The material in this set of notes is based on S&S Chapter 3, specifically 3.1-3.2. We’re finally going to define our first time series model! The first time series model we will define is the autoregressive (AR) model. We will then consider a different simple time series model, the moving average (MA) model. Putting both models together to create one more general model will give us the autoregressive moving average (ARMA) model.

The AR Model

The first kind of time series model we’ll consider is an autoregressive (AR) model. This is one of the most intuitive models we’ll consider. The basic idea is that we will model the response at time $t$ $x_t$ as a linear function of its $p$ previous values and some independent random noise, e.g.

$$x_t = 0.5x_{t-1} + w_t,$$

where $x_t$ is stationary and $w_t \sim N(0, \sigma_w^2)$. This kind of model is especially well suited to forecasting, as

$$E[x_{t+1}|x_t] = 0.5x_{t-1}.$$
We explicitly define an **autoregressive model of order** $p$, abbreviated as $\text{AR} (p)$ as:

\[(x_t - \mu_x) = \phi_1 (x_{t-1} - \mu_x) + \phi_2 (x_{t-2} - \mu_x) + \cdots + \phi_p (x_{t-p} - \mu_x) + w_t, \quad (3)\]

where $\phi_p \neq 0$, $x_t$ is stationary with mean $\mathbb{E}[x_t] = \mu_x$, and $w_t \overset{i.i.d.}{\sim} \mathcal{N} (0, \sigma^2_w)$. For convenience:

- We’ll often assume $\mu_x = 0$, so

\[x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t. \quad (4)\]

- We’ll introduce the **autoregressive operator** notation:

\[\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \quad (5)\]

where $B^p x_t = x_{t-p}$ is the **backshift operator**. This allows us to rewrite (3) and (4) more concisely as $\phi(B) (x_t - \mu_x) = w_t$ and

\[\phi(B) (x_t) = w_t, \quad (6)\]

respectively.

An $\text{AR} (p)$ model looks like a linear regression model, but the covariates are also random variables. We’ll start building an understanding of the $\text{AR} (p)$ model by starting with the simpler special case where $p = 1$.

The $\text{AR} (1)$ model with $\mu_x = 0$ is a special case of (3)

\[x_t = \phi_1 x_{t-1} + w_t. \quad (7)\]

A natural thing to do is to try to rewrite $x_t$ as a function of $\phi_1$ and the previous values of the random errors. Then (7) will look more like a classical regression problem, as it will no longer have random variables as as covariates. Furthermore, if we could rewrite $x_t$ as a function of $\phi_1$ and the random errors $w$, then $x_t$ would be a **linear process**.

A **linear process** $x_t$ is defined to be a linear combination of white noise $w_t$ and is given
by

\[ x_t = \mu_x + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \]

where the coefficients satisfy \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \), \( w_t \) are independent and identically distributed with mean 0 and variance \( \sigma_w^2 \), and \( \mu_x = \mathbb{E}[x_t] < \infty \). The condition \( \sum_{j=-\infty}^{\infty} |\psi_j| \) ensures that \( x_t = \mu_x + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j} < \infty \). Importantly, it can be shown that the autocovariance function of a linear process is

\[ \gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}, \quad (8) \]

for \( h \geq 0 \), recalling that \( \gamma_x(h) = \gamma_x(-h) \). This means that once we know the linear process representation of any time series process, we can easily compute its autocovariance (and autocorrelation) functions. We will often use the infinite moving average operator shorthand \( 1 + \psi_1 B + \psi_2 B^2 + \ldots \psi_j B^j + \cdots = \psi(B) \).

We can start rewriting \( x_t \) as follows:

\[ x_t = \phi_1^2 x_{t-1} + \phi_1 w_{t-1} + w_t \]
\[ = \phi_1^3 x_{t-2} + \phi_1^2 w_{t-2} + \phi_1 w_{t-1} + w_t \]
\[ = \phi_1^k x_{t-k} + \sum_{j=0}^{k-1} \phi_1^j w_{t-j}. \]

We can see that we can almost lagged values of \( x \) out of the right hand side. Fortunately, when \( |\phi_1| < 1 \), then

\[ \lim_{k \to \infty} \mathbb{E} \left[ \left( x_t - \sum_{j=0}^{k-1} \phi_1^j w_{t-j} \right)^2 \right] = \lim_{k \to \infty} \phi_1^{2k} \mathbb{E} \left[ x_{t-k}^2 \right] = 0, \]

because \( \mathbb{E} \left[ x_{t-k}^2 \right] \) is constant as long as \( x_t \) is stationary is assumed. This means that when \( |\phi_1| < 1 \), then we can write elements of the response \( x_t \) as a linear function the previous
values of the random errors:

\[ x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}. \]  \hspace{1cm} (9)

(9) is the \textbf{linear process} representation of an \textbf{AR} (1) model. It follows that the autocovariance function

\[ \gamma_x (h) = \sigma_w^2 \sum_{j=0}^{\infty} \phi_1^{j+h} \phi_1^j = \sigma_w^2 \phi_1^h \sum_{j=0}^{\infty} \phi_1^{2j} = \sigma_w^2 \phi_1^h \left( \frac{1}{1 - \phi_1^2} \right). \]  \hspace{1cm} (10)

and the autocorrelation function is

\[ \rho_x (h) = \phi^h. \]  \hspace{1cm} (11)

Note that this is not the only way to compute the values of autocovariance function. We could compute them directly from (4),

\[ \gamma_x (h) = \mathbb{E} [x_{t-h}x_t] \]  \hspace{1cm} (12)

\[ = \mathbb{E} [x_{t-h} (\phi_1 x_{t-1} + w_t)] \]

\[ = \phi_1 \mathbb{E} [x_{t-1-h} x_{t-1}] + \mathbb{E} [x_{t-h} w_t] \]

\[ = \phi_1 \gamma_x (h - 1). \]

This gives us a recursive relation that we can use to compute the autocovariance function
\[ \gamma_x(h), \text{starting from } \gamma_x(0). \text{ We can compute } \gamma_x(0) \text{ using substitution:} \]

\[
\begin{align*}
\gamma_x(0) &= \mathbb{E} \left[ x_t^2 \right] \\
&= \mathbb{E} \left[ x_t^2 \right] \\
&= \mathbb{E} \left[ (\phi_1 x_{t-1} + w_t)^2 \right] \\
&= \mathbb{E} \left[ \phi_1^2 x_{t-1}^2 + 2\phi_1 w_t x_{t-1} + w_t^2 \right] \\
&= \phi_1^2 \mathbb{E} \left[ x_{t-1}^2 \right] + \sigma_w^2 \\
&= \sigma_w^2 \sum_{j=0}^{\infty} \phi_1^{2j} \quad \text{(follows from continued substitution)} \\
&= \frac{\sigma_w^2}{1 - \phi_1^2}, \quad \text{if } |\phi_1| < 1, \gamma_x(0) = \infty \text{ otherwise!}
\end{align*}
\]

If \( |\phi_1| < 1 \), then it is easy to see that the AR(1) model \( x \) is stationary because the mean of each \( x_t \) is zero and the autocovariance function \( \gamma_x(h) = \sigma_w^2 \phi_h \left( \frac{1}{1 - \phi^2} \right) \) depends only on the lag, \( h \). What happens when \( |\phi_1| > 1 \)? (9) does not give a linear process representation if \( |\phi_1| > 1 \), because \( \sum_{j=0}^{\infty} |\phi_1|^j = +\infty \).

However when \( |\phi_1| > 1 \), we can revisit (7) and note that \( x_{t+1} = \phi_1 x_t + w_{t+1} \). Rearranging gives

\[ x_t = \left( \frac{1}{\phi_1} \right) x_{t+1} - \left( \frac{1}{\phi_1} \right) w_{t+1}. \]

If \( \phi > 1 \), then \( \left( \frac{1}{\phi_1} \right) < 1 \) and we can use the same approach we used previously to write

\[ x_t = -\sum_{j=1}^{\infty} \left( \frac{1}{\phi_1} \right)^j w_{t+j}. \]

The problem, however, is that this requires that \( x_t \) is a function of future values, which may not be known at time \( t \). We call such a time series non-causal. Using a non-causal model will rarely make sense in practice, and furthermore makes forecasting impossible - in the future, whenever we talk about AR(\( p \)) models we restrict our attention to causal models.

Understanding when a AR(\( p \)) model is causal is more difficult than understanding when
an \( \text{AR}(1) \) model is causal. We figured out when an \( \text{AR}(1) \) model is causal by finding the coefficients \( \ldots, \psi_{-j}, \ldots, \psi_j, \ldots \) of its linear process representation as a function of the AR coefficient \( \phi_1 \), and showing that all of the coefficients \( \psi_{-\infty}, \ldots, \psi_{-1} \) for future errors are exactly equal to zero.

The linear process representation is especially useful for an \( \text{AR}(p) \) model when \( p > 1 \), because computing the autocovariance function \( \gamma_x(h) \) directly as we did in (12) and (13) gets much more cumbersome when \( p > 1 \). We can see this in the \( \text{AR}(2) \) case, where we have

\[
x_t = \phi_2 x_{t-2} + \phi_1 x_{t-1} + w_t.
\]  

We can get a recursive relation for the autocovariance function \( \gamma_x(h) \) starting from \( \gamma_x(0) \) and \( \gamma_x(1) \) as follows:

\[
\gamma_x(h) = \mathbb{E}[x_{t-h} x_t] \\
= \mathbb{E}[x_{t-h} (\phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t)] \\
= \phi_1 \mathbb{E}[x_{t-1-(h-1)} x_{t-1}] + \phi_2 \mathbb{E}[x_{t-2-(h-2)} x_{t-2}] + \mathbb{E}[x_{t-h} w_t] \\
= \phi_1 \gamma_x(h-1) + \phi_2 \gamma_x(h-2).
\]

We can try to compute \( \gamma_x(0) \) and \( \gamma_x(1) \) using substitution:

\[
\gamma_x(0) = \mathbb{E}[x_t^2] \\
= \mathbb{E}[(\phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t)^2] \\
= \mathbb{E}[\phi_1^2 x_{t-1}^2 + \phi_2^2 x_{t-2}^2 + 2\phi_1 \phi_2 x_{t-1} x_{t-2} + 2\phi_1 x_{t-1} w_t + 2\phi_2 x_{t-2} w_t + w_t^2] \\
= \mathbb{E}[\phi_1^2 x_{t-1}^2 + \phi_2^2 x_{t-2}^2 + 2\phi_1 \phi_2 x_{t-1} x_{t-2}] + \sigma_w^2.
\]

However, this gets very complicated, even though we only have two lags!

Unfortunately, it’s much harder to find the linear process representation of an \( \text{AR}(p) \) model by simple substitution as we did with an \( \text{AR}(1) \) model. Substituting according to
Again, this is not working out nicely!

Instead, we can find the values of \( \psi_1, \ldots, \psi_j, \ldots \) that satisfy \( \phi(B) \psi(B) w_t = w_t \), which follows from substituting \( x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j} \) into (20). This is equivalent to finding the inverse function \( \phi^{-1}(B) \) that satisfies \( \phi(B) \phi^{-1}(B) w_t = w_t \).

We can see how this method for finding the values of \( \psi_1, \ldots, \psi_j, \ldots \) works by returning to the AR(1) case. The values \( \psi_1, \ldots, \psi_j, \ldots \) that satisfy \( \phi(B) \psi(B) w_t = w_t \) solve:

\[
1 + (\psi_1 - \phi_1) B + (\psi_2 - \psi_1 \phi_1) B^2 + \cdots + \psi_j B^j + \ldots = 1, \tag{15}
\]

where (15) follows from expanding \( \phi(B) \) and \( \psi(B) \). This allows us to recover the linear process representation of the AR(1) process in a different way, as (15) holds if all of the coefficients for \( B^j \) with \( j > 0 \) are equal to zero, i.e. \( \psi_k - \psi_{k-1} \phi_1 = 0 \) for \( k > 1 \).

Now let’s try this approach for the AR(2) case. We have

\[
1 = (1 - \phi_1 B - \phi_2 B^2) \left( 1 + \psi_1 B + \psi_2 B^2 + \cdots + \psi_j B^j + \ldots \right) \\
= 1 + (\psi_1 - \phi_1) B + (\psi_2 - \phi_2 - \phi_1 \psi_1) B^2 + (\psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1) B^3 + \cdots + \\
(\psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2}) B^j + \ldots
\]
We see that we can compute the values of $\psi_1, \ldots, \psi_j, \ldots$ recursively,

$$
\begin{align*}
\psi_1 &= \phi_1 \\
\psi_2 &= \phi_2 + \phi_1^2 \\
\psi_3 &= 2\phi_1 (\phi_2 + \phi_1^2),
\end{align*}
$$

and so on.

It’s also very tricky to figure out when AR ($p$) model is causal for $p > 1$. An AR ($p$) model is causal for $p > 1$ model is causal when all of the roots of the AR polynomial

$$
\phi (z) = 1 - \phi_1 z - \cdots - \phi_p z^p,
$$

lie outside the unit circle, i.e. $\phi (z) \neq 0$ for $|z| \leq 1$. This condition ensures that the $\sum_{j=1}^{\infty} |\psi_j| < \infty$. This is not very intuitive. If we want to try to get a handle on why the roots of the AR polynomial need to lie outside the unit circle for a AR($p$) model to be causal, we need to take a look at the proof. You won’t be tested on your understanding of this - we’ll just go through it here in case you are curious following along the proof of Theorem 3.2 in Chan (2010).

Let’s suppose that $\phi (z)$ has roots $r_1, \ldots, r_p$ that satisfy $1 < |r_1| \leq \cdots \leq |r_p|$, i.e. $\phi (r_j) = 0$ for $j = 1, \ldots, p$. Then this ensures that we can invert $\phi (z)$ when $z \leq |r_1|$. Recalling that $\psi (B)$ can be thought of as the inverse of $\phi (B)$, this means that

$$
\frac{1}{\phi (z)} = \sum_{j=0}^{\infty} \psi_j z^j < \infty \text{ if } |z| \leq |r_1|,
$$

where $\psi_0 = 1$. Then we can invert $\phi (z)$ at any value of $z < |r_1|$, e.g. at $z = 1 + \delta < |r_1|$, where $\delta > 0$. Writing this out, we have

$$
\frac{1}{\phi (1 + \delta)} = \sum_{j=0}^{\infty} \psi_j (1 + \delta)^j < \infty. \quad (16)
$$

If (16), then there must be some constant $M > 0$ that gives an upper bound for all $|\psi_j (1 + \delta)^j|$, i.e. $|\psi_j (1 + \delta)^j| \leq M$ for all $j = 0, 1, 2, \ldots$. Shifting things around, this is
equivalent to \(|\psi_j| \leq M (1 + \delta)^{-j}\). Then
\[
\sum_{j=1}^{\infty} |\psi_j| \leq M \sum_{j=1}^{\infty} \left( \frac{1}{1 + \delta} \right)^j
= M \left( \sum_{j=0}^{\infty} \left( \frac{1}{1 + \delta} \right)^j - 1 \right)
= M \left( \frac{1}{1 - \frac{1}{1 + \delta}} - 1 \right)
= M \left( \frac{1 + \delta}{1 + \delta - 1} - 1 \right) = M \left( \frac{1}{\delta} \right) < \infty.
\]

The MA Model

Instead of assuming that elements of a time series \(x_t\) are linear function of previous elements of the time series \(x_1, \ldots, x_{t-1}\) and independent, identically distributed noise \(w_t\), we might assume that elements of a time series \(x_t\) are a linear function of all of the current and previous noise variates, \(w_1, \ldots, w_{t-1}\). The latter gives us the moving average model of order \(q\), abbreviated as MA \((q)\). The MA \((q)\) model is explicitly defined as
\[
x_t - \mu_x = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}, \tag{17}
\]
where \(\theta_q \neq 0\), \(\mathbb{E} [x_t] = \mu_x\), and \(w_t \sim i.i.d. \mathcal{N}(0, \sigma_w^2)\). For convenience:

- We’ll often assume \(\mu_x = 0\), so
\[
x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}. \tag{18}
\]

- We’ll introduce the moving average operator notation:
\[
\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^p, \tag{19}
\]
which allows us to rewrite (17) and (18) more concisely as \(x_t - \mu_x = \theta(B) w_t\) and
\[
x_t = \theta(B) w_t, \tag{20}
\]
respectively.

Again, the \( \text{MA} (q) \) model looks like a linear regression model. Importantly, the \( \text{MA} (q) \) model is stationary for any values of the parameters \( \theta_1, \ldots, \theta_q \).

Like we did with the \( \text{AR} (p) \) model, we’ll start building an understanding of the \( \text{MA} (q) \) by starting with the simpler special case where \( q = 1 \),

\[
x_t = \theta_1 w_{t-1} + w_t. \tag{21}
\]

It is easy to see that this \( \text{MA} (q) \) model is mean zero. We can compute the autocovariance function as follows:

\[
\gamma_x (h) = \mathbb{E} [x_t x_{t-h}]
= \mathbb{E} [(\theta_1 w_{t-1} + w_t) (\theta_1 w_{t-h} + w_{t-h})]
= \mathbb{E} [\theta_1^2 w_{t-1} w_{t-h} + \theta_1 w_{t-1} w_{t-h} + \theta_1 w_{t-1} w_{t-h} + w_t w_{t-h}]
= \mathbb{E} [\theta_1^2 w_{t-1} w_{t-h} + \theta_1 w_{t-1} w_{t-h} + w_t w_{t-h}]
= \begin{cases} 
\sigma_w^2 (\theta_1^2 + 1) & h = 0 \\
\theta_1 & h = 1 \\
0 & h > 1 
\end{cases} \tag{22}
\]

The corresponding autocorrelation function is

\[
\rho_x (h) = \begin{cases} 
\frac{\theta_1}{\theta_1^2 + 1} & h = 1 \\
0 & h > 1 
\end{cases} \tag{23}
\]

The autocovariance and autocorrelation functions of the \( \text{MA} (q) \) model are noteworthy in two ways:

\( \bullet \) The autocorrelation function \( \rho_x (h) \) is bounded, \( \rho_x (h) \leq 1/2 \) for \( h = 1 \).

\( \ast \) The parameters of the \( \text{MA} (q) \) model do not uniquely determine the autocovariance and autocorrelation function values. \( \theta_1 \) and \( \sigma_w^2 \) do not uniquely determine the value
of the autocovariance function $\gamma_x(h)$, and $\theta_1$ does not determine the value of the autocorrelation function.

It is easiest to understand (*) via some examples. First, we compute $\gamma_x(h)$ and $\rho_x(h)$ for a MA(1) process with $\theta_1 = 5$ and $\sigma_w^2 = 1$,

$$\gamma_x(h) = \begin{cases} 5^2 + 1 = 26 & h = 0 \\ 5 & h = 1 \\ 0 & h > 1 \end{cases} \quad \text{and} \quad \rho_x(h) = \begin{cases} \frac{5}{5^2+1} = \frac{5}{26} & h = 1 \\ 0 & h > 1 \end{cases}.$$

Compare this to $\gamma_x(h)$ and $\rho_x(h)$ for a MA(1) process with $\theta_1 = 1/5$ and $\sigma_w^2 = 25$,

$$\gamma_x(h) = \begin{cases} 25 \left(\frac{1}{5^2} + 1\right) = 25 \left(\frac{1+25}{25}\right) = 26 & h = 0 \\ 25 \left(\frac{1}{5}\right) = 5 & h = 1 \\ 0 & h > 1 \end{cases} \quad \text{and} \quad \rho_x(h) = \begin{cases} \frac{\frac{5}{2}}{\frac{5}{2^2}+1} = \frac{5}{26} & h = 1 \\ 0 & h > 1 \end{cases}.$$

Both sets of MA(1) parameters give the values of the autocovariance and autocorrelation functions! This is undesirable - it means that even if we know that our time series is mean zero with a specific autocovariance function $\gamma_x(h)$ autocorrelation function $\rho_x(h)$, we can’t find a unique pair of corresponding MA(1) parameter values $(\theta_1, \sigma_w^2)$. 

We solve this problem by requiring that our MA(1) model be invertible, which means that it has an infinite autoregressive representation $(1 + \pi_1 B + \pi_2 B^2 + \cdots + \pi_j B^j + \cdots) x_t = w_t$ with $\sum_{j=1}^{\infty} |\pi_j| < \infty$. We can find a unique pair of corresponding MA(1) parameter values $(\theta_1, \sigma_w^2)$ if we restrict our attention to the parameter values that give an invertible MA(1) model. What we mean by this is that we can rearrange (21) to resemble a AR(1)
model for \( w_t \),

\[
\begin{align*}
  w_t &= -\theta_1 w_{t-1} + x_t \\
  &= \theta_1^2 w_{t-2} - \theta_1 x_{t-1} + x_t \\
  &= -\theta_1^3 w_{t-3} + \theta_1^2 x_{t-2} - \theta_1 x_{t-1} + x_t \\
  &= (-\theta_1)^k w_{t-k} + \sum_{j=0}^{k} (-\theta_1)^j x_{t-j},
\end{align*}
\]

where \( \lim_{k \to \infty} (-\theta_1)^k w_{t-k} + \sum_{j=0}^{k} (-\theta_1)^j x_{t-j} = \sum_{j=0}^{\infty} (-\theta_1)^j x_{t-j} \). Recalling the AR(1) model, this will be the case when \( |\theta_1| < 1 \). Going back to our example where we considered the MA(1) parameters \((\theta_1, \sigma_w^2) = (5, 1)\) and \((\theta_1, \sigma^2_w) = \left(\frac{1}{5}, 25\right)\), this means that only the latter pair \((\theta_1, \sigma_w^2) = \left(\frac{1}{5}, 25\right)\) satisfy our definition of a MA(1) model.

More generally, requiring that an MA(q) model be invertible ensures that we can find a unique set of corresponding MA(q) parameter values \((\theta_1, \ldots, \theta_q, \sigma^2_w)\) if we know that our time series is MA(q) with mean zero, a specific autocovariance function \( \gamma_x(h) \), and autocorrelation function \( \rho_x(h) \). We introduce some additional notation for this; an MA(q) model is invertible if we can write \( w_t = \pi(B) x_t \), where \( \pi(B) = 1 + \pi_1 B + \cdots + \pi_j B^j + \cdots \) and \( \sum_{j=0}^{\infty} |\pi_j| < \infty \). This looks a lot like the problem of ensuring that a AR(p) model is causal, and it turns out that an MA(q) model is invertible if when all of the roots of the MA polynomial

\[
\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,
\]

lie outside the unit circle, i.e. \( \theta(z) \neq 0 \) for \( |z| \leq 1 \).
The ARMA Model

The autoregressive moving average (ARMA) model combines the AR and MA models. We define an ARMA \((p, q)\) model as:

\[
(x_t - \mu_x) = \phi_1 (x_{t-1} - \mu_x) + \cdots + \phi_p (x_{t-p} - \mu_x) + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} + w_t,
\]

(24)

where \(w_t \sim i.i.d. \mathcal{N}(0, \sigma_w^2)\), \(x_t\) is stationary, \(\phi_p \neq 0, \theta_q \neq 0, \sigma_w^2 > 0\), and the MA and AR polynomials \(\theta(B)\) and \(\phi(B)\) have no common roots. We refer to \(p\) as the autoregressive order and \(q\) as the moving average order. Again, for convenience we will usually assume \(\mu_x = 0\), so

\[
x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}.
\]

(25)

Using operator notation becomes especially beneficial for ARMA \((p, q)\) models; we can just write \(\phi(B) x_t = \theta(B) w_t\) instead of (25). Note that:

- Setting \(p = 0\) gives a MA \((q)\) model;
- Setting \(q = 0\) gives an AR \((p)\).

As with AR \((p)\) and MA \((q)\) models, we will need to figure out when an ARMA \((p, q)\) is causal and invertible. Fortunately, this is simple given the work we’ve already done for MA \((q)\) and AR \((p)\) models. An ARMA \((p, q)\) is:

- **Causal**, i.e. we can find \(\psi_1, \ldots, \psi_j, \ldots\) such that \(\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}\) that satisfy \(\sum_{j=0}^{\infty} |\psi_j| < \infty\) for \(|z| < 1\), if \(\phi(z) \neq 0\) for \(|z| \leq 1\);

- **Invertible**, i.e. we can find \(\pi_1, \ldots, \pi_j, \ldots\) such that \(\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}\) that satisfy \(\sum_{j=0}^{\infty} |\pi_j| < \infty\) for \(|z| < 1\), if \(\theta(z) \neq 0\) for \(|z| \leq 1\).

Returning to the definition of an ARMA \((p, q)\) model, it is not immediately obvious why we require that the moving average and autoregressive polynomials \(\theta(B)\) and \(\phi(B)\) have no
common roots. Consider the following model, which resembles an \textbf{ARMA} \((p, q)\) model:

\begin{equation}
x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t,
\end{equation}

where \(x_t\) is stationary and \(w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_w)\). It’s easy to see that the mean function \(\mu_x = 0\). The autocovariance function \(\gamma_x (h)\) satisfies:

\begin{equation}
\gamma_x (h) = \mathbb{E}[x_t x_{t-h}]
= \mathbb{E}[(0.5x_{t-1} - 0.5w_{t-1} + w_t) x_{t-h}]
= 0.5\mathbb{E}[x_{t-1} x_{t-h}] - 0.5\mathbb{E}[w_{t-1} x_{t-h}] + \mathbb{E}[w_t x_{t-h}]
= \begin{cases} 
0.5\gamma_x (0) - 0.5\sigma^2_w & h = 1 \\
0.5\gamma_x (h - 1) & h > 1 
\end{cases}
\end{equation}

We just need to combine this with a starting value, \(\gamma_x (0)\):

\begin{equation}
\gamma_x (0) = \mathbb{E}[x_t^2]
= \mathbb{E}[0.5^2 x_{t-1}^2 + 0.5^2 w_{t-1}^2 + w_t^2 - (2) (0.5)^2 w_{t-1}^2]
= 0.5^2 \gamma_x (0) + (1 - 0.5^2) \sigma^2_w \implies \gamma_x (0) = \sigma^2_w
\end{equation}

Plugging this in to (27), for \(h > 0\) we get

\[\gamma_x (h) = 0!\]

This means that (26) is equivalent to the white noise model, \(x_t = w_t!\)

If we examine the corresponding AR and MA polynomials, we see that they share the common factor \(1 - 0.5B\), \(\theta (B) = 1 - 0.5B\) and \(\phi (B) = 1 - 0.5B\). Dividing each by the common factor yields \(\theta (B) = 1\) and \(\phi (B) = 1\), which gives us the familiar definition of the white noise model, \(x_t = w_t\). This is why we require that the the moving average and autoregressive polynomials \(\theta (B)\) and \(\phi (B)\) have no common roots, otherwise we could mistake a white noise process for an \textbf{ARMA}(p, q) process with \(p, q > 0\).

As with the \textbf{AR}(p) model, the linear process representation of an \textbf{ARMA}(p, q) model
is especially useful for computing the autocovariance function of an \textbf{ARMA} \((p, q)\) model. Using the same approach we used for the \textbf{AR} \((p)\) model, the values of \(\psi_1, \ldots, \psi_j, \ldots\) that satisfy \(x_t = \psi(B) w_t\) with \(\sum_{j=0}^{\infty} |\psi_j| < \infty\) can be computed by substituting \(\psi(B) w_t\) into the equation that defines the \textbf{ARMA} \((p, q)\) model, \(\phi(B) x_t\), and matching the coefficients for each power of \(B\) on each side, i.e.

\[
\phi(B) \psi(B) w_t = \theta w_t
\]

\[
\implies (1 - \phi_1 B - \ldots \phi_p B^p) \left(1 + \psi_1 B + \ldots \psi_j B^j\right) w_t = (1 + \theta_1 B + \ldots + \theta_q B^q) w_t.
\]

This yields a sequence of equations that would start with

\[
\psi_1 - \phi_1 = \theta_1 \\
\psi_2 - \phi_2 - \phi_1 \psi_1 = \theta_2,
\]

and continue on for \(\psi_3, \ldots, \psi_j, \ldots\). We will not be computing \(\psi_1, \ldots, \psi_j, \ldots\) by hand in class - this requires a knowledge of differential equations that goes above and beyond the prerequisites for this course. However, statistical software like \texttt{R} will often include functions that can be used to compute the \(\phi_1, \ldots, \phi_K\) for some user specified value \(K > 1\) given values for \(\phi_1, \ldots, \phi_p\) and \(\theta_1, \ldots, \theta_p\).