The ARIMA Model

The ARIMA model combines three types of time series models.

1. The **autoregressive model** relates each observation in a series to the observations that immediately precede it. For example, an autoregressive model with two terms, or AR(2), is:

   $$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} = \varepsilon_t$$

   where φ_1 and φ_2 are parameters and ε_t is a random error.

2. The **integrated model** is often referred to as a **random walk model**. For example:

   $$X_t - X_{t-1} = \varepsilon_t$$

   The value $X_t - X_{t-1}$ is the change or **difference** in the observations.

3. The **moving average model** relates to a moving average of random errors. For example, a moving average model with two terms, or MA(2), is:

   $$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

   where θ_1 and θ_2 are parameters and ε_t, ε_{t-1}, and ε_{t-2} are random errors.

The ARIMA model can combine any of these three types of terms.

To account for seasonality, the ARIMA model can be expanded to include similar parameters at seasonal frequency—that is, every four quarters for quarterly data, or every 12 months for monthly data. This type of model is known as a **seasonal ARIMA model**. For example, a seasonal autoregressive model for monthly data with one term can be expressed (using capitalized Greek letters to denote the seasonal parameters) as:

$$X_t - \Phi_1 X_{t-12} = \varepsilon_t$$

Similarly, a seasonal moving average model with one term is:

$$X_t = \varepsilon_t + \Theta_1 \varepsilon_{t-12}$$

The seasonal and non-seasonal components can be combined.

We will now present notation to describe the ARIMA model mathematically. First, we will introduce a special mathematical operator called the **lag operator**. The lag operator, *B*, operates on an observation of a time series to produce the observation for the previous period. That is, *B* is defined by:

$$X_{t-1} = BX_t$$

Note that $B(BX_t) = B^2 X_t = X_{t-2}$.

The ARIMA model is written as follows:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D X_t = \theta(B)\Theta(B^s)\varepsilon_t$$

Where:

X_t	is the original series (possibly preadjusted for outliers, calendar effects, etc.)
s	is the seasonal frequency, which is 4 for quarterly series and 12 for monthly series
$\phi(B)$	$= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the regular AR operator of order *p*
$\Phi(B^s)$	$= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$ is the seasonal AR operator of order *P*
$(1 - B)$	describes the differencing of a time series—that is, $(1 - B)X_t = (X_t - X_{t-1})$
$(1 - B)^d$	describes a time series that is differenced *d* times
$(1 - B^s)$	describes seasonal differencing of a time series—for example, $(1 - B^4)X_t = (X_t - X_{t-4})$
$\theta(B)$	$= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the regular MA operator of order *q*
$\Theta(B^s)$	$= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$ is the seasonal MA operator of order *q*
ε_t	is a random (white noise) process

Identifying an ARIMA model means determining the orders (number of parameters) for the regular and seasonal AR, integration, and MA operators—that is, determining *p, P, d, D, q,* and Q. An ARIMA model without any seasonal terms is usually denoted by ARIMA (*p,d,q*), where *p* is the number of autoregressive terms, *d* is the number of times the observations are differenced, and *q* is the number of moving average terms. The seasonal ARIMA model is generally denoted ARIMA (*p,d,q*),(*P,D,Q*)$_s$, where *p, d,* and *q* represent the number of non-seasonal autoregressive, differencing, and moving average terms; *P, D,* and *Q* represent the number of seasonal autoregressive, differencing, and moving average terms; and *s* represents the seasonal frequency (that is, 4 for quarterly data and 12 for monthly data).

While the ARIMA model is quite general, in practice the number of terms of each type is usually small (0, 1, or 2), and models often include only AR or MA terms, rather than both.

An example of a specification that is often used for economic time series is ARIMA (0,1,1),(0,1,1)$_s$. This means that the series is differenced and seasonally differenced, and there is one regular MA term and one seasonal MA term in the model (and no AR terms). This specification is sometimes referred to as the "airline model" (referencing an example based on airline passenger data that appeared in an early textbook by Box and Jenkins). This model is parsimonious, in that only two parameters need to be estimated (θ_1 and Θ_1).

The order (number of parameters) is selected by comparing values of a statistical information criterion up to a maximum order, which can be specified by the user. A series of statistical tests is used to select and validate the final model. The ARIMA order that is selected is central to the SEATS decomposition, because the seasonal and trend filters are derived from the coefficients of the estimated ARIMA model.

Users should consider the ARIMA order automatically selected by the software tool as the baseline model. Changes can be made to the baseline model when model statistics are not satisfactory and suggest that the model may be misspecified. Experience suggests that the differencing orders *d* and *D* are usually 0 or 1, though occasionally double differencing may be necessary for some series with persistent movements. In most cases, the user should avoid models with both AR and MA operators, as models that mix AR and MA operators tend to have unstable parameter estimates. The airline model $(0,1,1)(0,1,1)_s$ often works well as a default model for many types of economic time series.

To illustrate, consider that the compiler has used the automatic option of the seasonal adjustment software with the X-13ARIMA-SEATS method. The image shows the ARIMA models automatically selected, (0,1,1) (0,1,1) and (1,1,0) (0,1,1) for two different time series, besides the estimated parameters and respective t-statistics and p-value.

Example of Automatic Selection of ARIMA Model

ARIMA Model [(0,1,1) (0,1,1)]			
	Coefficients	T-Stat	P[|T|>t]
Theta (1)	0,2377	2,40	0,0181
BTheta (1)	-0,5493	-6.09	0,0000

ARIMA Model [(1,1,0) (0,1,1)]			
	Coefficients	T-Stat	P[|T|>t]
Phi (1)	-0,3140	-3,26	0,0015
BTheta (1)	-0,5349	-5,69	0,0000