ARIMA models for time series forecasting

Introduction to ARIMA: nonseasonal models

ARIMA(p,d,q) forecasting equation: ARIMA models are, in theory, the most general class of models for forecasting a time series which can be made to be "stationary" by differencing (if necessary), perhaps in conjunction with nonlinear transformations such as logging or deflating (if necessary). A random variable that is a time series is stationary if its statistical properties are all constant over time. *A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion*, i.e., its short-term random time patterns always look the same in a statistical sense. The latter condition means that its *autocorrelations* (correlations with its own prior deviations from the mean) remain constant over time, or equivalently, that its *power spectrum* remains constant over time. A random variable of this form can be viewed (as usual) as a combination of signal and noise, and the signal (if one is apparent) could be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also have a seasonal component. An ARIMA model can be viewed as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

The ARIMA forecasting equation for a stationary time series is a *linear* (i.e., regression-type) equation in which the predictors consist of *lags of the dependent* variable and/or *lags of the forecast errors*. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

If the predictors consist only of lagged values of Y, it is a pure autoregressive ("self-regressed") model, which is just a special case of a regression model and which could be fitted with standard regression software. For example, a first-order autoregressive ("AR(1)") model for Y is a simple regression model in which the independent variable is just Y lagged by one period (LAG(Y,1) in Statgraphics or Y_LAG1 in RegressIt). If some of the predictors are lags of the errors, an ARIMA model it is NOT a linear regression model, because there is no way to specify "last period's error" as an independent variable: the errors must be computed on a period-to-period basis when the model is fitted to the data. From a technical standpoint, the problem with using lagged errors as predictors is that *the model's predictions* *are not linear functions of the coefficients*, even though they are linear functions of the past data. So, coefficients in ARIMA models that include lagged errors must be estimated by *nonlinear* optimization methods ("hill-climbing") rather than by just solving a system of equations.

The acronym ARIMA stands for **Auto-Regressive Integrated Moving Average**. Lags of the stationarized series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. **Random-walk and random-trend models, autoregressive models, and exponential smoothing models are all special cases of ARIMA models.**

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.

The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which means:

If d=0: $y_t = Y_t$

If d=1: $y_t = Y_t - Y_{t-1}$

If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the *first-difference-of-the-first difference*, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.

In terms of y, the general forecasting equation is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

Here the moving average parameters (θ's) are defined so that their signs are negative in the equation, following the convention introduced by Box and Jenkins. Some authors and software (including the R programming language) define them so that they have plus signs instead. When actual numbers are plugged into the equation, there is no ambiguity, but it's important to know which convention your software uses when you are reading the output. Often the parameters are denoted there by AR(1), AR(2), …, and MA(1), MA(2), … etc..

To identify the appropriate ARIMA model for Y, you begin by determining the order of differencing (d) needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. If you stop at this point and predict that the differenced series is constant, you have merely fitted a random walk or random trend model. However, the stationarized series may still have autocorrelated errors, suggesting that some number of AR terms (p ≥ 1) and/or some number MA terms (q ≥ 1) are also needed in the forecasting equation.

The process of determining the values of p, d, and q that are best for a given time series will be discussed in later sections of the notes (whose links are at the top of this page), but a preview of some of the types of *nonseasonal* ARIMA models that are commonly encountered is given below.

ARIMA(1,0,0) = first-order autoregressive model: if the series is stationary and autocorrelated, perhaps it can be predicted as a multiple of its own previous value, plus a constant. The forecasting equation in this case is

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1}$$

...which is Y regressed on itself lagged by one period. This is an "ARIMA(1,0,0)+constant" model. If the mean of Y is zero, then the constant term would not be included.

If the slope coefficient ϕ_1 is positive and less than 1 in magnitude (it *must* be less than 1 in magnitude if Y is stationary), the model describes mean-reverting behavior in which next period's value should be predicted to be ϕ_1 times as far away from the mean as this period's value. If ϕ_1 is negative, it predicts mean-reverting behavior with alternation of signs, i.e., it also predicts that Y will be below the mean next period if it is above the mean this period.

In a *second-order* autoregressive model (ARIMA(2,0,0)), there would be a Y_{t-2} term on the right as well, and so on. Depending on the signs and magnitudes of the coefficients, an ARIMA(2,0,0) model could describe a system whose mean reversion takes place in a *sinusoidally oscillating* fashion, like the motion of a mass on a spring that is subjected to random shocks.

ARIMA(0,1,0) = random walk: If the series Y is not stationary, the simplest possible model for it is a random walk model, which can be considered as a limiting case of an AR(1) model in which the autoregressive coefficient is equal to 1, i.e., a series with infinitely slow mean reversion. The prediction equation for this model can be written as:

$$\hat{Y}_t - Y_{t-1} = \mu$$

or equivalently

$$\hat{Y}_t = \mu + Y_{t-1}$$

...where the constant term is the average period-to-period change (i.e. the long-term drift) in Y. This model could be fitted as a *no-intercept regression model* in which the first difference of Y is the dependent variable. Since it includes (only) a nonseasonal difference and a constant term, it is classified as an "ARIMA(0,1,0) model with constant." The random-walk-*without*-drift model would be an ARIMA(0,1,0) model *without* constant

---

ARIMA(1,1,0) = differenced first-order autoregressive model: If the errors of a random walk model are autocorrelated, perhaps the problem can be fixed by adding one lag of the dependent variable to the prediction equation--i.e., by regressing *the first difference of Y* on itself lagged by one period. This would yield the following prediction equation:

$$\hat{Y}_t - Y_{t-1} = \mu + \phi_1(Y_{t-1} - Y_{t-2})$$

$$\hat{Y}_t - Y_{t-1} = \mu$$

which can be rearranged to

$$\hat{Y}_t = \mu + Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2})$$

This is a first-order autoregressive model with one order of nonseasonal differencing and a constant term--i.e., an ARIMA(1,1,0) model.

ARIMA(0,1,1) without constant = simple exponential smoothing: Another strategy for correcting autocorrelated errors in a random walk model is suggested by the simple exponential smoothing model. Recall that for some nonstationary time series (e.g., ones that exhibit noisy fluctuations around a slowly-varying mean), the random walk model does not perform as well as a moving average of past values. In other words, rather than taking the most recent observation as the forecast of the next observation, it is better to use an *average* of the last few observations in order to filter out the noise and more accurately estimate the local mean. The simple exponential smoothing model uses an *exponentially weighted moving average* of past values to achieve this effect. The prediction equation for the simple exponential smoothing model can be written in a number of mathematically equivalent forms, one of which is the so-called "error correction" form, in which the previous forecast is adjusted in the direction of the error it made:

$$\hat{Y}_t = \hat{Y}_{t-1} + \alpha e_{t-1}$$

Because $e_{t-1} = Y_{t-1} - \hat{Y}_{t-1}$ by definition, this can be rewritten as:

$$\hat{Y}_t = Y_{t-1} - (1-\alpha)e_{t-1}$$

$$= Y_{t-1} - \theta_1 e_{t-1}$$

which is an ARIMA(0,1,1)-without-constant forecasting equation with $\theta_1 = 1-\alpha$. This means that you can fit a simple exponential smoothing by specifying it as an ARIMA(0,1,1) model without constant, and the estimated MA(1) coefficient corresponds to 1-minus-alpha in the SES formula. Recall that in the SES model, the *average age* of the data in the 1-period-ahead forecasts is 1/α, meaning that they will tend to lag behind trends or turning points by about 1/α periods. It follows that the average age of the data in the 1-period-ahead forecasts of an ARIMA(0,1,1)-without-constant model is 1/(1−θ_1). So, for example, if $\theta_1 = 0.8$, the average age is 5. As θ_1 approaches 1, the ARIMA(0,1,1)-without-constant model becomes a very-long-term moving average, and as θ_1 approaches 0 it becomes a random-walk-without-drift model.

What's the best way to correct for autocorrelation: adding AR terms or adding MA terms? In the previous two models discussed above, the problem of autocorrelated errors in a random walk model was fixed in two different ways: by adding a lagged value of the differenced series to the equation or adding a lagged value of the forecast error. Which approach is best? A rule-of-thumb for this situation, which will be discussed in more detail later on, is that *positive* autocorrelation is usually best treated by adding an AR term to the model and *negative* autocorrelation is usually best treated by adding an MA term. In business and economic time series, *negative* autocorrelation often arises as *an artifact of differencing*. (In general, differencing reduces positive autocorrelation and may even cause a switch from positive to negative autocorrelation.) So, the ARIMA(0,1,1) model, in which differencing is accompanied by an MA term, is more often used than an ARIMA(1,1,0) model.

ARIMA(0,1,1) with constant = simple exponential smoothing with growth: By implementing the SES model as an ARIMA model, you actually gain some flexibility. First of all, the estimated MA(1) coefficient is allowed to be *negative*: this corresponds to a smoothing factor larger than 1 in an SES model, which is usually not allowed by the SES model-fitting procedure. Second, you have the option of including a constant term in the ARIMA model if you wish, in order to estimate an average non-zero trend. The ARIMA(0,1,1) model *with* constant has the prediction equation:

$$\hat{Y}_t = \mu + Y_{t-1} - \theta_1 e_{t-1}$$

The one-period-ahead forecasts from this model are qualitatively similar to those of the SES model, except that the trajectory of the long-term forecasts is typically a sloping line (whose slope is equal to mu) rather than a horizontal line.

ARIMA(0,2,1) or (0,2,2) without constant = linear exponential smoothing: Linear exponential smoothing models are ARIMA models which use *two* nonseasonal differences in conjunction with MA terms. The second difference of a series Y is not simply the difference between Y and itself lagged by two periods, but rather it is the *first difference of the first difference*--i.e., the change-in-the-change of Y at period t. Thus, **the second difference of Y at period t is equal to** $(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$. A second difference of a discrete function is analogous to a second derivative of a continuous function: it measures the "acceleration" or "curvature" in the function at a given point in time.

The ARIMA(0,2,2) model without constant predicts that the second difference of the series equals a linear function of the last two forecast errors:

$$\hat{Y}_t - 2Y_{t-1} + Y_{t-2} = - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

which can be rearranged as:

$$\hat{Y}_t = 2\, Y_{t-1} - Y_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

where θ_1 and θ_2 are the MA(1) and MA(2) coefficients. This is a general *linear exponential smoothing model*, essentially the same as Holt's model, and Brown's model is a special case. It uses exponentially weighted moving averages to estimate both a *local level* and a *local trend* in the series. The long-term forecasts from this model converge to a straight line whose slope depends on the average trend observed toward the end of the series.

ARIMA(1,1,2) without constant = damped-trend linear exponential smoothing:

$$\hat{Y}_t = Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2}) - \theta_1 e_{t-1} - \theta_1 e_{t-1}$$

This model is illustrated in the accompanying slides on ARIMA models. It extrapolates the local trend at the end of the series but flattens it out at longer forecast horizons to introduce a note of conservatism, a practice that has empirical support. See the article on "Why the Damped Trend works" by Gardner and McKenzie and the "Golden Rule" article by Armstrong et al. for details.

It is generally advisable to stick to models in which at least one of p and q is no larger than 1, i.e., do not try to fit a model such as ARIMA(2,1,2), as this is likely to lead to overfitting and "common-factor" issues that are discussed in more detail in the notes on the mathematical structure of ARIMA models.

Spreadsheet implementation: ARIMA models such as those described above are easy to implement on a spreadsheet. The prediction equation is simply a linear equation that refers to past values of original time series and past values of the errors. Thus, you can set up an ARIMA forecasting spreadsheet by storing the data in column A, the forecasting formula in column B, and the errors (data minus forecasts) in column C. The forecasting formula in a typical cell in column B would simply be a linear expression referring to values in preceding rows of columns A and C, multiplied by the appropriate AR or MA coefficients stored in cells elsewhere on the spreadsheet.

Go to next topic: Identifying the order of differencing
