2.575) = -1.2875,\\
(4b)\ \hat{\varepsilon}_4 &= X_4 - \hat{X}_4 = -2.2 - (-1.2875) = -0.9125,\\
(5a)\ \hat{X}_5 &= \theta\hat{\varepsilon}_4 = (0.5)(-0.9125) = -0.4563,\\
(5b)\ \hat{\varepsilon}_5 &= X_5 - \hat{X}_5 = 0.4 - (-0.4563) = 0.8563.
\end{aligned}
$$

So our predicted value of X_6 is

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

Note that the predicted value of X_7 is $\theta\varepsilon_6$. Now, estimate of ε_6 is $X_6 - \hat{X}_6$. Since we do not have X_6, we have to substitute it by $\hat{X}_6$. So $\hat{\varepsilon}_6 = \hat{X}_6 - \hat{X}_6 = 0$. So the predicted value of X_7 is $\theta\hat{\varepsilon}_6 = 0$. **As a matter of fact, $\hat{X}_7, \hat{X}_8, ...$ are all zeros.**

Forecasting for MA(1) for general n

If $\mu \neq 0$, then the estimate of ε_t at each stage will be $\hat{\varepsilon}_t = X_t - \mu - \theta\hat{\varepsilon}_{t-1}, -1 < \theta < 1$. The argument is basically the same as before. The method requires an initial guess $\hat{\varepsilon}_0$ of ε_0 (usuall $\hat{\varepsilon}_0 = 0$). Then iteratively calculate

$$
\begin{aligned}
&\hat{X}_1 = \mu + \theta\hat{\varepsilon}_0,\ \hat{\varepsilon}_1 = X_1 - \hat{X}_1,\\
&\hat{X}_2 = \mu + \theta\hat{\varepsilon}_1,\ \hat{\varepsilon}_2 = X_2 - \hat{X}_2,\\
&\quad\vdots,\\
&\hat{X}_n = \mu + \theta\hat{\varepsilon}_{n-1},\ \hat{\varepsilon}_n = X_n - \hat{X}_n.
\end{aligned}
$$

So the forecasted values of X_{n+1} is $\hat{X}_{n+1} = \mu + \theta\hat{\varepsilon}_n$. Since X_{n+1} is not available, $\hat{\varepsilon}_{n+1} = \hat{X}_{n+1} - \hat{X}_{n+1} = 0$, and hence $\hat{X}_{n+2} = \mu$. The same argument shows $\hat{X}_{n+2} = \hat{X}_{n+3} = \cdots = \mu$. **Recall that when n is not small and $-1 < \theta < 1$, the choice of ε_0 has negligible effect on $\hat{X}_{n+1}$.**

A simple iterative method for forecasting MA(2) series

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.

For simplicity, let us assume that $\mu = 0, \theta_1 = 0.5, \theta_2 = 0.4$ and we have a series of length $n = 5$. The values of the observed series $\{X_1, ..., X_5\}$ are $\{1.5, 2.1, -1.9, -2.2, 0.4\}$. We want to forecast $X_{n+1}, X_{n+2}, \ldots$.

Suppose we make a guess $\hat{\varepsilon}_{-1}$ and $\hat{\varepsilon}_0$ of ε_{-1} and ε_0 (the usual guesses are $\hat{\varepsilon}_{-1} = 0$, $\hat{\varepsilon}_0 = 0$, though other complicated methods are also possible). Let us take $\hat{\varepsilon}_{-1} = 0$ and $\hat{\varepsilon}_0 = 0$ here. Then

$$
\begin{aligned}
(1a)\ \hat{X}_1 &= \theta_1\hat{\varepsilon}_0 + \theta_2\hat{\varepsilon}_{-1} = 0,\\
(1b)\ \hat{\varepsilon}_1 &= X_1 - \hat{X}_1 = 1.5 - 0 = 1.5,\\
(2a)\ \hat{X}_2 &= \theta_1\hat{\varepsilon}_1 + \theta_2\hat{\varepsilon}_0 = (0.5)(1.5) + (0.4)(0) = 0.75,\\
(2b)\ \hat{\varepsilon}_2 &= X_2 - \hat{X}_2 = 2.1 - 0.75 = 1.35,\\
(3a)\ \hat{X}_3 &= \theta_1\hat{\varepsilon}_2 + \theta_2\hat{\varepsilon}_1 = (0.5)(1.35) + (0.4)(1.5) = 0.675 + 0.6 = 1.275,\\
(3b)\ \hat{\varepsilon}_3 &= X_3 - \hat{X}_3 = -1.9 - 1.275 = -3.175,\\
(4a)\ \hat{X}_4 &= \theta_1\hat{\varepsilon}_3 + \theta_2\hat{\varepsilon}_2 = (0.5)(-3.175) + (0.4)(1.35) = -1.0475,\\
(4b)\ \hat{\varepsilon}_4 &= X_4 - \hat{X}_4 = -2.2 - (-1.0475) = -1.1525,\\
(5a)\ \hat{X}_5 &= \theta_1\hat{\varepsilon}_4 + \theta_2\hat{\varepsilon}_3 = (0.5)(-1.1525) + (0.4)(-3.175) = -1.8463,\\
(5b)\ \hat{\varepsilon}_5 &= X_5 - \hat{X}_5 = 0.4 - (-1.8463) = 2.2463.
\end{aligned}
$$

Now the forecasts are

$$
\begin{aligned}
(6a)\ \hat{X}_6 &= \theta_1\hat{\varepsilon}_5 + \theta_2\hat{\varepsilon}_4 = (0.5)(2.2463) + (0.4)(-1.1525) = 0.6622,\\
(6b)\ \hat{\varepsilon}_6 &= \hat{X}_6 - \hat{X}_6 = 0,\\
(7a)\ \hat{X}_7 &= \theta_1\hat{\varepsilon}_6 + \theta_2\hat{\varepsilon}_5 = (0.5)(0) + (0.4)(2.2463) = 0.8986,\\
(7b)\ \hat{\varepsilon}_7 &= \hat{X}_7 - \hat{X}_7 = 0,\\
(8a)\ \hat{X}_8 &= \theta_1\hat{\varepsilon}_7 + \theta_2\hat{\varepsilon}_6 = 0.
\end{aligned}
$$

Note that the expression for $\hat{\varepsilon}_6$ is $X_6 - \hat{X}_6$. However, since X_6 is unknown, we substitute X_6 by $\hat{X}_6$ and thus $\hat{\varepsilon}_6 = 0$. The same is also true for $\hat{\varepsilon}_7, \hat{\varepsilon}_8$ etc. Also note that predicted values of $X_8, X_9, \ldots$ are zero. In general for an MA(2) series with $\mu = 0$, $\hat{X}_{n+3} = \hat{X}_{n+4} = \cdots = 0$.

If the MA(2) series has mean $\mu \neq 0$, then the method starts with initial guesses $\hat{\varepsilon}_{-1}$ and $\hat{\varepsilon}_0$ of ε_{-1} and $\hat{\varepsilon}_0$ (often one takes $\hat{\varepsilon}_{-1} = \hat{\varepsilon}_0 = 0$). Then the iterations are

$$
\begin{aligned}
&\hat{X}_1 = \mu + \theta_1\hat{\varepsilon}_0 + \theta_2\hat{\varepsilon}_{-1},\ \hat{\varepsilon}_1 = X_1 - \hat{X}_1,\\
&\hat{X}_2 = \mu + \theta_1\hat{\varepsilon}_1 + \theta_2\hat{\varepsilon}_0,\ \hat{\varepsilon}_2 = X_2 - \hat{X}_2,\\
&\quad\vdots,\\
&\hat{X}_n = \mu + \theta_1\hat{\varepsilon}_{n-1} + \theta_2\hat{\varepsilon}_{n-2},\ \hat{\varepsilon}_n = X_n - \hat{X}_n.
\end{aligned}
$$

The forecated values of $X_{n+1}, X_{n+2}, \ldots$ are

$$\hat{X}_{n+1} = \mu + \theta_1\hat{\varepsilon}_n + \theta_2\hat{\varepsilon}_{n-1},\ \hat{\varepsilon}_{n+1} = 0,$$
$$\hat{X}_{n+2} = \mu + \theta_1\hat{\varepsilon}_{n+1} + \theta_2\hat{\varepsilon}_n = \mu + \theta_2\hat{\varepsilon}_n,\ \hat{\varepsilon}_{n+2} = 0,$$
$$\hat{X}_{n+3} = \mu + \theta_1\hat{\varepsilon}_{n+2} + \theta_2\hat{\varepsilon}_{n+1} = \mu,\ \hat{\varepsilon}_{n+3} = 0.$$

Note that $\mu = \hat{X}_{n+3} = \hat{X}_{n+4} = \cdots$.

A simple iterative method for forecasting ARMA(1,1) series

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

Consider an ARMA model $X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}$. In order for $\{X_t\}$ to be stationary, we must have $-1 < \phi < 1$ and if we want to write this series as a long AR model. We should also have $-1 < \theta < 1$, so that the guess of an initial ε has negligible effect on the forecasts. We want to show how the iterative method for forecasting works here. The forecasted values of X_{n+1} depends on $X_n, X_{n-1}, \ldots, X_1$ plus a guess of an initial ε, but the basic ingredient is that the dependence on past observations decrease rapidly as we go back farther in time.

As before we take $n = 5, \mu = 0, \theta = 0.5$ and we now assume in addition that $\phi = 0.8$. The data is the same as before: the values of the observed series $\{X_1, ..., X_5\}$ are $\{1.5, 2.1, -1.9, -2.2, 0.4\}$.We now want to forecast $X_6, X_7, \ldots$. Note that the series now is

$$X_t = 0.8X_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t.$$

So if we have a guess $\hat{\varepsilon}_{t-1}$ of ε_{t-1}, then the predicted value of X_t would be

$$\hat{X}_t = 0.8X_t + 0.5\hat{\varepsilon}_{t-1}.$$

Here we start with an initial guess $\hat{\varepsilon}_1 = 0$ of ε_1 instead of making a guess of ε_0 as in the MA(1) case. Then

$$\hat{X}_2 = 0.8X_1 + 0.5\hat{\varepsilon}_1 = (0.8)(1.5) + (0.5)(0) = 1.2$$
$$\hat{\varepsilon}_2 = X_2 - \hat{X}_2 = 2.1 - 1.2 = 0.9,$$
$$\hat{X}_3 = 0.8X_2 + 0.5\hat{\varepsilon}_2 = (0.8)(1.2) + (0.5)(0.9) = 1.41,$$
$$\hat{\varepsilon}_3 = X_3 - \hat{X}_3 = -1.9 - 1.41 = -3.31,$$
$$\hat{X}_4 = 0.8X_3 + 0.5\hat{\varepsilon}_3 = (0,8)(-1.9) + (0.5)(-3.31) = -3.175,$$
$$\hat{\varepsilon}_4 = X_4 - \hat{X}_4 = -2.2 - (-3.175) = 0.975,$$
$$\hat{X}_5 = 0.8X_4 + 0.5\hat{\varepsilon}_4 = (0.8)(-2.2) + (0.5)(0.975) = -1.2725,$$
$$\hat{\varepsilon}_5 = X_5 - \hat{X}_5 = 0.4 - (-1.2725) = 1.6725.$$

So the forecasts are

$$\begin{aligned}
\hat{X}_6 &= 0.8X_5 + 0.5\hat{\varepsilon}_5 = (0.8)(0.4) + (0.5)(1.6725) = 1.1563,\\
\hat{\varepsilon}_6 &= \hat{X}_6 - \hat{X}_6 = 0,\\
\hat{X}_7 &= 0.8\hat{X}_6 + 0.5\hat{\varepsilon}_6 = (0.8)(1.1563) + (0.5)(0) = 0.9250,\\
\hat{\varepsilon}_7 &= \hat{X}_7 - \hat{X}_7 = 0,\\
\hat{X}_8 &= 0.8\hat{X}_7 + 0.5\hat{\varepsilon}_7 = (0.8)(0.9250) + (0.5)(0) = 0.7400.
\end{aligned}$$

Note that in the forecast of X_6, the MA part has a contribution. **But there is no contribution of the MA part in the forecasts of $X_7, X_8, \ldots$., only the AR part contributes.**

Forecasting for ARMA(1,1): general n.

Now look at the case the general case instead of $n = 5$ for ARMA(1,1). Let us assume that we have an ARMA(1,1) model $X_t - \mu = 0.8(X_{t-1} - \mu) + \varepsilon_t + 0.5\varepsilon_{t-1}$ and we want to predict X_{n+1} when the observed data is $X_1, ..., X_n$. Note the best predictor for X_{n+1} is $\mu + 0.8(X_n - \mu) + 0.5\varepsilon_n$. 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

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

So the predicted value of X_{n+1} is

$$\hat{X}_{n+1} = \mu + 0.8(X_n - \mu) + 0.5\hat{\varepsilon}_n.$$

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

$$\hat{X}_{n+2} = \mu + 0.8(\hat{X}_{n+1} - \mu) + 0.5\hat{\varepsilon}_{n+1} = \mu + 0.8(\hat{X}_{n+1} - \mu).$$

.Similarly, $\hat{X}_{n+3} = \mu + 0.8(\hat{X}_{n+2} - \mu)$ and so on. **Note that only the AR part contributes to the forecasted values of $X_{n+2}, X_{n+3}, \ldots$.**

Remark: 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 it is considerably more complicated. For instance, if $\theta = -1$, then the best predictor of X_{n+1} is of the form

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

Non-uniqueness of moving average models

Moving average models are not unique. For instance, consider an $MA(1)$ model $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$, where $\{\varepsilon_t\}$ white noise with variance σ^2. Under the assumption of normality (i.e. $\{\varepsilon_t\}$ are normally distributed), any stationary series is completely characterized by the mean and autocovariances. Thus if if two sequences have the same mean and autocovariance functions, they are equally good descriptions of the data, i.e., they provide the same fit and they have the same predictive performances. Consider the following two models

$$X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1} \tag{1}$$
$$X_t - \mu = \varepsilon'_t + (1/\theta)\varepsilon'_{t-1}, \tag{2}$$

where $\theta \neq 0$, $\{\varepsilon_t\}$ white noise with variance σ^2, and $\{\varepsilon'_t\}$ is white noise with variance $\theta^2\sigma^2$. Note that we only observe the data $\{X_t\}$, not ε_t or ε'_t. Both models have the same mean μ. All the autocovariances of lag 2 or higher are zero for both models. Note that for model 1, we have

$$\gamma(0) = (1+\theta^2)\sigma^2, \gamma(1) = \theta\sigma^2, 0 = \gamma(2) = \gamma(3) = \cdots.$$

For model 2,

$$\gamma(0) = (1 + (1/\theta)^2)(\theta^2\sigma^2) = (1+\theta^2)\sigma^2,$$
$$\gamma(1) = (1/\theta)(\theta^2\sigma^2) = \theta\sigma^2, 0 = \gamma(2) = \gamma(3) = \cdots.$$

So both the models have identical mean and autocovariance structures. Hence they will provide identical fits and predictions. This non-uniqueness is called lack of "identifiability". This can pose a big problem for numerical calculations. Computer packages will often fail to provide estimates of the parameters θ and σ^2 if there is more than one "correct" solution.

What do we do about it? To see this let us examine it more closely. If the value of θ is larger than 1 in magnitude and $Var(\varepsilon_t) = \sigma^2$, then we may as well consider a model (2) for which coefficient associated with ε_{t-1} is $1/\theta$ whose magnitude is less than 1 and $Var(\varepsilon'_t) = \theta^2\sigma^2$. So if $\theta = 2$, then we should work with model (2) in which the coefficient associated with ε'_{t-1} is $1/\theta = 1/2$ and $Var(\varepsilon'_t) = \theta^2\sigma^2 = 4\sigma^2$. Thus we can always choose a model where the coefficient associated with ε_{t-1} is no larger than 1 in magnitude and this is what is done in practice.

How about general MA(q) models? The same issue on non-identifiability comes up. One can restrict attention to those models with appropriate conditions on the MA parameters $\theta_1, ..., \theta_q$ needed for identifiably and this is what is usually done. We will address this issue soon.

Summary of technical issues for $AR(p)$, $MA(q), ARMA(p,q)$models

1. For $AR(p)$ models, we need conditions on the autoregressive coefficients $\phi_1, ..., \phi_p$ in order to guarantee stationarity.
2. For $MA(q)$ models, we need conditions on the moving average coefficients $\theta_1, ..., \theta_q$ in order to guarantee "identifiability".
3. For $ARMA(p, q)$ models, the coefficients in the AR part must satisfy constrains to guarantee stationarity (as in (1) above) and the MA coefficients must satisfy constraints in order to guarantee identifiability of the model (as in (2) above). In addition, we need conditions on the AR and the MA coefficients in order to guarantee "non-redundancy". We will discuss this issue in detail soon. But let us first see what is meant by the problem of redundancy in the ARMA(1,1) case.

Redundancy issue for ARMA(1,1) model.

Suppose that the series $\{X_t\}$ can be described as $X_t = \varepsilon_t$, where $\{\varepsilon_t\}$ is white noise. So $\{X_t\}$ is a white noise. Subtract $0.5X_{t-1} = 0.5\varepsilon_{t-1}$ from this series to get

$$X_t - 0.5X_{t-1} = \varepsilon_t - 0.5\varepsilon_{t-1}, i.e., X_t = 0.5X_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}.$$

Now it seems that the series $\{X_t\}$ is $ARMA(1, 1)$, whereas in reality it is white noise. As a matter of fact we can rewrite X_t as

$$X_t = \phi X_{t-1} + \varepsilon_t - \phi\varepsilon_{t-1},$$

for any $-1 < \phi < 1$. Once again it looks as if $\{X_t\}$ is ARMA(1,1), however there is a redundant parameter ϕ. Also note that the number of such redundant models is infinite. In general if $\theta = -\phi$ in an $ARMA(1, 1)$ model $X_t = \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, then there is redundancy, since $\{X_t\}$ can be described as white noise. So it is important to fit an ARMA model with constraints on the parameters in order to avoid any redundancy.

Some technical results for MA(q)

For any $MA(q)$ model, all autocorrelations of lag $q + 1$ and higher are zero. Recall that X_t is $MA(q)$ then it has the form

$$X_t - \mu = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_{t-q}\varepsilon_{t-q},$$

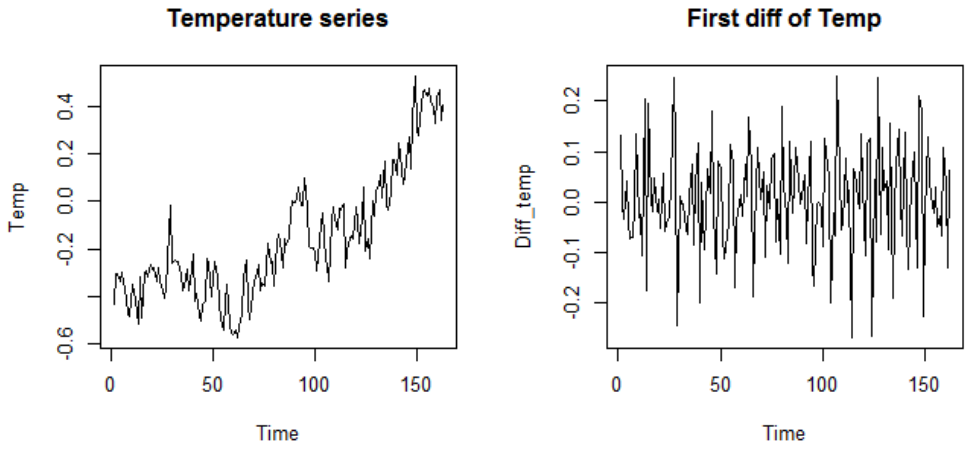
where $\{\varepsilon_t\}$ is white noise with variance σ^2. We will now write down the the autocovariance and autocorrelations for the sequence $\{X_t\}$. For notational convenience we will rewrite the MA(q) model as

$$X_t - \mu = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_{t-q}\varepsilon_{t-q}, \text{ with } \theta_0 = 1.$$

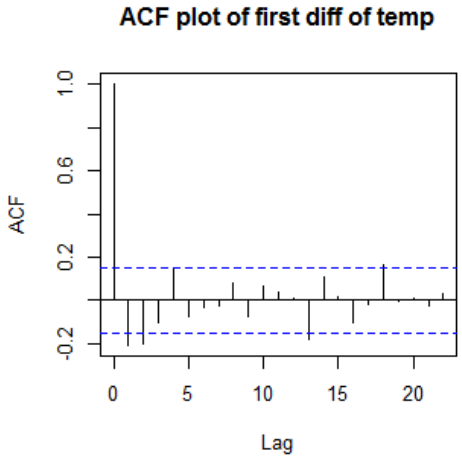
Then

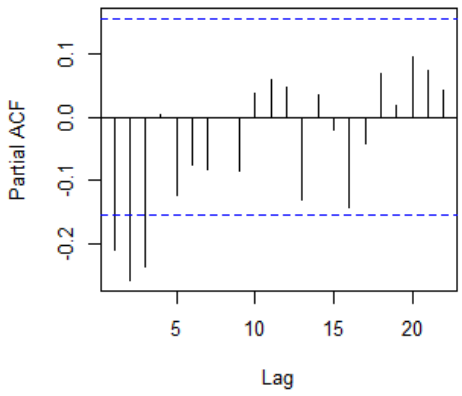
$$\gamma(h) = \begin{cases} \sigma^2 \sum_{i=0}^{q-h} \theta_i\theta_{i+h} & h = 0, ..., q \\ 0 & h = q + 1, q + 2, ... \end{cases},$$

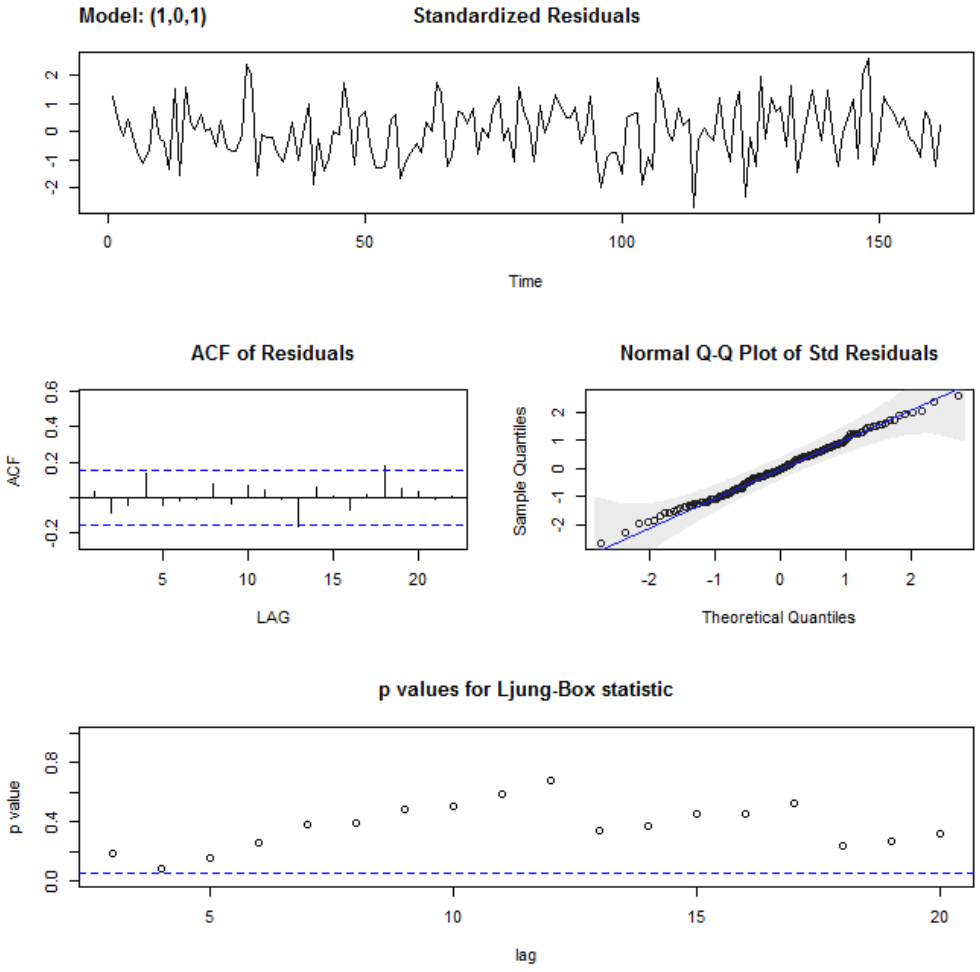
$$\rho(h) = \begin{cases} [\sum_{i=0}^{q-h} \theta_i\theta_{i+h}]/[\sum_{i=0}^{q} \theta_i^2] & h = 0, ..., q \\ 0 & h = q + 1, q + 2, ... \end{cases}$$


Temperature series
Temp
Time
-0.6
-0.2
0.0
0.2
0.4
0
50
100
150

First diff of Temp
Diff_temp
Time
-0.2
-0.1
0.0
0.1
0.2
0
50
100
150


ACF plot of first diff of temp
ACF
Lag
-0.2
0.2
0.6
1.0
0
5
10
15
20


PACF plot of first diff of temp
Partial ACF
Lag
-0.2
-0.1
0.0
0.1
5
10
15
20


Model: (1,0,1)
Standardized Residuals
Time
ACF of Residuals
ACF
LAG
Normal Q-Q Plot of Std Residuals
Sample Quantiles
Theoretical Quantiles
p values for Ljung-Box statistic
p value
lag