11 ARMA Models

In this section we study an important set of time series models; the autoregressive moving average (ARMA) models.

11.1 Reasons for Studying Models

In addition to being a simple description of a data set, models are useful for many reasons, including the following which we illustrate with an $MA(1, \beta, \sigma^2)$ model

$$X(t) = \epsilon(t) + \beta \epsilon(t), \quad \epsilon \sim WN(\sigma^2).$$

1. A model can be thought of as a rule for simulating data sets like the one we have if we can determine what model generated our data. For example, the $MA(1)$ model says that realizations are generated by first generating a white noise realization and then doing the simple linear combination $\epsilon(t) + \beta \epsilon(t - 1)$ to get a realization.

2. A model makes it easy to find the spectral density function of the process. For example, the filter theorem lets us immediately say for an $MA(1)$ model that

$$f(\omega) = \sigma^2 \left| 1 + \beta e^{2\pi i \omega} \right|^2.$$

3. A model makes it possible to do prediction in the sense that we can find the autocovariances $R(v)$ that are needed to solve the
prediction normal equations $\Gamma_n \lambda_{nh} = r_{nh}$. For the $MA(1)$ we have

$$R(v) = \begin{cases} 
\sigma^2 (1 + \beta^2), & v = 0 \\
\sigma^2 \beta, & v = \pm 1 \\
0, & |v| > 1.
\end{cases}$$

### 11.2 The Moving Average Model of Order $q$

A time series $X$ is said to be a moving average process of order $q$ with coefficients $\beta$ and noise variance $\sigma^2$ if

$$X(t) = \epsilon(t) + \beta_1 \epsilon(t - 1) + \cdots + \beta_q \epsilon(t - q),$$

where $\epsilon \sim WN(\sigma^2)$. Thus to simulate an $MA(q)$ realization (via the MADT command) we must know $q$, the $\beta$'s, and $\sigma^2$ and then we can simulate a $WN(\sigma^2)$ realization and use the $MA$ formula to get the realization from $X$. Thus $q$, $\beta$, and $\sigma^2$ are the parameters of the model and we write $X \sim MA(q, \beta, \sigma^2)$.

We see that $X$ is a filtered version of white noise and thus we know it is covariance stationary with

$$f(\omega) = \sigma^2 \left| 1 + \beta_1 e^{2\pi i \omega} + \cdots + \beta_q e^{2\pi i q \omega} \right|^2,$$

and

$$R(v) = \begin{cases} 
\sum_{k=0}^{q-|v|} |v| \beta_k \beta_{k+|v|}, & |v| \leq q, \\
0, & |v| > q.
\end{cases}$$
Qualitative Features of MA Processes

There are two facts about MA processes that we will use repeatedly when trying to find a model for a data set:

1. **R is truncated at \( v = q \):** The formula above shows that \( R(v) \), and thus \( \rho(v) = R(v)/R(0) \) is zero past lag \( q \).

2. **f can't have sharp peaks unless \( q \) is large:** This is easy to see if we write

\[
f(\omega) = \sum_{v=-\infty}^{\infty} R(v) e^{-2\pi i v \omega}
\]

\[
= R(0) + 2 \sum_{v=1}^{\infty} R(v) \cos(2\pi v \omega)
\]

\[
= R(0) + 2 \sum_{v=1}^{q} R(v) \cos(2\pi v \omega)
\]

which means that unless \( q \) is large, \( f \) can't have sharp peaks.

As an example, the figure below has the results of the following commands for an \( MA(2, \beta = (.3, .9), \sigma^2 = 1) \) process:
beta=<.3,.9>

; get data, true rho and f:

x=madt(beta,2,1,0,100)
rho=macorr(beta,2,1,30,r0)
f=masp(beta,2,1,100)

; get rho hats and periodogram:

rhohat=corr(x,n,30,0,1,r0hat,per)
rhohat=corr(x,n,0,100,1,r0hat,per)

; plot results:

plot(x)
plotsp(f,100,r0,per100,r0hat)
plot2(rho,rhohat,30,30,1,2)
Inverting an MA Model

An $MA$ process is a filtered version of present and past values of white noise. The question arises as to whether we can invert the model and write $\epsilon(t)$ as a filtered version of present and past $X$’s.

For example, if we have an $MA(1, \beta = -0.5)$ model

$$X(t) = \epsilon(t) - 0.5\epsilon(t - 1)$$
we have by successive substitution

\[
\epsilon(t) = X(t) + 0.5\epsilon(t-1) \\
= X(t) + 0.5 [X(t-1) + 0.5\epsilon(t-2)] \\
= \ldots \\
= \sum_{j=0}^{t-1} (0.5)^j X(t-j) + (0.5)^t \epsilon(0).
\]

As \( t \) gets large, this last term goes to zero since \(|\beta| < 1\), and we write

\[
\sum_{j=0}^{\infty} (0.5)^j X(t-j) = \epsilon(t),
\]

and we have thus in fact inverted the \( MA(1) \) model into a different model (which we will see is called an infinite order autoregressive model). Notice that this only worked since \(|\beta| < 1\).

There is an easier way to look at this.

**Def:** Define the lag or backshift operator \( L \) to be such that \( L^j X(t) = X(t-j) \), that is \( L^j \) applied to a time series shifts the series back in time by \( j \) units (if \( j \) is negative, it shifts the series forward in time).

Thus we can write the \( MA(q) \) model as (defining \( \beta_0 = 1 \))

\[
X(t) = \sum_{k=0}^{q} \beta_k \epsilon(t-k) = \sum_{k=0}^{q} \beta_k L^k \epsilon(t) = h(L)\epsilon(t),
\]
where

\[ h(L) = \sum_{k=0}^{q} \beta_k L^k. \]

For our \( MA(1, \beta = -0.5) \) example, we have

\[ X(t) = (1 - 0.5L)\epsilon(t), \]

and so we would like to write

\[
\begin{align*}
\epsilon(t) &= \frac{X(t)}{1 - 0.5L} \\
&= \left( \sum_{j=0}^{\infty} (0.5)^j L^j \right) X(t) \\
&= \sum_{j=0}^{\infty} (0.5)^j X(t - j)
\end{align*}
\]

in analogy with the geometric series \( 1/(1 - z) = \sum_{j=0}^{\infty} z^j \), and in fact we can since \(|\beta| < 1\).

In general, we can invert an \( MA(q) \) process if and only if the zeros of \( h(z) \) are all greater than one in modulus, in which case we can write

\[ \epsilon(t) = \sum_{j=0}^{\infty} \gamma_j X(t - j), \]

where \( \gamma_j \) is the coefficient of \( z^j \) in \( 1/h(z) \).
Ex: For an $MA(1)$, $h(z) = 1 + \beta z$, the zero of which is $z = -1/\beta$ and $| - 1/\beta | > 1$ if and only if $|\beta| < 1$ as required.

Zeros and the unit circle: A complex number $z = a + bi$ has modulus greater than one if and only if the point $(a, b)$ plotted in a (real part, imaginary part) axis falls outside a circle of radius one centered at the origin, that is, the unit circle. This is because $|z| = \sqrt{a^2 + b^2}$. Thus the zeros of $h$ are all greater than one in modulus if and only if they all fall outside the unit circle.

Finding the coefficients of the reciprocal polynomial: We find the coefficients of the reciprocal of $h(z) = 1 + .3z + .9z^2$ for the $MA(2, \beta = (.3, .9))$ model by long division:

\[
\begin{align*}
1 - .3z - .81z^2 + .513z^3 & \text{ etc.} \\
\hline
1+.3z+.9z^2 & | 1 \\
1 + .3z + .9z^2 & \\
\hline
- .3z - .9z^2 & \\
- .3z - .09z^2 - .27z^3 & \\
\hline
- .81z^2 + .27 z^3 & \\
- .81z^2 - .243z^3 - .729z^4 & \\
\hline
+ .513z^3 + .729z^4 & 
\end{align*}
\]

which can be verified by

\[
\begin{align*}
? \ beta=\langle .3, .9 \rangle \\
? \ gamma=\text{invpoly}(beta, 2, 10)
\end{align*}
\]
? list(gamma,10,4,4f7.3)

reciprocal of polynomial

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.300</td>
<td>-0.810</td>
<td>0.513</td>
<td>0.575</td>
</tr>
<tr>
<td>5</td>
<td>-0.634</td>
<td>-0.327</td>
<td>0.669</td>
<td>0.094</td>
</tr>
<tr>
<td>9</td>
<td>-0.630</td>
<td></td>
<td>0.105</td>
<td></td>
</tr>
</tbody>
</table>

(Note that invpoly uses something more sophisticated than long division.)

11.3 The Autoregressive Model of Order \( p \)

A time series is said to follow an autoregressive model of order \( p \), coefficients \( \alpha \), and noise variance \( \sigma^2 \) if it satisfies the \( p \)th degree difference equation

\[
X(t) + \alpha_1 X(t - 1) + \cdots + \alpha_p X(t - j) = \epsilon(t),
\]

where \( \epsilon \sim WN(\sigma^2) \). Note that using the lag operator \( L \) and defining \( \alpha_0 = 1 \), we can write

\[
g(L)X(t) = \epsilon(t),
\]

where \( g(z) = \sum_{j=0}^{p} \alpha_j z^j \) (note that we have used \( h \) for the MA polynomial and \( g \) for the autoregressive polynomial).

The parameters of the model are \( p, \alpha, \) and \( \sigma^2 \) and we write \( X \sim AR(p, \alpha, \sigma^2) \).
Many people find the AR model to be very natural as we can write

\[ X(t) = -\alpha_1 X(t - 1) - \cdots - \alpha_p X(t - j) + \epsilon(t), \]

that is, the value at time \( t \) is a linear combination of the values at the previous \( p \) times plus a random error. In fact this is the origin of the name auto-regression as it looks like a regression of \( X(t) \) on lagged values of itself.

**What Coefficients Are Legal?**

For the AR difference equation to be valid for values of \( t \) from the distant past into the distant future, then the zeros of the indicial polynomial

\[ \sum_{j=0}^{p} \alpha_j z^{p-j} \]

must all be inside the unit circle, as we have seen that any difference equation having any zero outside the unit circle must explode. (This leaves cases where zeros are on the unit circle which are not legal either because then we would have a process like a random walk which is not stationary). Notice that the indicial polynomial of the difference equation is different from the characteristic polynomial \( g \) of the AR process. In fact the zeros of the two polynomials are reciprocals of each other and so we require the zeros of \( g \) to all be outside the unit circle.

**Ex:** We can look at the zeros for the \( AR(2, \alpha = (.3, .9)) \) process

\[
\begin{align*}
? \ alpha=&<.3,.9> \\
? \ polyroots&(\alpha,1,2,100,rr,ri,ier) \\
? \ mz=&{rr^2+ri^2}^.5
\end{align*}
\]
Thus these $\alpha$'s are legal.

\[ f(\omega) \text{ for an } AR(p) \]

Since the zeros of $g(z)$ are all outside the unit circle, we can invert the model to get

\[ X(t) = \sum_{k=0}^{\infty} \gamma_k \epsilon(t - k), \]

where $\gamma_k$ is the coefficient of $z^k$ in $1/g(z)$. This means that $X$ is a filtered version of white noise and is thus covariance stationary. Thus from the filter theorem and the fact that $g(L)X(t) = \epsilon(t)$, we have

\[ |g(e^{2\pi i \omega})|^{-2} f(\omega) = \sigma^2, \]

or

\[ f(\omega) = \sigma^2 \frac{1}{|g(e^{2\pi i \omega})|^2}. \]

Note that this is the reciprocal of an $MA(p, \alpha, 1/\sigma^2)$ spectral density and so it could have sharp peaks.

**Correlations and the Yule-Walker Equations**

It is easy to show that $\alpha$ and $\sigma^2$ are related to the autocovariances $R(v)$ by equations called the Yule-Walker equations

\[ R(v) + \alpha_1 R(v - 1) + \cdots + \alpha_p R(v - p) = \delta_v \sigma^2, \quad v \geq 0, \]
that is, for $v > 0$, the $R$ sequence satisfies a homogeneous version of the same difference equation that $X$ itself satisfies.

If we divide both sides of this equation by $R(0)$ we also get that the correlation sequence satisfies the same difference equation. Thus if we go back through the solution to a difference equation we can see that $\rho$ must decay to zero either exponentially (like $z^v$) or sinusoidally depending on whether the zeros of $g$ are real or complex or both.

The Yule-Walker equations are a set of linear equations that can be solved for $\alpha$ and $\sigma^2$ if we know $R(0), \ldots, R(p)$ or for the $R$’s if we know $\alpha$ and $\sigma^2$.

For an $AR(1)$, we can write the Yule Walker equations as

$$
R(0) + \alpha R(1) = \sigma^2, \quad v = 0 \\
R(1) + \alpha R(0) = 0, \quad v > 0
$$

and solving the ones for $v = 0$ and $v = 1$ gives

$$
R(0) = \frac{\sigma^2}{1 - \alpha^2},
$$

while the rest of the equations give

$$
R(v) = (-\alpha)^v \frac{\sigma^2}{1 - \alpha^2}, \quad v > 1.
$$
On the other hand, we can solve the equations the other way to get

\[ \alpha = -\rho(1), \quad \sigma^2 = R(0)(1 - \alpha^2). \]

For general \( p \), the equations for \( v = 1, \ldots, p \) give

\[ \Gamma_p \alpha = -r_p, \]

where \( \Gamma_p = \text{Toepl}(R(0), R(1), \ldots, R(p-1)) \) and

\[ r_p = (R(1), \ldots, R(p))^T, \]

while the equation for \( v = 0 \) gives

\[ \sigma^2 = R(0) + \alpha_1 R(1) + \cdots + \alpha_p R(p). \]

### 11.4 Autoregressive Moving Average Models

While the \( AR \) and \( MA \) models can be used for many data sets, there are some data for which they are not adequate, and a more general set of models is needed.

**Def:** The time series \( X \) is said to be an autoregressive moving average (ARMA) process of orders \( p \) and \( q \), coefficients \( \alpha \) and \( \beta \) and noise variance \( \sigma^2 \) if \( X \) satisfies the difference equation

\[
X(t) + \alpha_1 X(t-1) + \cdots + \alpha_p X(t-p) = \epsilon(t) \\
+ \beta_1 \epsilon(t-1) + \cdots + \beta_q \epsilon(t-q),
\]

where \( \epsilon \sim WN(\sigma^2) \).
Remarks:

1. This can be written as \( g(L)X(t) = h(L)\epsilon(t) \)

2. Again, the zeros of \( g \) must all be outside the unit circle for the difference equation to be valid far into the past. Further, to be able to write \( \epsilon(t) \) as a filtered version of present and past \( X \)'s, the zeros of \( h \) must all be outside the unit circle. This means we need to be able to find the coefficients of either \( h(z)/g(z) \) (to write \( X \) as a filtered version of \( WN \)) or \( g(z)/h(z) \) (to write \( \epsilon \) as a filtered version of present and past \( X \)'s) (see the \texttt{DIVPOLY} command).

3. Note that \( ARMA(p, 0) = AR(p), ARMA(0, q) = MA(q) \), and \( ARMA(0, 0) = WN \).

4. If we write \( g(L)X(t) = h(L)\epsilon(t) = Y(t) \), we can use the filter theorem to write down

\[
|g(e^{2\pi i\omega})|^2 f(\omega) = |h(e^{2\pi i\omega})|^2 \sigma^2 = f_Y(\omega),
\]

which gives the spectral density of \( X \) as

\[
f(\omega) = \sigma^2 \frac{|h(e^{2\pi i\omega})|^2}{|g(e^{2\pi i\omega})|^2},
\]

that is, as a constant times the product of an \( MA \) type spectrum times an \( AR \) type spectrum. Thus \( f \) can have sharp peaks and/or sharp troughs.
5. It is easy to derive Yule-Walker type equations relating parameters and autocovariances or autocorrelations:

\[
\sum_{j=0}^{p} \alpha_j R(v - j) = \begin{cases} 
\sigma^2 \sum_{k=v}^{q} \beta_k \gamma_{k-v}, & v = 0, \ldots, q \\
0, & v > q 
\end{cases}
\]

where the \( \gamma \)'s are the coefficients of \( h(z)/g(z) \). Thus when \( v > q \), the sequence of \( R \)'s satisfies a homogeneous difference equation of order \( p \) and thus it decays exponentially or sinusoidally to zero for \( v > q \).

### 11.5 Prediction for ARMA Models

For a general ARMA model, there are no simple formulas for \( \hat{X}_{nh} \) or \( \hat{\sigma}^2_{nh} \) although there is an extensive literature on how to efficiently solve the prediction normal equations (see Section 2.6 of the text). There are several important facts that come from this literature, including

1. If \( X \sim AR(p, \alpha, \sigma^2) \), then the BLUPs of \( X(n+1), X(n+2), \ldots \) given \( X(1), \ldots, X(n) \) are found using the AR difference equation (with the \( \epsilon \)'s set to zero), for example

\[
\begin{align*}
\hat{X}_{n1} & = -\alpha_1 X(n) - \cdots - \alpha_p X(n - p + 1), \\
\hat{X}_{n2} & = -\alpha_1 \hat{X}_{n1} - \cdots - \alpha_p X(n - p + 2),
\end{align*}
\]
and so on, until

$$\hat{X}_{nh} = - \sum_{j=1}^{p} \alpha_j \hat{X}_{n,h-j}, \quad h \geq p,$$

which means that the predictors are decaying exponentially or sinusoidally to zero.

2. It can be shown that the predictors for any $ARMA$ model behave this same way, that is, they decay exponentially or sinusoidally to zero.

3. It can be shown that for large $n$

$$\text{Var}(\hat{X}_{nh} - X(n + h)) \approx \sigma^2 \sum_{j=0}^{h-1} \gamma_j^2,$$

where the $\gamma$'s are the coefficients of $h(z)/g(z)$. This summation converges to $R(0)$ as $h \to \infty$.

11.6 Partial Autocorrelations for ARMA Models

There are also no simple formulas for the partial autocorrelations for general $ARMA$ models. However, it can be shown that there is an interesting duality between $\rho$ and $\theta$ for $AR$, $MA$ and $ARMA$ models; namely

1. While the $\rho$'s for an $MA$ process are truncated (are zero) for lags past $q$, the $\theta$'s for an $AR$ process are truncated past lag $p$. 

2. While the $\rho$'s for an $AR$ process decay exponentially or sinusoidally, the $\theta$'s for an $MA$ process decay exponentially or sinusoidally.

3. While the $\rho$'s for an $ARMA$ process decay exponentially or sinusoidally for lags past $q$, the $\theta$'s for an $ARMA$ decay exponentially or sinusoidally past lag $p$.

**Identifying ARMA Models**

Thus we have seen that the $\rho$'s, $\theta$'s, and $f$ for each of the $MA$, $AR$, and $ARMA$ models have certain qualitative features (for example, $\rho$ is truncated for an $MA$). Given a data set, if we calculate the $\hat{\rho}$'s, $\hat{\theta}$'s, and $\hat{f}$, we can look at them to see if we can detect qualitative features that match a particular model.

Thus an important part of time series analysis is to look at $\rho$'s, $\theta$'s and $f$'s for a wide variety of different $ARMA$ models to gain experience in identifying models. The macro `IDARMA` macro is ideally suited for gaining this experience.