STAT 5170: Applied Time Series

Course notes for part A of learning unit 3

In this learning unit, our focus is on the theoretical properties of autoregressive and moving-average time series. In the next part of the unit, we will combine our understanding of these types of time series models into an expanded class of time series models known as autoregressive moving-average models.

Section 3A.1: Infinite-sum representations.

In our previous exploration with these time series, we worked with induction arguments in order to derive representations that allowed us to explore their properties. Partly to re-familiarize us to these ideas, and to prepare us for what is to come, let us clarify and extend some of the properties that we had previously explored.

To begin, suppose (x_t) is an $AR(1)$ time series, and repeatedly apply the autoregressive relationship to see that

$$\begin{aligned} x_t &= \phi_1 x_{t-1} + w_t \\ &= \phi_1\{\phi_1 x_{t-2} + w_{t-1}\} + w_t \\ &= \phi_1^2 x_{t-2} + \phi_1 w_{t-1} + w_t \\ &= \phi_1^2\{\phi_1 x_{t-3} + w_{t-2}\} + \phi_1 w_{t-1} + w_t \\ &= \phi_1^3 x_{t-3} + \phi_1^2 w_{t-2} + \phi_1 w_{t-1} + w_t \end{aligned}$$

Such manipulations set up an induction argument that would establish the representation

$$x_t = \phi_1^k x_{t-k} + \sum_{j=0}^{k-1} \phi_1^j w_{t-j},$$

for any value k. Recall that we had previously derived a similar representation, but there the initial value x_0 replaces the quantity x_{t-k} in the first term of the above formula. Here, rather than anchoring our thinking to an initial value, the time series (x_t) is to be conceptualized as a doubly-infinite sequence $\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots$ From the representation, above, the following property is readily deduced.

Property: Suppose $|\phi_1| < 1$ and (x_t) is stationary. It follows that

$$\phi_1^k x_{t-k} \to 0 \quad \text{as} \quad k \to 0$$

and

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j}.$$

$\square$

The first conclusion of this property follows immediately from the convergence of $\phi_1^k \to 0$ as $k \to 0$ when $|\phi_1| < 1$. The second conclusion is interesting for expressing the autoregressive times series, (x_t) completely in terms of the white noise times series, (w_t), as an infinite sum.

Working in the other direction, we may instead study the time series (x_t) by starting with the infinite-sum representation, and then deducing the above property's converse.

Property: If $|\phi_1| < 1$ and (w_t) is Gaussian white noise, then the time series (x_t) defined from the infinite-sum representation

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j}$$

is well-defined, stationary, and satisfies the autoregressive relationship.

$$x_t = \phi_1 x_{t-1} + w_t.$$

□

To establish that the time series is well-defined requires we check that the infinite-sum representation defines a valid joint distribution for (x_t). Doing so requires some rather technical tools from probability theory and is not discussed. Establishing that (x_t) satisfies the $AR(1)$ autoregressive relationship may be carried out by direct substitution:

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j} = \phi_1 \sum_{j=1}^{\infty} \phi_1^{j-1} w_{t-j} + w_t = \phi_1 \left(\sum_{j=0}^{\infty} \phi_1^j w_{t-j-1} \right) + w_t = \phi_1 x_{t-1} + w_t.$$

Working further with the infinite-sum representation, the autocovariance function is derived as

$$\begin{aligned} \gamma(s,t) &= Cov\left[\sum_{j=0}^{\infty} \phi_1^j w_{s-j}, \sum_{k=0}^{\infty} \phi_1^k w_{t-k} \right] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_1^j \phi_1^k Cov[w_{s-j}, w_{t-k}] \\ &= \sigma_w^2 \sum_{k=0}^{\infty} \phi_1^{s-t+k} \phi_1^k. \end{aligned}$$

The second step of this derivation follows from the observation that the covariance $Cov[w_{s-j}, w_{t-k}]$ is nonzero only if $s - j = t - k$, which implies $j = s - t + k$. This autocovariance formula confirms that the autocovariance function $\gamma(s,t)$ depends on s and t only through $s - t$, which completes the deduction that (x_t) is stationary.

If we continue working with the autocovariance formula, however, we see that it is simplified by applying the geometric-series formula

$$\sum_{i=0}^{\infty} u^i = \frac{1}{1-u}, \text{ for } |u| \le 1,$$

from which it follows that

$$\gamma(s,t) = \sigma_w^2 \phi_1^{s-t} \sum_{k=0}^{\infty} \phi_1^{2k} = \frac{\sigma_w^2 \phi_1^{s-t}}{1 - \phi_1^2}.$$

Having established stationarity, it makes sense to express the autocovariance formula more simply as the autocovariance function

$$\gamma(h) \quad = \quad \frac{\sigma_w^2 \phi_1^h}{1-\phi_1^2}.$$

The corresponding autocorrelation function is

$$\rho(h) \quad = \quad \frac{\gamma(h)}{\gamma(0)} = \left(\frac{\sigma_w^2 \phi_1^h}{1-\phi_1^2}\right)\left(\frac{1-\phi_1^2}{\sigma_w^2 \phi_1^0}\right) = \phi_1^h.$$

For the sake of completeness in our exploration of the infinite-sum representation, the mean function of (x_t) is quickly deduced:

$$\mu_t \quad = \quad E[x_t] = \sum_{j=0}^{\infty} \phi_1^j E[w_{t-j}] = 0.$$

We see, therefore, that (x_t) is a mean-zero time series. If we want to specify a non-zero mean, μ, redefine the times series as

$$x_t = \mu + \sum_{j=0}^{\infty} \phi_1^j w_{t-j}.$$

An equivalent expression is

$$x_t - \mu = \sum_{j=0}^{\infty} \phi_1^j w_{t-j},$$

which, by our previous work, identifies that the time series (y_t) defined by $y_t = x_t - \mu$ is a mean-zero time series that satisfies the $AR(1)$ relationship

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + w_t.$$

Rearranging terms, it follows that (x_t) satisfies the relationship

$$x_t = \alpha + \phi_1 x_{t-1} + w_t,$$

where

$$\alpha = (1 - \phi_1)\mu.$$

A generalization of the latter observation to non-zero mean $AR(p)$ autoregressive processes is easy to obtain. Here, the mean-zero time series (y_t), defined by $y_t = x_t - \mu$, satisfies

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t.$$

from which it follows that

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t,$$

where

$$\alpha = (1 - \phi_1 - \cdots - \phi_p)\mu.$$

Similar use of induction will produce infinite-sum representations of other time-series models, though the relevant formulas can become quite complicated. The next example illustrates this for the $AR(p)$ models.

Example: Infinite-sum representation of an $AR(p)$ time series

Suppose the time series (x_t) is such that

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t,$$

where $\phi_1, \ldots, \phi_p$ are autoregressive parameters and (w_t) is Gaussian white noise. With a bit of effort, it is possible to deduce the formula

$$\begin{aligned} x_t &= \sum_{\boldsymbol{j} \in C_k} \binom{k}{j_1, \ldots, j_p} \phi_1^{j_1} \cdots \phi_p^{j_p} x_{t-g(\boldsymbol{j})} \\ &\qquad + \sum_{l=0}^{k-1} \sum_{\boldsymbol{j} \in C_l} \binom{l}{j_1, \ldots, j_p} \phi_1^{j_1} \cdots \phi_p^{j_p} w_{t-g(\boldsymbol{j})}, \end{aligned}$$

where $\boldsymbol{j} = (j_1, \ldots, j_p)$ is a p-tuple of non-negative integers, C_k is the set of such p-tuples that satisfy $j_1 + \cdots + j_p = k$,

$$\binom{k}{j_1, \ldots, j_p} = \frac{k!}{j_1! \cdots j_p!}$$

is the multinomial coefficient, and $g(\boldsymbol{j}) = j_1 + 2j_2 + \ldots + pj_p$.

For example, when $p = 4$, this formula evaluated at $k = 1, 2, 3$ yields

$$\begin{aligned} x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \phi_4 x_{t-4} + w_t \\ &= \phi_1^2 x_{t-2} + 2\phi_1\phi_2 x_{t-3} + (2\phi_1\phi_3 + \phi_2^2) x_{t-4} + 2(\phi_1\phi_4 + \phi_2\phi_3) x_{t-5} \\ &\quad + w_t + \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4}. \\ &= \phi_1^3 x_{t-3} + 3\phi_1^2\phi_2 x_{t-4} + 3(\phi_1^2\phi_3 + \phi_2^2\phi_1) x_{t-5} + (\phi_2^3 + 3\phi_1^2\phi_4 + 6\phi_1\phi_2\phi_3) x_{t-6} \\ &\quad + 3(\phi_2^2\phi_3 + \phi_3^2\phi_1 + 2\phi_1\phi_2\phi_4) x_{t-7} + 3(\phi_2^2\phi_4 + \phi_3^2\phi_2 + 2\phi_1\phi_3\phi_4) x_{t-8} \\ &\quad + (3\phi_4^2\phi_1 + \phi_3^3 + 6\phi_2\phi_3\phi_4) x_{t-9} + 3(\phi_3^2\phi_4 + \phi_4^2\phi_2) x_{t-10} + 3\phi_4^2\phi_3 x_{t-11} + \phi_4^3 x_{t-12} \\ &\quad + w_t + \phi_1 w_{t-1} + (\phi_2 + \phi_1^2) w_{t-2} + (\phi_3 + 2\phi_1\phi_2) w_{t-3} + (\phi_4 + 2\phi_1\phi_3 + \phi_2^2) w_{t-4} \\ &\quad + 2(\phi_1\phi_4 + \phi_2\phi_3) w_{t-5}. \end{aligned}$$

Notice in this example that the terms are grouped in terms of distinct values of $g(\boldsymbol{j})$, which determines the subscript of the associated time-series value. This is distinct from the general formula's implied ordering of terms.

Under certain conditions, one of which is stationarity, the first term in the general formula would tend to zero as $k \to \infty$, and the time series may be represented by the infinite sum

$$x_t = \sum_{l=0}^{\infty} \sum_{\boldsymbol{j} \in C_l} \binom{l}{j_1, \ldots, j_p} \phi_1^{j_1} \cdots \phi_p^{j_p} w_{t-g(\boldsymbol{j})}.$$

We have not yet determined conditions under which the $\phi_1, \ldots, \phi_p$ induce the time series to be stationary; this will be addressed in part B of the learning unit.

It is possible to rewrite the infinite-sum representation through an equivalent formula that group terms by distinct subscript values of the white-noise terms:

$$x_t = \sum_{g=0}^{\infty} \sum_{\boldsymbol{j} \in B_g} \binom{|\boldsymbol{j}|}{j_1, \ldots, j_p} \phi_1^{j_1} \cdots \phi_p^{j_p} w_{t-g}.$$

where B_g is the set of non-negative integer p-tuples, $\boldsymbol{j} = (j_1, \ldots, j_p)$, such that $g(\boldsymbol{j}) = g$, and $|\boldsymbol{j}| = j_1 + \cdots + j_p$.

With this, the autocovariance function may be deduced as

$$\begin{aligned} Cov[x_s, x_t] &= \sum_{g=0}^{\infty} \sum_{h=0}^{\infty} \sum_{\boldsymbol{j} \in B_g} \sum_{\boldsymbol{k} \in B_h} \binom{|\boldsymbol{j}|}{j_1, \ldots, j_p} \binom{|\boldsymbol{k}|}{k_1, \ldots, k_p} \phi_1^{j_1+k_1} \cdots \phi_p^{j_p+k_p} Cov[w_{s-g}, w_{t-h}] \\ &= \sigma_w^2 \sum_{g=0}^{\infty} \sum_{\boldsymbol{j} \in B_g} \sum_{\boldsymbol{k} \in B_{g+s-t}} \binom{|\boldsymbol{j}|}{j_1, \ldots, j_p} \binom{|\boldsymbol{k}|}{k_1, \ldots, k_p} \phi_1^{j_1+k_1} \cdots \phi_p^{j_p+k_p}, \end{aligned}$$

where the second step follows from the property $Cov[w_{s-g}, w_{t-h}] = 0$ unless $s - g = t - h$.

This formula is quite challenging to work with directly. However it does identify that the autocovariance $Cov[x_s, x_t]$ depends on s and t only through $t - s$, which is a reflection of stationarity. Additional work with this formula will establish that the autocovariance function $\gamma(h) = Cov[x_{t+h}, x_t]$ obeys the relationship

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p),$$

for $h = 1, 2, \ldots$. Equations of this sort are sometimes called *difference equations*. Upon dividing each by $\gamma(0)$ we arrive an equivalent set of difference equations defined by

$$\rho(h) = \phi_1 \rho(h-1) + \cdots + \phi_p \rho(h-p).$$

Because $\rho(0) = 1$ and $\rho(-h) = \rho(h)$, we can identify starting values with which to calculate a recursive solution to the difference equations.

For instance, in the case $p = 2$, the difference equations are

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2).$$

We know $\rho(0) = 1$; and, also, by applying $\rho(-1) = \rho(1)$ to the difference equation $\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1)$ and solving, that $\rho(1) = \phi_1/(1 - \phi_2)$. From this, the difference equations provide the solutions

$$\begin{aligned}
\rho(0) &= 1 \\
\rho(1) &= \phi_1/(1-\phi_2) \\
\rho(2) &= \phi_1\rho(1) + \phi_2\rho(0) = \phi_1^2/(1-\phi_2) + \phi_1 \\
\rho(3) &= \phi_1\rho(2) + \phi_2\rho(1) = \phi_1^3/(1-\phi_2) + \phi_1\phi_2(2-\phi_2)/(1-\phi_2)
\end{aligned}$$

and so on.

In the case of general p, one would first determine initial values by solving the system of difference equations that are defined at the settings $h = 0, \ldots, p-1$, given by

$$\begin{aligned}
\rho(0) &= 1 \\
\rho(1) &= \phi_1\rho(0) + \phi_2\rho(-1) + \cdots + \phi_p\rho(1-p) \\
&\vdots \\
\rho(p-1) &= \phi_1\rho(p-2) + \cdots + \phi_{p-1}\rho(0) + \phi_p\rho(-1),
\end{aligned}$$

together with the property $\rho(-h) = \rho(h)$. Subsequently, the autocorrelation function for $h \geq p$ would be determined recursively. $\square$

Section 3A.2: The backshift operator.

As the last example illustrates, working directly with infinite-sum representations can be challenging. Fortunately, a special notational device called the *backshift operator*, B, is available that simplifies some related deductions, and helps to identify important time-series concepts.

The action of the backshift operator is to shift the subscript of a time series variable one unit backward in time; thus $Bx_t = x_{t-1}$. An exponentiated backshift operator is to be interpreted as multiple applications of the operator; thus, $B^k x_t = x_{t-k}$.

The backshift operator may be used to define an autoregressive time series by first defining an *autoregressive operator* according to

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p.$$

Subsequently, the autoregressive relationship may be expressed as $\phi(B)x_t = w_t$; observe

$$\begin{aligned}
\phi(B)x_t &= (1 - \phi_1 B - \cdots - \phi_p B^p)x_t \\
&= x_t - \phi_1 Bx_t - \cdots - \phi_p B^p x_t \\
&= x_t - (\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p}) = w_t.
\end{aligned}$$

The backshift operator may similarly be used to define a *moving-average operator* according to

$$\theta(B) = 1 + \theta_1 B + \cdots - \theta_q B^q,$$

from which the moving-average relationship is $x_t = \theta(B)w_t$, since

$$\begin{aligned} x_t &= w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-p} \\ &= w_t + \theta_1 B w_t + \cdots + \theta_q B^q w_t \\ &= (1 + \theta_1 B + \cdots + \theta_q B^q) w_t = \theta(B) w_t. \end{aligned}$$

For purposes of studying infinite-sum representations, let us define the *infinite moving-average operator*

$$\psi(B) \quad = \quad 1 + \phi_1 B + \phi_1^2 B^2 + \cdots ,$$

for absolutely summable coefficients $\psi_0, \psi_1, \ldots$. This allows us to conveniently reference the one-sided linear process

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

as $x_t = \psi(B)w_t$. Among other uses, we will use this operator to better understand infinite-sum representations of $AR(p)$ time series. We know, for example, that a stationary $AR(1)$ time series, which is characterized by the autoregressive operator $\phi(B) = 1 - \phi_1 B$, has the infinite-sum representation

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j};$$

it is therefore alternatively characterized by the infinite moving-average operator with coefficients $\psi_j = \phi_1^j$.

With these concepts on hand, the backshift operator notation offers an alternative approach to deriving the infinite-sum $AR(1)$ representation that avoids the use of induction arguments. To see this, suppose (x_t) is $AR(1)$, with autoregressive operator $\phi(B) = 1 - \phi_1 B$, and we wish to derive its infinite-sum representation in terms of the infinite moving-average operator $\psi(B)$. That is, from the relationship $\phi(B)x_t = w_t$ we wish to derive the relationship $x_t = \psi(B)w_t$.

A step in this direction is to substitute $x_t = \psi(B)w_t$ into $\phi(B)x_t = w_t$ and rearrange terms:

$$\phi(B)\psi(B)w_t = w_t \quad \Leftrightarrow \quad \psi(B)w_t = \phi^{-1}(B)w_t$$

Here, the operator $\phi^{-1}(B)$ is the inverse of $\phi(B)$, which is defined by the relationship

$$\phi^{-1}(B)\phi(B)x_t = x_t.$$

Assuming the inverse operator, $\phi^{-1}(B)$, exists, its form is implied by power series expansion: observe

$$\psi(B) \quad = \quad \phi^{-1}(B) = \frac{1}{1 - \phi_1 B} = 1 + \phi_1 B + \phi_1^2 B^2 + \cdots .$$

From this, we see that $\psi(B)$ is the infinite moving-average operator with coefficients $\psi_j = \phi_1^j$, just as we had observed before.

Part of the point of this deduction is to illustrate that operators such as $\phi(B)$ and $\psi(B)$, which are defined from backshift operators, may be manipulated in parallel ways as analytical formulas, provided they are ultimately understood as (finite or infinite) polynomials, so that the exponentiated backshift operators make sense.

Section 3A.3: Causality.

Working with infinite-series representations leads to a concept that we have not yet carefully considered, called *causality.* This is illustrated as follows.

Example: AR(1) time series and causality

In the case $|\phi_1| > 1$, the expression

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j}.$$

does not make sense, since the magnitude of ϕ_1^j explodes as $j \to \infty$.

To handle this case, rewrite the autoregressive relationship, and rearrange from

$$x_{t+1} = \phi_1 x_t + w_{t+1}$$

to

$$x_t = \phi_1^{-1} x_{t+1} - \phi_1^{-1} w_{t+1}.$$

Next, apply a familiar recursive argument to see that

$$\begin{aligned}
x_t &= \phi_1^{-1} x_{t+1} - \phi_1^{-1} w_{t+1} \\
&= \phi_1^{-1}\{\phi_1^{-1} x_{t+2} - \phi_1^{-1} w_{t+2}\} - \phi_1^{-1} w_{t+1} \\
&= \phi_1^{-2} x_{t+2} - \phi_1^{-2} w_{t+2} - \phi_1^{-1} w_{t+1} \\
&= \phi_1^{-2}\{\phi_1^{-1} x_{t+3} - \phi_1^{-1} w_{t+3}\} - \phi_1^{-2} w_{t+2} - \phi_1^{-1} w_{t+1} \\
&= \phi_1^{-3} x_{t+3} - \phi_1^{-3} w_{t+3} - \phi_1^{-2} w_{t+2} - \phi_1^{-1} w_{t+1},
\end{aligned}$$

and so on, which establishes the formula

$$x_t \;=\; \phi_1^{-k} x_{t+k} - \sum_{j=1}^{k} \phi_1^{-j} w_{t+j}.$$

Subsequently, the assumption $|\phi_1| > 1$ implies

$$\phi_1^{-k} x_{t+k} \to 0 \quad \text{as} \quad k \to 0,$$

from which the infinite-sum representation

$$x_t = -\sum_{j=1}^{\infty} \phi_1^{-j} w_{t+j}$$

follows.

The interpretation of this particular the infinite-sum representation is that x_t has a stationary representation in terms of *future* realizations of a white-noise time series. Thus, while it is possible to deduce properties of this time series, such as stationarity, the time series itself is impractical to work with (*e.g.*, for forecasting), because it depends on events that would occur in the future. This autoregressive time series would be called a *future-dependent* time series.

In contrast, when $|\phi_1| < 1$, the representation

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j}$$

is valid, and is in terms of events that would only occur in the past. This autoregressive time series would be called a *causal* time series.

The restriction of the autoregressive parameter to $|\phi_1| < 1$ is thus seen as necessary to induce in an $AR(1)$ time series two important properties: (i) stationarity and (ii) causality. □

Causality is defined as the requirement that x_t has an infinite-sum of white noise representation that is absent any future values $w_{t+1}, w_{t+2}, \ldots$. Such lack of dependence on future values captures a piece of intuition that we would expect to have available when working with a time series. Moreover, when causality holds, some mathematical derivations become more manageable. For example, the difference equation that defines the ACF of an $AR(p)$ times series becomes quite simple to derive. Consider that the difference equation

$$\gamma(h) = \phi_1\gamma(h-1) + \cdots + \phi_p\gamma(h-p),$$

follows quickly from the autoregressive relationship

$$x_{t+h} = \phi_1 x_{t+h-1} + \cdots + \phi_p x_{t+h-p} + w_{t+h},$$

upon multiplying each side by x_t and taking expectations. The result is

$$E[x_{t+h}x_t] = \phi_1 E[x_{t+h-1}x_t] + \cdots + \phi_p E[x_{t+h-p}x_t] + E[w_{t+h}x_t],$$

which is identical to the above equation, provided $E[w_{t+h}x_t] = 0$; that property follows from causality, since w_{t+h} is a future value, relative to x_t, and so $E[w_{t+h}x_t] = E[w_{t+h}]E[x_t] = 0$. Notice, here, that we have made use of the identity $Cov[x_s, x_t] = E[x_s x_t]$, which holds for zero-mean time series.

In addition, it becomes relatively straightforward to extend the set of difference equations to the case $h = 0$, which allows us to deduce the variance of the time series.

Example: Variance of an $AR(p)$ time series

Suppose (x_t) is an $AR(p)$ time series such that

$$x_t \quad = \quad \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t,$$

where $\phi_1, \ldots \phi_p$ are autoregression parameters and (w_t) is Gaussian white noise.

Observe that

$$\begin{aligned} E[x_t^2] &= E[x_t(\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t)] \\ &= \phi_1 E[x_{t-1}x_t] + \cdots + \phi_p E[x_{t-p}x_t] + E[w_t x_t] \end{aligned}$$

and that

$$\begin{aligned} E[w_t x_t] &= E[w_t(\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t)] \\ &= \phi_1 E[w_t x_{t-1}] + \cdots + \phi_p E[x_t x_{t-p}] + E[w_t^2] = \sigma_w^2 \end{aligned}$$

since $E[w_t^2] = \sigma_w^2$ and, by causality, $E[w_t x_{t-j}] = 0$ for $j = 1, \ldots, p$. Putting these observations together provides the equation

$$\gamma(0) = \phi_1 \gamma(1) + \cdots + \phi_p \gamma(p) + \sigma_w^2$$

Supposing that we have already solved the difference equations for $h = 1, 2, \ldots, p$, and so have available the ACF values $\rho(1), \ldots, \rho(p)$, the above equation would be interpreted as

$$\gamma(0) = \gamma(0)\{\phi_1 \rho(1) + \cdots + \phi_p \rho(p)\} + \sigma_w^2,$$

which ultimately provides the following formula for the variance

$$\gamma(0) = \frac{\sigma_w^2}{1 - \phi_1 \rho(1) - \cdots - \phi_p \rho(p)}.$$

In the case $p = 2$, recall that in a previous example we deduced that the ACF-values at lags $h = 1, 2$ of an $AR(2)$ time series are

$$\rho(1) = \phi_1/(1 - \phi_2) \quad \text{and} \quad \rho(2) = \phi_1^2/(1 - \phi_2) + \phi_2.$$

The variance of the time series is therefore

$$\gamma(0) = \frac{\sigma_w^2}{1 - \phi_1^2/(1 - \phi_2) - \phi_1^2 \phi_2/(1 - \phi_2) - \phi_2^2} = \frac{\sigma_w^2(1 - \phi_2)}{(1 - \phi_2)(1 - \phi_2^2) - \phi_1^2(1 + \phi_2)}.$$

$\square$

Causality also simplifies derivations of the partial autocorrelation function, as the next two examples illustrate.

Example: Partial autocorrelation of an $AR(1)$ time series

Suppose (x_t) is a causal $AR(1)$ time series such that

$$x_t = \phi_1 x_{t-1} + w_t,$$

where ϕ_1 is the autoregressive parameter and (w_t) is white noise. Let us deduce the partial autocorrelation at lag $h = 2$,

$$\phi_{22} = Corr[x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t],$$

for fitted values $\hat{x}_{t+2} = \hat{x}_t = a_1 x_{t+1}$, where a_1 solves $\rho(k) = a_1\rho(k-1)$, and is therefore

$$a_1 = \frac{\gamma(1)}{\gamma(0)} = \rho(1) = \phi_1.$$

Note that a_1 may instead be found, easily, as the minimizing value of $E[(x_{t+h} - a_1 x_{t+h-1})^2]$. To do so, first observe that

$$\begin{aligned} E[(x_{t+2} - a_1 x_{t+1})^2] &= E[x_{t+2}^2 - 2a_1 x_{t+2}x_{t+1} + a_1^2 x_{t+1}^2] \\ &= E[x_{t+2}^2] - 2a_1 E[x_{t+2}x_{t+1}] + a_1^2 E[x_{t+1}^2], \\ &= \gamma(0) - 2a_1\gamma(1) + a_1^2\gamma(0), \end{aligned}$$

and then complete the square to rewrite

$$\gamma(0) - 2a_1\gamma(1) + a_1^2\gamma(0) = \gamma(0)\left(1 - \frac{\gamma(1)^2}{\gamma(0)^2}\right) + \gamma(0)\left(a_1 - \frac{\gamma(1)}{\gamma(0)}\right)^2,$$

which is minimized at the setting above, at which the second (always non-negative) term is zero.

The partial autocorrelation is

$$\begin{aligned} \phi_{22} &= Corr[x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t] \\ &= Corr[x_{t+2} - \phi_1 x_{t+1}, x_t - \phi_1 x_{t+1}] \\ &= Corr[w_{t+2}, x_t - \phi_1 x_{t+1}] = 0, \end{aligned}$$

where the last step is implied from causality, since w_{t+2} is a future value relative to x_t and x_{t+1}. $\square$

The last example establishes that the PACF at lag $h = 2$ of an $AR(1)$ time series is zero. This reflects a general property of $AR(p)$ time series, which is that the PACF is zero at any lag h such that $h > p$. This property is carefully explored in the next example.

Example: Partial autocorrelation of an $AR(p)$ time series

Suppose (x_t) is an $AR(p)$ time series, for which

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t,$$

where $\phi_1, \ldots \phi_p$ are autoregression parameters and (w_t) is Gaussian white noise.

At lag h, the coefficients $a_1, \ldots, a_h$ of the fitted values $\hat{x}_{t+h} = a_1 x_{t+h-1} + \cdots + a_h x_{t+1}$ and $\hat{x}_t = a_1 x_{t+1} + \cdots + a_h x_{t+h-1}$ may be determined from the criterion that identifies the best linear predictor, $x_{h+1}^h = \phi_{h1}x_h + \cdots + \phi_{hh}x_1$. These coefficients must satisfy

$$\rho(k) = \phi_{h1}\rho(k-1) + \cdots + \phi_{hh}\rho(k-h),$$

for $k = 1, \ldots, h-1$.

In the case $h > p$, we are already familiar with a parallel set of equations

$$\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p).$$

These are the difference equations that we encountered in our examination of the autocorrelation function of an $AR(p)$ time series. It follows that

$$\phi_{hj} \quad = \quad \begin{cases} \phi_j & \text{if } j \leq p \\ 0 & \text{if } j > p \end{cases}$$

gives the coefficients of the fitted-value formulas. That is, $a_j = \phi_{hj}$, and the formulas are

$$\begin{aligned} \hat{x}_{t+h} &= \phi_1 x_{t+h-1} + \cdots + \phi_p x_{t+h-p} \\ \hat{x}_t &= \phi_1 x_{t+1} + \cdots + \phi_p x_{t+h-1}. \end{aligned}$$

It follows that

$$x_{t+h} - \hat{x}_{t+h} = (\hat{x}_{t+h} + w_{t+h}) - \hat{x}_{t+h} = w_{t+h},$$

and so the partial autocorrelation at lag h is

$$Corr[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t] \quad = \quad Corr[w_{t+h}, x_t - \hat{x}_t] = 0,$$

by causality, since w_{t+h} is a future value of all of the terms in

$$x_t - \hat{x}_t = x_t - (\phi_1 x_{t+1} + \cdots + \phi_p x_{t+h-1}).$$

In the case $h \leq p$, the partial autocorrelation at lag h may be calculated by the Durbin-Levinson algorithm, as the coefficient ϕ_{hh} in the best linear predictor, $x_{h+1}^h = \phi_{h1}x_h + \cdots + \phi_{hh}x_1$, after first solving the difference equations $\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p)$, for $k = 1, \ldots, h$, to determine the ACF. □

The property that the PACF of an $AR(p)$ time series is zero at lags $h > p$ and non-zero at lags $h \leq p$ motivates the sample PACF to assist in identifying a suitable value of p. Used together, the sample ACF and sample PACF may possibly be useful for distinguishing between $AR(p)$ and $MA(q)$ time series models, according to patterns summarized in the following table:

Diagnostic	$AR(p)$	$MA(q)$
ACF	decay	cutoff at lag q
PACF	cutoff at lag p	decay

Section 3A.4: Invertibility.

The remainder of this part of the learning unit explores moving-average time series in a parallel manner of the above exploration of autoregressive time series, and introduces a new concept that complements the notion of causality.

Recall that $MA(q)$ moving-average time series (x_t) satisfies the relationship

$$x_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $\theta_1, \ldots, \theta_q$ are constant parameters and (w_t) is Gaussian white noise. Alternatively, the moving-average relationship may be expressed as $x_t = \theta(B)$, where $\theta(B) = 1+\theta_1 B+\cdots+\theta_q B^q$ is the moving-average operator.

Our previous analysis of autoregressive time series began with an inquiry into conditions by which such time series are stationary. As it turns out, the stationarity of moving-average time series is much simpler to investigate:

Property: An $MA(q)$ time series is always stationary. □

This property is established by deriving the mean function and autocovariance function of the moving-average. The mean function is $\mu_t = E[x_t] = 0$ since each $E[w_t] = 0$, by which $E[x_t] = E[w_t] + \theta_1 E[w_{t-1}] + \cdots + \theta_q E[w_{t-q}] = 0$. The autocovariance function is readily derived by expressing (x_t) as a one-sided linear process, and applying the associated ACF formula. The latter derivation is collected into the following example.

Example: ACF of a $MA(q)$ time series

Consider the $MA(q)$ time series (x_t) such that

$$x_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

wherein the $\theta_1, \ldots, \theta_q$ are constant parameters and (w_t) is Gaussian white noise. This is a one-sided linear process $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$, with $\psi_0 = 1$, $\psi_j = \theta_j$ for $j = 1, \ldots, q$, and $\psi_j = 0$ for $j > q$. The autocovariance function of a one-sided linear process is

$$\gamma(t+h, t) \quad = \quad \sigma_w^2 \sum_{k=0}^{\infty} \psi_{k+h} \psi_k,$$

which establishes that (x_t) is stationary.

Going further, the autocovariance function evaluates to

$$\gamma(h) \quad = \quad \begin{cases} (1+\theta_1^2+\cdots+\theta_q^2)\sigma_w^2 & \text{if } h = 0 \\ (\theta_h + \theta_1\theta_{h+1} + \cdots + \theta_{q-h}\theta_q)\sigma_w^2 & \text{if } h = 1, \ldots, q \\ 0 & \text{if } h = q+1, q+2, \ldots \end{cases}$$

The accompanying ACF is

$$\rho(h) \quad = \quad \gamma(h)/\gamma(0) = \begin{cases} 1 & \text{if } h = 0 \\ \dfrac{\theta_h + \theta_1\theta_{h+1} + \cdots + \theta_{q-h}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2} & \text{if } h = 1, \ldots, q \\ 0 & \text{if } h = q+1, q+2, \ldots \end{cases}$$

□

Having found that the issue of stationarity is straightforward for moving-average time series, let us probe the $MA(1)$ model, which will lead us to deeper insights. We suppose the time series (x_t) has

$$x_t = w_t + \theta_1 w_{t-1}.$$

From the ACF formula derived in the example, the autocovariance and autocorrelation functions are

$$\gamma(h) = \begin{cases} (1+\theta_1^2)\sigma_w^2 & \text{if } h=0 \\ \theta_1\sigma_w^2 & \text{if } h=1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \rho(h) = \begin{cases} 1 & \text{if } h=0 \\ \dfrac{\theta_1}{1+\theta_1^2} & \text{if } h=1 \\ 0 & \text{otherwise} \end{cases}$$

Looking carefully at the autocorrelation function, observe the identity

$$\frac{\theta_1}{1+\theta_1^2} = \frac{1/\theta_1}{1+(1/\theta_1)^2}$$

This simple result implies that the exact same ACF could be produced by two different parameterizations. For example, the $\rho(h)$ specified by the setting $\theta_1 = 5$ is the same as that by $\theta_1 = 1/5$. Moreover, the phenomenon extends to the autocovariance function: observe that the $\gamma(h)$ specified by the setting $\theta_1 = 5$ and $\sigma_w^2 = 1$ is the same as that by $\theta_1 = 1/5$ and $\sigma_w^2 = 25$. In summary, we have deduced that the $MA(1)$ models

$$x_t = w_t + 5w_{t-1} \qquad w_t \sim iid\ N(0,1)$$

and

$$y_t = v_t + \frac{1}{5}v_{t-1} \qquad v_t \sim iid\ N(0,25)$$

are the same.

As it turns out, this is a rather general phenomenon in moving-average time series, and is captured by the following terminology.

Definition: A model is *non-identifiable* if it may be specified by at least two distinct parameter settings. □

We can explore non-identifiability broadly by working with the moving-average operator $\theta(B) = 1 + \theta_1 B$, and the corresponding expression $x_t = \theta(B)w_t$ of the moving-average relationship.

To this end, define the *infinite autoregressive operator* $\pi(B)$ according to

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots,$$

and suppose the time series (x_t) satisfies the relationship $\pi(B)x_t = w_t$, for unknown coefficients π_j. Now substitute $\pi(B)x_t = w_t$ into $x_t = \theta(B)w_t$ and solve for unknown $\pi(B)$:

$$x_t = \theta(B)\pi(B)x_t \quad \Leftrightarrow \quad \theta^{-1}(B)x_t = \pi(B)x_t$$

The inverse operator, $\theta^{-1}(B)$, if it exists, would be found by power series expansion, which yields

$$\pi(B) \quad = \quad \theta^{-1}(B) = \frac{1}{1+\theta_1 B} = 1 - \theta_1 B + \theta_1^2 B^2 - \theta_1^3 B^3 + \cdots = \sum_{j=0}^{\infty} (-\theta_1)^j B^j.$$

This produces the infinite autoregressive representation of (x_t) given by

$$w_t \quad = \quad \pi(B) x_t = \sum_{j=0}^{\infty} (-\theta_1)^j x_{t-j},$$

which makes sense only when $|\theta_1| < 1$.

These manipulations imply that when an infinite autoregressive representation exists it is unique. Thus, we have a suitable strategy for managing the non-identifiability issue: should there be distinct parameterizations of the same moving-average model, select the one with an autoregressive representation. In the case of the $MA(1)$ model, if the choice is between $\theta_1 = 5$ or $\theta_1 = 1/5$, select the value $\theta_1 = 1/5$ since it would satisfy the condition $|\theta_1| < 1$ that guarantees the existence of an autoregressive representation.

A corresponding terminology is that a moving-average time-series model with an autoregressive representation is said to be *invertible*.

For insight into the sensibility of the term "invertible," consider the $MA(1)$ formula

$$x_t = w_t + \theta_1 w_{t-1}$$

and rearrange in such a way that "inverts" the roles of (x_t) and (w_t):

$$w_t = -\theta_1 w_{t-1} + x_t.$$

This presents (w_t) as if it follows an $AR(1)$ autoregressive relationship with parameter $-\theta_1$ and (dependent) residual time series (x_t). Continued substitutions of the autoregressive-type formula produces

$$\begin{aligned}
w_t &= -\theta_1 w_{t-1} + x_t \\
&= -\theta_1 \{-\theta_1 w_{t-2} + x_{t-1}\} + x_t \\
&= \theta_1^2 w_{t-2} - \theta_1 x_{t-1} + x_t \\
&= \theta_1^2 \{-\theta_1 w_{t-3} + x_{t-2}\} - \theta_1 x_{t-1} + x_t \\
&= -\theta_1^3 w_{t-3} + \theta_1^2 x_{t-2} - \theta_1 x_{t-1} + x_t,
\end{aligned}$$

which after k steps is

$$w_t = (-\theta_1)^k w_{t-k} + \sum_{j=0}^{k-1} (-\theta_1)^j x_{t-j}.$$

When $|\theta_1| < 1$, this line of deduction offers another route to the previously established infinite-sum representation

$$w_t \quad = \quad \pi(B) x_t = \sum_{j=0}^{\infty} (-\theta_1)^j x_{t-j}.$$