Univariate ARIMA Forecasts (Theory)

Al Nosedal
University of Toronto

April 16, 2019

All univariate forecasting methods (including ARIMA methods) are based on the same logic. First, the expected value of the time series process is calculated and, second, the expected value is extrapolated into the future. If the current time series observation is Y_t, then we are interested in predicting the values of Y_{t+1}, Y_{t+2}, . . . , Y_{t+n}. We will denote our ARIMA forecast of Y_{t+n} by $Y_t(n)$. We call $Y_t(n)$ the origin-t forecast of Y with a lead time of n observations.

As a first step in generating an estimate of $Y_t(n)$, we calculate the expected value of the Y_t process. Our calculations will be simplified considerably if we work in terms of the deviate process, y_t. Noting that the Y_t and y_t processes are related by

$$y_t = Y_t - \mu,$$

(where $\mu = E(Y_t)$) we can translate our calculations back into the Y_t process simply by adding a constant to our result.

Now there are actually two expected values of a time series process which can be used for univariate forecasts: the unconditional and the conditional process expectations.

Example: ARIMA(1,0,0)

To illustrate the differences between these two expectations, consider the ARIMA(1, 0, 0) process

$$y_t = \phi_1 y_{t-1} + w_t$$

(where $|\phi_1| < 1$), this process can be expressed as a sum of past "shocks" (Remember?)

$$y_t = w_t + \phi_1 w_{t-1} + \phi_1^2 w_{t-2} + \phi_1^3 w_{t-3} + \ldots$$

Example (cont.)

Taking the expected value of this expression,

$$E[y_t] = E[w_t] + \phi_1 E[w_{t-1}] + \phi_1^2 E[w_{t-2}] + \phi_1^3 E[w_{t-3}] + ...$$

$$E[y_t] = 0$$

and thus

$$E[Y_t] = E[y_t] + \mu = \mu.$$

Example (cont.)

Extrapolating this term into the future,

$y_t(1) = E[y_{t+1}] = 0$

$y_t(2) = E[y_{t+2}] = 0$

. . .

$y_t(n) = E[y_{t+n}] = 0$

When the unconditional expectation of the process is used as a univariate forecast, the process mean is the forecast regardless of lead time.

(The problem with forecasts based on the unconditional expectation of a process is that much valuable information is ignored.)

Example (cont.)

The conditional expectation of y_{t+1} is:

$$E[y_{t+1}|y_t, y_{t-1}, ..., y_2, y_1].$$

The conditional expectation of y_{t+1} is conditional upon the t preceding observations of the time series process.
(Recalling that $y_{t+1} = w_{t+1} + \phi_1 y_t$ and
$y_{t+1} = w_{t+1} + \phi_1 w_t + \phi_1^2 w_{t-1} + \phi_1^3 w_{t-2} + \phi_1^4 w_{t-3} + ...$)

$$E[y_{t+1}|y_t, y_{t-1}, ..., y_2, y_1] = E[w_{t+1}] + \phi_1 y_t = \phi_1 y_t.$$

Example (cont.)

Conditional expectation forecasts of the ARIMA(1,0,0) process are, then,
$y_t(1) = E[w_{t+1} + \phi_1 y_t] = \phi_1 y_t$

. . .

$y_t(n) = E[w_{t+n} + \phi_1 w_{t+n-1} + ... + \phi_1^{n-1} w_{t+1} + \phi_1^n y_t] = \phi_1 y_t = \phi_1^n y_t.$

It should be intuitively plausible that the "best" forecast of a time series process is the **conditional expectation** of the process. What we mean by "best" in this context is that the conditional expectation forecast has the lowest possible mean-square forecast error (MSFE) of any expectation-based forecast.

Example (cont.)

Using this conditional expectation as a forecast of y_{t+1}, the error in forecasting is:

$$e_{t+1} = y_{t+1} - y_t(1) = w_{t+1}.$$

This error will always be equal to the "random shock", w_{t+1}, and the **forecast variance** is thus

$$VAR(1) = E[w_{t+1}^2] = \sigma_w^2,$$

which is the variance of the white noise process.

Example (cont.)

A 95% interval forecast of y_{t+1} is thus

$$y_t(1) \pm 1.96\sigma_w.$$

(We expect y_{t+1} to lie in this interval 95% of the time).

Example (cont.)

If we now wish to forecast the next value of the process, we begin with the expression for y_{t+2}.

$$y_{t+2} = w_{t+2} + \phi_1 y_{t+1}$$

$$y_{t+2} = w_{t+2} + \phi_1 w_{t+1} + \phi_1^2 y_t = w_{t+2} + \phi_1 w_{t+1} + y_t(2).$$

Example (cont.)

The error in forecasting is:

$$e_{t+2} = y_{t+2} - y_t(2) = w_{t+2} + \phi_1 w_{t+1}.$$

The **forecast variance** is thus

$$VAR(2) = E[(w_{t+2} + \phi_1 w_{t+1})^2] = \sigma_w^2(1 + \phi_1^2).$$

Example (cont.)

A 95% interval forecast of y_{t+2} is thus

$$y_t(2) \pm 1.96\sqrt{\sigma_w^2(1+\phi_1^2)}.$$

(We expect y_{t+2} to lie in this interval 95% of the time).

Example (cont.)

It can be shown that
$VAR(1) = \sigma_w^2$
$VAR(2) = (1 + \phi_1^2)\sigma_w^2$
. . .
$VAR(n) = (1 + \phi_1^2 + ... + \phi_1^{2n-2})\sigma_w^2$.
Noting that the expression for $VAR(n)$ is a geometric progression, forecast variance approaches a limit of

$$lim_{n\to\infty} VAR(n) = \frac{\sigma_w^2}{1 - \phi_1^2}$$

which is the variance of the y_t autoregressive process.

Example: Moving averages

An ARIMA(0, 0, 1) process written as

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

Example (cont.)

Point forecasts for the ARIMA(0, 0, 1) are:

$y_t(1) = \theta_1 w_t$

$y_t(2) = 0$

. . .

$y_t(n) = 0$

Example (cont.)

Forecast variance is given by:
$VAR(1) = \sigma_w^2$
$VAR(2) = (1 + \theta_1^2)\sigma_w^2$
. . .
$VAR(n) = (1 + \theta_1^2)\sigma_w^2$

Example (cont.)

After the second step into the future, forecast variance is constant. The limit of $VAR(n)$ is thus

$$lim_{n \to \infty} VAR(n) = (1 + \theta_1^2)\sigma_w^2,$$

which is the variance of the ARIMA(0, 0, 1) process.

Example: Integrated Processes

An ARIMA(0, 1, 0) process, or random walk, written as

$$y_t = y_{t-1} + w_t$$

Example (cont.)

Point forecasts for the ARIMA(0, 1, 0) are:

$y_t(1) = y_t$

$y_t(2) = y_t$

. . .

$y_t(n) = y_t$

Example (cont.)

Forecast variance is given by:
$VAR(1) = \sigma_w^2$
$VAR(2) = 2\sigma_w^2$
. . .
$VAR(n) = n\sigma_w^2$

Example (cont.)

After two or three steps into the future, the confidence intervals become so large as to render the interval forecast meaningless.

APPENDIX

Definition

If X and Y are any two random variables, the conditional expectation of $g(Y)$, given that $X = x$, is defined to be

$$E[g(Y)|X = x] = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

if X and Y are jointly continuous and

$$E[g(Y)|X = x] = \sum_{\text{all } y} g(y)p(y|x)$$

if X and Y are jointly discrete.

Theorem

Let X and Y denote random variables. Then

$$E[Y] = E_X[E(Y|X)]$$

Properties

- $E[aY + bZ + c|X = x] = aE[Y|X = x] + bE[Z|X = x] + c.$
- $E[h(X)|X = x] = h(x).$
- If X and Y are independent, then $E[Y|X] = E[Y]$.

Minimum Square Error Prediction (1)

Suppose Y is a random variable with mean μ_Y and variance σ_Y^2. If our object is to predict Y using only a constant c, what is the best choice for c? A common criterion is to choose c to minimize the **mean square error of prediction**, that is, to minimize

$$g(c) = E[(Y - c)^2]$$

It turns out that the optimal c is $c = E(Y) = \mu$.
(Remember? we showed this together).

Minimum Square Error Prediction (2)

Now consider the situation where a second random variable X is available and we wish to use the observed value of X to help predict Y. Let $\rho = corr(X, Y)$. Suppose that only linear functions $a + bX$ can be used for the prediction. The mean square error is then given by

$$g(a, b) = E[(Y - a - bX)^2]$$

It turns out that the optimal b and a are given by

$$b = \frac{Cov(X, Y)}{Var(X)} = \rho \frac{\sigma_Y}{\sigma_X}$$

$$a = E(Y) - bE(X) = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X$$

(HW?).

Minimum Square Error Prediction (3)

Let us now consider the more general problem of predicting Y with an **arbitrary** function of X. Once more our criterion will be to minimize the mean square error of prediction. We need to choose the function $h(X)$, that minimizes

$$E[Y - h(X)]^2$$

Using properties of conditional expectation and what we proved in Minimum Square Error Prediction (1), we showed that the best choice of $h(X)$ is

$$h(X) = E[Y|X]$$