STAT 5170: Applied Time Series

Course notes for part B of learning unit 3

Section 3B.1: ARMA models.

The insights made in part A of the learning unit into stationarity, causality, and invertibility of $AR(p)$ and $MA(q)$ models has prepared us to consider the more general class of *autoregressive moving-average* (ARMA) models:

Definition: The time series (x_t) is an $ARMA(p,q)$ autoregressive moving-average time series if

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are constant parameters and (w_t) is Gaussian white noise. $\square$

In backshift operator notation, the ARMA relationship may be expressed as

$$\phi(B)x_t = \theta(B)w_t$$

for the autoregressive operator

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

and moving-average operator

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

As the following example illustrates, ARMA models are subject to subtle issues related to parameterization:

<u>Example:</u> Parameter redundancy in ARMA models

Suppose the time series (x_t) is defined by the $ARMA(1,1)$ relationship

$$x_t = \phi_1 x_{t-1} + w_t + \theta_1 w_{t-1}$$

where (w_t) is Gaussian white noise. Suppose further ϕ_1 is the negative of $-\theta_1$, which we express by defining the common value $\zeta = \phi_1 = -\theta_1$. It follows that

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

The relevant autoregressive and moving-average operators are

$$\phi(B) = 1 - \zeta B \quad \text{and} \quad \theta(B) = 1 - \zeta B,$$

and the $ARMA(1,1)$ relationship $\phi(B)x_t = \theta(B)w_t$ is

$$(1 - \zeta B)x_t = (1 - \zeta B)w_t.$$

The crucial observation in this example is that the factor $(1 - \zeta B)$ cancels, hence the relationship reduces to

$$x_t = w_t.$$

It is therefore seen that the $ARMA(1,1)$ parameterization has disguised that (x_t) is identical to the white noise time series (w_t).

This issue in ARMA models is called *parameter redundancy*: the ARMA parameterization can sometimes hide a simpler model structure. Fortunately, parameter redundancy is easily detected by looking for common factors in the autoregressive and moving-average operators, $\phi(B)$ and $\theta(B)$. Consider the following example, which is a little more complicated than the previous one:

Suppose the time series (x_t) is defined by the $ARMA(2,2)$ relationship

$$x_t = 0.40x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

where (w_t) is Gaussian white noise. Rewritten, this is

$$x_t - 0.40x_{t-1} - 0.45x_{t-2} = w_t + w_{t-1} + 0.25w_{t-2},$$

or, using backshift-operator notation,

$$(1 - 0.40B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t.$$

The latter identifies the relevant operators

$$\phi(B) = 1 - 0.40B - 0.45B^2 \quad \text{and} \quad \theta(B) = 1 + B + 0.25B^2.$$

In the manipulations that follow, we treat the operators $\phi(B)$ and $\theta(B)$ as polynomials in B, which may seem a bit strange since B is an abstract object, rather than a member of a familiar set of numbers. As was suggested when we worked with inverse operators and power-series expansions of $\phi(B)$ and $\theta(B)$, these operators can be manipulated *as if* they define polynomials of a more familiar type. In what follows, all of our manipulations will make sense when we treat the argument in any of these operators as a *complex number*. When convenient, this is made explicit by substituting z for B in the notation. For example, by this notation, the specific operators that are presently relevant are $\phi(z) = 1 - 0.40z - 0.45z^2$ and $\theta(z) = 1 + z + 0.25z^2$, where z is regarded as a complex number. In general, autoregressive and moving-average operators are sometimes written

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

and referred to as *autoregressive and moving-average polynomials*, respectively.

Returning to the problem at hand, let us factor the relevant polynomials to deduce the formulas

$$\begin{aligned} \phi(z) &= 1 - 0.40z - 0.45z^2 = (1 + 0.5z)(1 - 0.9z) \\ \theta(z) &= 1 + z + 0.25z^2 = (1 + 0.5z)(1 + 0.5z). \end{aligned}$$

Observe that $\phi(z)$ and $\theta(z)$ have a *common factor*, $(1 + 0.5z)$. By allowing this factor to cancel in the ARMA relationship, one is led to a simpler parameterization of the model. To see this, rewrite $\phi(B)x_t = \theta(B)w_t$ in terms of the factored polynomials,

$$(1 + 0.5B)(1 - 0.9B)x_t = (1 + 0.5B)(1 + 0.5B)w_t;$$

then cancel the common factor to yield the simpler parameterization

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t.$$

This shows that the $ARMA(2, 2)$ parameterization has disguised that (x_t) may be defined through a simpler $ARMA(1, 1)$ relationship, specifically

$$x_t = 0.90x_{t-1} + w_t + 0.5w_{t-1}.$$

In general, to avoid the issue of parameter redundancy, it is typically required that the autoregressive and moving-average polynomials, $\phi(z)$ and $\theta(z)$, have no common factors. □

In order to avoid the parameter-redundancy issue illustrated in the above example, we make use of the concepts we had separately previously developed for autoregressive and moving-average models. The following examples summarize these concepts within the ARMA context:

Definition: The $ARMA(p, q)$ time series (x_t) is *causal* if it has an $MA(\infty)$ representation

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

such that $\sum_{j=1}^{\infty} |\psi_j| < \infty$. It is *invertible* if it has an $AR(\infty)$ representation

$$x_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t$$

such that $\sum_{j=1}^{\infty} |\pi_j| < \infty$. □

The idea here is that if an ARMA model is both causal and invertible, then its parameterizations in terms of the infinite moving-average operator $\psi(B)$, by which $x_t = \psi(B)w_t$, and the infinite autoregressive operator $\pi(B)$, by which $\pi(B)x_t = w_t$, are each *unique*. The two infinite operators are

$$\psi(B) \;\; = \;\; 1 + \psi_1 B + \psi_2 B^2 + \cdots$$

and

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

Conceptually, these representations are available to avoid any problems having to do with parameter redundancy.

The following property indicates how working with the autoregressive and moving-average polynomials can help to determine an ARMA model's status in terms of causality and invertibility:

Property: Suppose (x_t) is an $ARMA(p, q)$ time series. It follows that...

(*i.*) (x_t) is causal if and only if $\phi(z) \neq 0$ for every $|z| \leq 1$, and

(*ii.*) (x_t) is invertible if and only if $\theta(z) \neq 0$ for every $|z| \leq 1$ □

When (x_t) is causal, the power-series expansion of $\theta(z)/\phi(z)$ exists on $|z| \leq 1$. The coefficients of the relevant infinite moving-average polynomial $\psi(z)$ are found by equating the ψ_j with the coefficients of that expansion; this can be done by matching terms in the expansion of $\phi(z)\psi(z) = \theta(z)$. Similarly, when (x_t) is invertible, the power-series expansion of $\phi(z)/\theta(z)$ exists on $|z| \leq 1$. The coefficients of the relevant infinite autoregressive polynomial $\pi(z)$ are found by equating the π_j with the coefficients of that expansion; this can be done by matching terms in the expansion of $\theta(z)\pi(z) = \phi(z)$.

The conceptual tools alluded to by this property are demonstrated in the following two examples.

Example: Causal $AR(2)$ time series

Suppose the time series (x_t) is defined by the $AR(2)$ relationship

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t,$$

where (w_t) is Gaussian white noise. The autoregressive polynomial is

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2,$$

whose two roots may be found using the quadratic formula,

$$z = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

Denote by z_1 and z_2 the two roots of $\phi(z)$. It follows that autoregressive polynomial factors according to

$$\phi(z) = \left(1 - \frac{1}{z_1} z\right)\left(1 - \frac{1}{z_2} z\right).$$

Expanding, the original form $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ is recovered in terms of z_1 and z_2 according to

$$\phi(z) = 1 - \left(\frac{1}{z_1} + \frac{1}{z_2}\right) z + \frac{1}{z_1 z_2} z^2,$$

in which the parameters ϕ_1 and ϕ_2 are identified as

$$\phi_1 = \frac{1}{z_1} + \frac{1}{z_2} \quad \text{and} \quad \phi_2 = -\frac{1}{z_1 z_2}.$$

By separately considering all possible cases in which $|z_1| > 1$ and $|z_2| > 1$ (for which z_1 and z_2 may be real and distinct, real and equal, or a complex conjugate pair), it is possible to deduce that the time series is causal if and only if the following three conditions are satisfied:

$$\phi_1 + \phi_2 < 1, \phi_1 - \phi_2 > -1, \text{ and } |\phi_2| < 1.$$

$\square$

<u>Example:</u> Causal and invertible $ARMA(1,1)$ time series

Suppose the time series (x_t) is defined by the $ARMA(1,1)$ relationship

$$x_t = 0.90x_{t-1} + w_t + 0.5w_{t-1}.$$

where (w_t) is Gaussian white noise. (Note this example picks up from where the last example left off.) In this case, the AR and MA polynomials are

$$\phi(z) = 1 - 0.9z \quad \text{and} \quad \theta(z) = 1 + 0.5z.$$

Solving $\phi(z) = 0$, one sees that

$$\phi(z) = 1 - 0.9z = 0 \quad \text{if, and only if,} \quad z = 1/0.9 = 1.1111\ldots > 1,$$

from which it follows that (x_t) is causal. Similarly, by solving $\theta(z) = 0$, one sees that

$$\theta(z) = 1 + 0.5z = 0 \quad \text{if, and only if,} \quad z = -1/0.5 = -2\ldots < -1,$$

from which it follows that (x_t) is invertible.

To find the $MA(\infty)$ representation, expand $\phi(z)\psi(z) = \theta(z)$ in z and match the coefficients of $\psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \cdots$. Observe,

$$\phi(z)\psi(z) = (1 - 0.9z)(\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots) = 1 + 0.5z = \theta(z)$$

becomes

$$\psi_0 + (\psi_1 - 0.9\psi_0)z + (\psi_2 - 0.9\psi_1)z^2 + \cdots = 1 + 0.5z.$$

Hence,

$$\begin{array}{rcl} \psi_0 & = & 1 \\ \psi_1 - 0.9\psi_0 & = & 0.5 \\ \psi_2 - 0.9\psi_1 & = & 0 \\ & \vdots & \\ \psi_j - 0.9\psi_{j-1} & = & 0 \end{array} \quad \text{if, and only if,} \quad \begin{array}{rcl} \psi_0 & = & 1 \\ \psi_1 & = & 0.9 + 0.5 = 1.4 \\ \psi_2 & = & (1.4)0.9 \\ & \vdots & \\ \psi_j & = & (1.4)0.9^{j-1} \end{array}$$

The $MA(\infty)$ representation is therefore

$$\begin{aligned} x_t &= w_t + 1.4w_{t-1} + (1.4)(0.9)w_{t-2} + (1.4)(0.9)^2 w_{t-3} + \cdots \\ &= w_t + (1.4)\sum_{j=1}^{\infty}(0.9)^{j-1}w_{t-j} \end{aligned}$$

A parallel argument will establish the $AR(\infty)$ representation

$$\begin{aligned} x_t &= w_t + 1.4x_{t-1} - (1.4)(0.5)x_{t-2} + (1.4)(0.5)^2 x_{t-3} \pm \cdots \\ &= (1.4)\sum_{j=1}^{\infty}(-0.5)^{j-1}x_{t-j} + w_t. \end{aligned}$$

In general, suppose that (x_t) is $ARMA(1,1)$ t.s., with

$$x_t = \phi_1 x_{t-1} + w_t + \theta_1 w_{t-1},$$

such that $\phi_1 \neq -\theta_1$ (so the representation does not disguise white noise). The AR and MA polynomials are

$$\phi(z) = 1 - \phi_1 z \quad \text{and} \quad \theta(z) = 1 + \theta_1 z,$$

whose roots are $z = 1/\phi_1$ for $\phi(z)$ and $z = -1/\theta_1$ for $\theta(z)$. From this, it is deduced that (x_t) is causal if $|1/\phi_1| > 1$ and invertible if $|1/\theta_1| > 1$.

The restrictions $|1/\phi_1| > 1$ and $|1/\theta_1| > 1$ are equivalently stated $|\phi_1| < 1$ and $|\theta_1| < 1$.

The coefficients of $\psi(z)$ and $\pi(z)$ are unique, so even if $\phi_1 \neq -\theta_1$, indicating that the representation is a disguise of white noise, the situation is easily managed. Moreover, any parameter redundancy would become obvious upon working with the expressions $\phi(z)\psi(z) = \theta(z)$ and $\theta(z)\pi(z) = \phi(z)$. $\square$

It is possible to work out general strategies for calculating the coefficients of an $MA(\infty)$ or $AR(\infty)$ representation of an $ARMA(p,q)$ time series. Suppose the autoregressive and moving-average polynomials are, respectively,

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \theta_q z^q,$$

and adopt the notational convention that $\phi_j = 0$ if $j > p$ and $\theta_j = 0$ if $j > q$.

For the $MA(\infty)$ representation, the equation to be solved is

$$\phi(z)\psi(z) = (1 - \phi_1 z - \cdots - \phi_p z^p)(1 + \psi_1 z + \psi_2 z^2 \cdots) = 1 + \theta_1 z + \theta_q z^q = \theta(z),$$

where $\psi(z) = 1 + \psi_1 z + \psi_2 z^2 \cdots$ is the infinite moving-average polynomial whose coefficients are to be determined. Expanding $\phi(z)\psi(z)$ and keeping track of the exponents leads to the following equivalent set of recursive equations:

$$\begin{aligned} \psi_1 &= \theta_1 + \phi_1 \\ \psi_j &= \theta_j + \phi_j + \phi_1\psi_{j-1} + \cdots + \phi_{j-1}\psi_1 \end{aligned}$$

for $j = 2, \ldots, p$, and

$$\psi_{p+k} \;=\; \theta_{p+k} + \phi_1 \psi_{p+k-1} + \cdots + \phi_p \psi_k$$

for $k = 1, 2, \ldots$ (hence $j = p + k > p$).

For the $AR(\infty)$ representation, the equation to be solved is

$$\theta(z)\pi(z) = (1 + \theta_1 z + \theta_q z^q)(1 - \pi_1 z - \pi_2 z^2 \cdots) = 1 - \phi_1 z - \cdots - \phi_p z^p = \phi(z),$$

where $\pi(z) = 1 - \pi_1 z - \pi_2 z^2 \cdots\cdots$ is the infinite autoregressive polynomial whose coefficients are to be determined. The equivalent set of recursive equations are as follows:

$$\begin{aligned} \pi_1 &= \phi_1 + \theta_1 \\ \pi_j &= \phi_j + \theta_j - \theta_1 \pi_{j-1} - \cdots - \theta_{j-1} \pi_1 \end{aligned}$$

for $j = 2, \ldots, q$, and

$$\pi_{q+k} \;=\; \phi_{q+k} - \theta_1 \pi_{q+k-1} - \cdots - \theta_q \pi_k$$

for $k = 1, 2, \ldots$ (hence $j = q + k > q$).

We have not discussed the ACF and PACF of ARMA models. Nevertheless, given the tools we have developed we would know how to formulate these diagnostics from the infinite autoregressive of moving-average representations of an ARMA time series. Recall from previous discussion that the sample ACF and PACF is sometimes useful as a diagnostic tool for selecting the form of a model. We specifically noted that the ACF may be helpful in identifying the q of a $MA(q)$ time series, and the PACF may be helpful in identifying the p of a $AR(p)$ time series. The following expands a previous table summarizing the relevant patterns by adding a column for ARMA models.

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

As indicated, though these diagnostics may be helpful in distinguishing an ARMA model from an autoregressive or moving-average model, identifying p and q is not as easy a task as in the simpler models.

Section 3B.2: Forecasting.

In our discussion of linear prediction, the relevant prediction formulas were derived within a framework that evaluates prediction performance according to mean squared prediction error. The same criterion is convenient for deriving prediction formulas under moving-average, autoregressive, and ARMA models. Recall its definition:

Definition: If x_{n+m}^n is a prediction of x_{n+m}, its *mean squared prediction error* is

$$P_{n+m}^n \;=\; E[(x_{n+m} - x_{n+m}^n)^2],$$

To be clear, the expectation operator $E[\cdot]$ is to be understood with respect to fixed values of all parameters. $\square$

One reason mean squared prediction error is so convenient is that it supports an intuitive set of formulas that results as the solution to an optimization problem. Specifically,

Property: The conditional expectation

$$x_{n+m}^n \quad = \quad E[x_{n+m}|x_1, \ldots, x_n]$$

minimizes mean squared prediction error. $\square$

The proof of this property is straightforward, as shown below.

Proof: Write $\boldsymbol{x} = (x_1, \ldots, x_n)$, and observe that

$$P_{n+m}^n \quad = \quad E[(x_{n+m} - x_{n+m}^n)^2] = E[E[\{x_{n+m} - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}]].$$

in which the notation $x_{n+m}^n = x_{n+m}^n(\boldsymbol{x})$ emphasizes the dependency of x_{n+m}^n on $\boldsymbol{x}$. This shows that the best prediction is found by minimizing

$$E[\{x_{n+m} - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}].$$

Next, observe

$$\begin{aligned}
&E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}] + E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}] \\
&\quad = \quad E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}]\}^2|\boldsymbol{x}] + E[\{E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}] \\
&\qquad\quad +2E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}]\}\{E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}|\boldsymbol{x}].
\end{aligned}$$

The third term is zero, since

$$\begin{aligned}
&E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}]\}\{E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}|\boldsymbol{x}] \\
&\quad = \quad E[x_{n+m} - E[x_{n+m}|\boldsymbol{x}]|\boldsymbol{x}]E[E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})|\boldsymbol{x}]
\end{aligned}$$

and

$$E[x_{n+m} - E[x_{n+m}|\boldsymbol{x}]|\boldsymbol{x}] = E[x_{n+m}|\boldsymbol{x}] - E[x_{n+m}|\boldsymbol{x}] = 0.$$

Therefore,

$$\begin{aligned}
&E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}] + E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}] \\
&\quad = \quad E[\{x_{n+m} - E[x_{n+m}|\boldsymbol{x}]\}^2|\boldsymbol{x}] + E[\{E[x_{n+m}|\boldsymbol{x}] - x_{n+m}^n(\boldsymbol{x})\}^2|\boldsymbol{x}].
\end{aligned}$$

Only the second term depends on $x_{n+m}^n(\boldsymbol{x})$. It is positive, and zero when

$$x_{n+m}^n(\boldsymbol{x}) \quad = \quad E[x_{n+m}|\boldsymbol{x}].$$

$\square$

The examples, below, offer several demonstrations of how prediction formula might be applied. Note in all of the examples the parameters have known values. In practice, these values would be replaced by estimated values, which are determined by techniques yet to be discussed.

Example: $AR(p)$ **forecasting**

Suppose (x_t) is a causal $AR(2)$ time series such that

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t,$$

where (w_t) is Gaussian white noise. Suppose further that a finite set of values $x_1, \ldots, x_n$ are is available from this time series, and that $n \geq 2$.

When $m = 1$, the prediction $x^n_{n+m} = x^n_{n+1}$ is typically called a *one-step ahead* prediction. Its formula is

$$\begin{aligned} x^n_{n+1} &= E[x_{n+1}|x_1, \ldots, x_n] \\ &= E[\phi_1 x_{(n+1)-1} + \phi_2 x_{(n+1)-2} + w_{n+1}|x_1, \ldots, x_n] \\ &= E[\phi_1 x_n + \phi_2 x_{n-1} + w_{n+1}|x_1, \ldots, x_n] \\ &= \phi_1 x_n + \phi_2 x_{n-1} + E[w_{n+1}] \\ &= \phi_1 x_n + \phi_2 x_{n-1}. \end{aligned}$$

For $m = 2$, write

$$\begin{aligned} x_{n+2} &= \phi_1 x_{n+1} + \phi_2 x_n + w_{n+2} \\ &= \phi_1\{\phi_1 x_n + \phi_2 x_{n-1} + w_{n+1}\} + \phi_2 x_n + w_{n+2} \\ &= (\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1} + \phi_1 w_{n+1} + w_{n+2}. \end{aligned}$$

and observe that the *two-step ahead* prediction formula is

$$\begin{aligned} x^n_{n+2} &= E(x_{n+2}|x_1, \ldots, x_n) \\ &= E[(\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1} + \phi_1 w_{n+1} + w_{n+2}|x_1, \ldots, x_n] \\ &= (\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1} + \phi_1 E(w_{n+1}) + E(w_{n+2}) \\ &= (\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1} \end{aligned}$$

By comparing formulas for x^n_{n+2} and x^n_{n+1}, and tracing the derivation above, it is readily seen that the two-step ahead prediction is

$$x^n_{n+2} = (\phi_1^2 + \phi_2) x_n + \phi_1 \phi_2 x_{n-1} = \phi_1\{\phi_1 x_n + \phi_2 x_{n-1}\} + \phi_2 x_n = \phi_1 x^n_{n+1} + \phi_2 x_n.$$

Such deduction extends to higher-step predictions, and provides a general recursive formula

$$x^n_{n+m} = \phi_1 x^n_{n+m-1} + \phi_2 x^n_{n+m-2},$$

where x^n_{n+m-j} is to be interpreted as $x^n_{n+m-j} = x_{n+m-j}$ when $j \geq m$.

More generally, when (x_t) is an $AR(p)$, such that

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

and $n \geq p$, it can be shown that the prediction formula is

$$x_{n+m}^n \;=\; \phi_1 x_{n+m-1}^n + \cdots + \phi_p x_{n+m-p}^n,$$

interpreting $x_{n+m-j}^n = x_{n+m-j}$ when $j \geq m$. □

Example: Innovations of an autoregressive model

A specialized use of forecasting in the context of an autoregressive time-series is to recover a portion of its underlying white-noise series. Suppose the $AR(p)$ time series (x_t) is observed at values $x_1, \ldots, x_n$; the autoregressive relationship is

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

where (w_t) is Gaussian white noise.

An *innovation* is the difference between an observed data value and its prediction derived from previous data values. At time t, this is

$$e_t = x_t - x_t^{t-1}.$$

For $t = p+1, \ldots, n$, the prediction formula given in the previous example reduces to

$$x_t^{t-1} \;=\; \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p}.$$

Comparing this formula to the definition of the autoregressive relationship, the innovations are quickly seen to simplify to

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

In other words, calculating the innovations of an autoregressive time-series yields the white-noise time series $w_{p+1}, \ldots, w_n$. Such use of forecasting can be useful for model checking, as will be demonstrated in later examples.

Example: $ARMA(p,q)$ **forecasting**

Suppose (x_t) is a causal and invertible $ARMA(p,q)$ time series such that

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are constant parameters and (w_t) is Gaussian white noise.

Invertibility implies the $AR(\infty)$ representation

$$w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}, \quad \text{or, equivalently,} \quad x_t = -\sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t,$$

such that $\pi_0 = 1$ and $\sum_{j=0}^{\infty} |\pi_j| < \infty$. This sets up a parallel deduction as in the previous example, which yields the prediction formula

$$x_{n+m}^n \;=\; -\sum_{j=1}^{\infty} \pi_j x_{n+m-j}^n,$$

This would be calculated first for $m = 1$, then for successive values $m = 2, 3, \ldots$

Because only a finite set of data is available, $x_1, \ldots, x_n$, the infinite sum in this formula may be impossible to calculate. An *ad hoc* modification, which feasible for use in practice, is to substitute the *truncated* sum

$$\tilde{x}^n_{n+m} \quad = \quad - \sum_{j=1}^{n+m-1} \pi_j x^n_{n+m-j}.$$

As we have seen, such substitution is irrelevant when the time series is $AR(p)$ and $n \geq p$, for in that case $\pi_j = -\phi_j$ if $j = 1, \ldots, p$ and $\pi_j = 0$ if $j > p$, and the infinite sum reduces to the truncated sum. In other cases, the π_j will decay to zero exponentially fast (*i.e.* very fast), and the truncated sum will approximate the infinite sum, provided that n is suitably large. □

Example: Innovations of an $ARMA(p, q)$ model

Consider the causal and invertible $ARMA(p, q)$ model of the previous example,

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q},$$

with the $AR(\infty)$ representation

$$w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}, \quad \text{or, equivalently,} \quad x_t = -\sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t$$

and one-step-ahead prediction formulas

$$x^n_{n+1} = -\sum_{j=1}^{\infty} \pi_j x^n_{n+1-j} \quad \text{and} \quad \tilde{x}^n_{n+1} = -\sum_{j=1}^{n} \pi_j x^n_{n+1-j},$$

wherein the second version is truncated to account for the finite number of available data points.

Innovations are defined from the unmodified prediction formula according to

$$e_t = x_t - x^{t-1}_t \quad \text{where} \quad x^{t-1}_t = -\sum_{j=1}^{\infty} \pi_j x_{t-j},$$

and by the truncated prediction formula according to

$$\tilde{e}_t = x_t - \tilde{x}^{t-1}_t \quad \text{where} \quad \tilde{x}^{t-1}_t = -\sum_{j=1}^{t-1} \pi_j x_{t-j}.$$

In the former version, the innovation reduces to

$$e_t = x_t - x^{t-1}_t = \left(-\sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t \right) - \left(-\sum_{j=1}^{\infty} \pi_j x_{t-j} \right) = w_t,$$

which implies that, if this calculation were possible, the innovations would recover a portion of the underlying white-noise time series. In the truncated version, the innovation is

$$\tilde{e}_t = x_t - \tilde{x}_t^{t-1} = \left(-\sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t \right) - \left(-\sum_{j=1}^{t-1} \pi_j x_{t-j} \right) = w_t - \sum_{j=t}^{\infty} \pi_j x_{t-j}.$$

The presence of the second term on the right side of this equation implies that the truncated innovations may not *exactly* recover a portion of the underlying white-noise time series. However, if t is large, then that second term is likely to be small, because the π_j will decay to zero exponentially fast, and the innovations would approximate the white-noise series. For example, if n is fairly large, and p and q are fairly small, a strategy for recovering a workable approximation to white noise is to fix a minimum time point t_0 well below n and well above the maximum of p and q, and calculate $\tilde{e}_t$ across $t = t_0, \ldots, n$, which provides an approximation to $w_{t_0}, \ldots, w_n$. □

Example: $ARMA(1,1)$ **forecasting**

When forecasting under simple ARMA models, it is sometimes easy to derive an exact prediction formula, rather than apply a prediction formula based on a truncated $AR(\infty)$ representation.

Suppose (x_t) is a causal and invertible $ARMA(1,1)$ time series such that

$$x_t = \phi_1 x_{t-1} + w_t + \theta_1 w_{t-1}$$

where ϕ_1 and θ_1 are constant parameters and (w_t) is Gaussian white noise. The evolution of the time series is such that w_{n+1} is a future error, which is independent of the data, $x_1, \ldots, x_n$, while invertibility implies that w_n may depend on the $x_1, \ldots, x_n$.

A forecast for w_n is

$$w_n^n = E[w_n | x_1, \ldots, x_n],$$

which is calculated recursively as follows: Set $w_0^n = 0$ and $x_0 = 0$. Then solve the $ARMA(1,1)$ relationship to deduce

$$w_k^n = x_k - \phi_1 x_{k-1} - \theta_1 w_{k-1}^n,$$

and successively calculate at $k = 1, \ldots, n$.

The one-step ahead prediction formula is then

$$\begin{aligned} x_{n+1}^n &= E[x_{n+1} | x_1, \ldots, x_n] \\ &= E[\phi_1 x_n + w_{n+1} + \theta_1 w_n | x_1, \ldots, x_n] \\ &= \phi_1 x_n + \theta_1 w_n^n. \end{aligned}$$

For higher-step predictions, the formula is

$$x_{n+m}^n = \phi_1 x_{n+m-1}^n + \theta_1 w_{n+m-1}^n.$$

where $w_{n+k}^n = 0$ for $k > 1$.

These formulas are to be applied in sequence, $m = 1, 2, \ldots$ $\square$

Another property that is sometimes useful to consider is the following.

Property: Suppose (x_t) is a causal and invertible $ARMA(p, q)$ time series, such that

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

and $\sigma_w^2 = Var(w_t)$. If $x_{n+m}^n = E[x_{n+m}|x_1, \ldots, x_n]$ then the m-step-ahead mean squared prediction error is

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2.$$

$\square$

The proof of this property is straightforward:

Proof: Set

$$w_{n+k}^n = E[w_{n+k}|x_1, \ldots, x_n],$$

For $k \geq 1$, causality implies that w_{n+k} is independent of $x_1, \ldots, x_n$, hence $w_{n+k}^n = E[w_{n+k}] = 0$. For $k \leq 0$, invertibility implies that $w_1, \ldots, w_n$ is a function of $x_1, \ldots, x_n$, hence $w_{n+k}^n = w_{n+k}$. Subsequently,

$$x_{n+m}^n = \sum_{j=0}^{\infty} \psi_j w_{n+m-j}^n = \sum_{j=m}^{\infty} \psi_j w_{n+m-j},$$

and

$$\begin{aligned}
P_{n+m}^n &= E[(x_{n+m} - x_{n+m}^n)^2] \\
&= E\left[\left(\sum_{j=0}^{\infty} \psi_j w_{n+m-j} - \sum_{j=m}^{\infty} \psi_j w_{n+m-j}\right)^2\right] \\
&= E\left[\left(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\right)^2\right] \\
&= E\left[\sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \psi_j \psi_k w_{n+m-j} w_{n+m-k}\right] \\
&= \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \psi_j \psi_k E[w_{n+m-j} w_{n+m-k}] \\
&= \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2.
\end{aligned}$$

$\square$

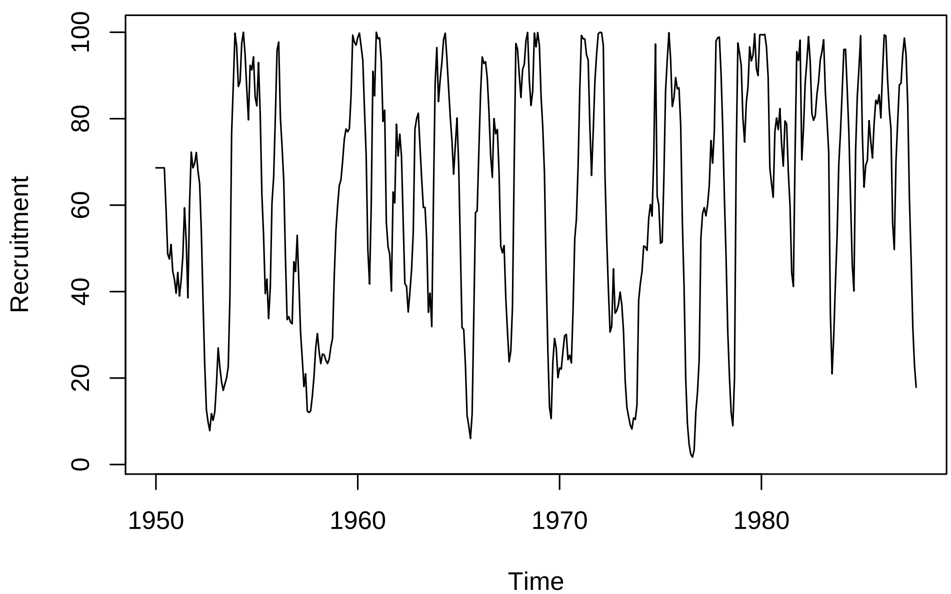
Related to this property is that the interval having endpoints

$$x_{n+m}^n \pm z_{\alpha/2}\sqrt{P_{n+m}^n}$$

is often approximately valid in practice as a $(1-\alpha)100\%$ prediction interval, at least when (x_t) is a causal and invertible $ARMA(p,q)$ time series, $x_{n+m}^n = E[x_{n+m}|x_1,\ldots,x_n]$, and n is large.

Example: Fish population count data

Recall the time series of $n = 453$ monthly fish population counts in the Pacific Ocean from 1950-1988, which is plotted below.



One plausible model for these data is $AR(2)$ with autoregressive parameters $\phi_1 = 1.35$ and $\phi_2 = -0.46$, and an added constant term of 6.80. The model is

$$x_t \quad = \quad 6.80 + 1.35x_{t-1} - 0.46x_{t-2} + w_t,$$

where (w_t) is Gaussian white noise. A workable value of the white-noise variance is $\sigma_w^2 = 89.33$.

In an $AR(2)$ model, the $MA(\infty)$ representation is deduced by substituting $w_t = \phi(B)x_t$ into $x_t = \psi(B)w_t$, resulting in $x_t = \psi(B)\phi(B)x_t$, to show that the coefficients of $\psi(z) = 1 + \psi_1 z + \psi_2 z^2 + \cdots$ are found by expanding

$$\begin{aligned}
1 &= (1 + \psi_1 z + \psi_2 z^2 + \cdots)(1 - \phi_1 z - \phi_2 z^2) \\
&= 1 + (\psi_1 - \phi_1)z + (\psi_2 - \phi_1\psi_1 - \phi_2)z^2 + (\psi_3 - \phi_1\psi_2 - \phi_2\psi_1)z^3 \\
&\quad +(\psi_4 - \phi_1\psi_3 - \phi_2\psi_2)z^4 + (\psi_5 - \phi_1\psi_4 - \phi_2\psi_3)z^5 + (\psi_6 - \phi_1\psi_5 - \phi_2\psi_4)z^6 \cdots
\end{aligned}$$

The relevant system of equations becomes

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2}$$

for $j \geq 2$, having set $\psi_0 = 0$ and $\psi_1 = \phi_1$.

By the formula

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2,$$

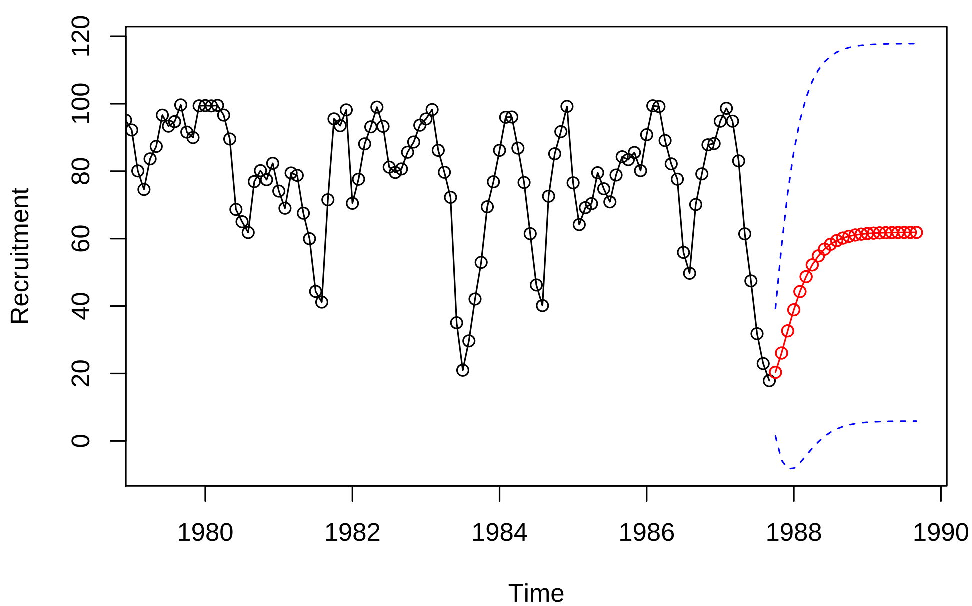
the first few prediction errors for the model at hand are

$$\begin{aligned}
P_{n+1}^n &= \sigma_w^2 = 89.33 \\
P_{n+2}^n &= \sigma_w^2(1+\phi_1^2) = 89.33(1+1.35^2) = 252.13 \\
P_{n+3}^n &= \sigma_w^2\{1+\phi_1^2+(\phi_1^2+\phi_2)^2\} = 89.33\{1+1.35^2+(1.35^2-0.46)^2\} = 417.97
\end{aligned}$$

A plot of forecast values, overlaid with 95% prediction errors given by the formula

$$x_{n+m}^n \pm 1.96\sqrt{P_{n+m}^n}$$

is displayed as follows



At least two interesting properties of the forecast are evident in this plot. One is that the forecast values ultimately level off to the presumed mean of the process. Another is that the width of prediction intervals start off small, at one-step ahead predictions, then grow larger and eventually remain at a steady width. □

The following algorithm admits the simultaneous calculation of predictions and the corresponding mean squared-prediction-errors from the covariance function of a time series.

Example: The Innovations Algorithm

Suppose that (x_t) is a mean-zero, stationary time series with covariance function $\gamma(h)$. Start by defining the initial values $x_1^0 = 0$ and $P_1^0 = \gamma(0)$, and then calculate iteratively as $t = 1, 2, \ldots$

$$\begin{aligned} P_{t+1}^t &= \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 P_{j+1}^j \\ x_{t+1}^t &= \sum_{j=1}^{t} \theta_{tj}(x_{t-j+1} - x_{t-j+1}^{t-j}) \end{aligned}$$

along the way, calculating $\theta_{tt} = \gamma(t)/\gamma(0)$ and

$$\theta_{t,t-j} = \left\{ \gamma(t-j) - \sum_{k=0}^{j-1} \theta_{j,j-k}\theta_{t,t-k}P_{k+1}^k \right\} / P_{j+1}^j,$$

for $j = 1, \ldots, t-1$. For a set of finite measurements, $x_1, \ldots, x_n$, the algorithm would stop at $t = n$. Predictions and mean squared prediction errors at further steps ahead are calculated iteratively according to

$$\begin{aligned} P_{n+m}^n &= \gamma(0) - \sum_{j=1}^{t-1} \theta_{n+m-1,j}^2 P_{n+m-j}^{n+m-j-1} \\ x_{n+m}^n &= \sum_{j=m}^{n+m-1} \theta_{n+m-1,j}\left(x_{n+m-j} - x_{n+m-j}^{n+m-j-1}\right), \end{aligned}$$

while iteratively calculating by $\theta_{n+m-1,j}$ by the same formula as before.

As a demonstration of the Innovations Algorithm, suppose (x_t) is an $MA(1)$ moving-average time series defined by the relationship $x_t = w_t + \theta w_{t-1}$, where (w_t) is Gaussian white noise. From previous discussion, we know that the corresponding autocovariance function is

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2 & \text{if } h = 0 \\ \theta\sigma_w^2 & \text{if } h = 1 \\ 0 & \text{otherwise} \end{cases}$$

Observe that the property $\gamma(h) = 0$ for $h > 1$ implies that $\theta_{t,t-j} = 0$ for $j < t-1$. To see this, start by deducing $\theta_{11} = \gamma(1)/\gamma(0)$, $\theta_{22} = \gamma(2)/\gamma(0) = 0$, and $\theta_{21} = \{\gamma(1) - \theta_{11}\theta_{22}P_1^0\}/P_2^1 = \gamma(1)/P_2^1$; then note how this pattern extends since, when $j < t-1$, terms $\gamma(t-j)$ and $\theta_{t,t-k}$ are zero in the formula for $\theta_{t,t-j}$. This observation greatly simplifies the algorithm's required calculations. Subsequently, starting with the values $x_1^0 = 0$ and $P_1^0 = \gamma(0)$, it is now straightforward to deduce

$$\begin{aligned} \theta_{t1} &= \gamma(1)/P_t^{t-1} \\ P_{t+1}^t &= \gamma(0) - \theta_{t1}^2 P_t^{t-1} \\ x_{t+1}^t &= \theta_{t1}(x_t - x_t^{t-1}). \end{aligned}$$

□