

# Technical Note: Moving Average Model

---

Occasionally, we receive requests for a technical issue about ARMA modeling beyond our regular NumXL support, which delves more into the mathematical formulation of ARMA. We are always happy to help our users with any question they may have, so we decided to share our internal technical notes with you.

These notes were originally composed when we sat in on a time series analysis class. Over the years, we've maintained these notes with new insights, empirical observations and intuitions acquired. We often go back to these notes for resolving development issues and/or to properly address a product support matter.

In this paper, we'll go over a simple, yet fundamental, econometric model: moving average. This model serves as a cornerstone for all serious discussion on ARMA/ARIMA models.

## Background

A moving average model of order  $q$  (i.e. MA( $q$ )) is defined as follows:

$$\begin{aligned}x_t &= \mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t \\a_t &= \sigma \times \varepsilon_t \\ \varepsilon_t &\sim \text{i.i.d} \sim N(0,1)\end{aligned}$$

Where

- $a_t$  is the innovations or shocks for our process
- $\sigma$  is the conditional standard deviation (aka volatility)

The output value ( $x_t$ ) is solely determined by a long run average ( $\mu$ ) and a weighted sum of past shocks or innovations ( $\{a_t\}$ ).

## Stability

By definition, the MA process is stable and has a finite long-run mean ( $\mu$ ) and variance:

- (1) The unconditional (i.e. long-run) mean is simply  $\mu$

$$E[x_t] = E[\mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t] = \mu$$

- (2) The unconditional (i.e. long run) variance is defined as follows:

$$\text{Var}[x_t] = E[(x_t - \mu)(x_t - \mu)] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t)^2]$$

$$\text{Var}[x_t] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma^2$$

- For a finite order  $q$ , the process is guaranteed to be stable (i.e. does not converge to infinity).
- For an infinite order  $q$  (i.e.  $MA(\infty)$ ), the process is stable only if the long-run variance is finite:

$$\text{Var}[x_t] = (1 + \theta_1^2 + \theta_2^2 + \dots) \sigma^2 = (1 + \sum_{i=1}^{\infty} \theta_i^2) \sigma^2$$

In other words, the sum of the squared values of the MA coefficients is finite.

$$\sum_{i=1}^{\infty} \theta_i^2 < \infty$$

## Forecast

Given an input sample data  $\{x_1, x_2, \dots, x_T\}$ , we can calculate values of the moving average process for future (i.e. out-of-sample) values as follows:

$$x_T = \mu + a_T + \theta_1 a_{T-1} + \theta_2 a_{T-2} + \dots + \theta_q a_{T-q}$$

$$E[x_{T+1} | a_T, a_{T-1}, \dots, a_{T-q}] = \mu + \theta_1 a_T + \theta_2 a_{T-1} + \dots + \theta_q a_{T-q+1}$$

$$E[x_{T+2} | a_T, a_{T-1}, \dots, a_{T-q}] = \mu + \theta_2 a_T + \theta_3 a_{T-1} + \dots + \theta_q a_{T-q+2}$$

...

$$E[x_{T+q} | a_T, a_{T-1}, \dots, a_{T-q}] = \mu + \theta_q a_T$$

$$E[x_{T+q+k} | a_T, a_{T-1}, \dots, a_{T-q}] = \mu$$

The variance (squared standard error) of the out-of-sample values is expressed as follows:

$$\text{Var}[x_{T+1} | a_T, a_{T-1}, \dots, a_{T-q}] = \text{Var}[a_{T+1} + \theta_1 a_T + \theta_2 a_{T-1} + \dots + \theta_q a_{T-q+1}] = \sigma^2$$

$$\text{Var}[x_{T+2} | a_T, a_{T-1}, \dots, a_{T-q}] = \text{Var}[a_{T+2} + \theta_1 a_{T+1} + \theta_2 a_T + \dots + \theta_q a_{T-q+2}] = (1 + \theta_1^2) \sigma^2$$

$$\text{Var}[x_{T+3} | a_T, a_{T-1}, \dots, a_{T-q}] = \text{Var}[a_{T+3} + \theta_1 a_{T+2} + \theta_2 a_{T+1} + \dots + \theta_q a_{T-q+3}] = (1 + \theta_1^2 + \theta_2^2) \sigma^2$$

...

$$\text{Var}[x_{T+q} | a_T, a_{T-1}, \dots, a_{T-q}] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{q-1}^2) \sigma^2$$

$$\text{Var}[x_{T+q+k} | a_T, a_{T-1}, \dots, a_{T-q}] = \text{Var}[x_{T+q+k}] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma^2$$

Note: The conditional variance grows cumulatively over q-steps to reach its long-run (unconditional) variance.

## Stationarity

By definition, a weak-sense stationary (weak stationary) process requires the first moment (i.e. mean) and covariance do not vary with respect to time.

For the first moment, the conditional mean is obviously time-invariant:

$$E[x_t] = E[x_{t+k}] = E[x_{t-k}] = \mu$$

For the second moment (i.e. variance and auto-covariance), let's examine this assumption. By definition, the auto-covariance for lag order j is expressed as follows:

$$\gamma_j = E[(x_t - \mu)(x_{t-j} - \mu)]$$

The auto-covariance of lag order zero ( $\gamma_0$ ) is the same as the unconditional variance:

$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma^2 \\ \gamma_1 &= (\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_4 + \dots + \theta_{q-1}\theta_q)\sigma^2 \\ \gamma_2 &= (\theta_2 + \theta_1\theta_3 + \theta_2\theta_4 + \theta_3\theta_5 + \dots + \theta_{q-2}\theta_q)\sigma^2 \\ \gamma_3 &= (\theta_3 + \theta_1\theta_4 + \theta_2\theta_5 + \theta_3\theta_6 + \dots + \theta_{q-3}\theta_q)\sigma^2 \\ &\dots \\ \gamma_q &= \theta_q\sigma^2 \\ \gamma_{k>q} &= 0 \end{aligned}$$

To examine for variability with respect to time, it is sufficient to examine the following:

$$\begin{aligned} \gamma_j &= E[(x_t - \mu)(x_{t-j} - \mu)] = \gamma_{-j} = E[(x_t - \mu)(x_{t+j} - \mu)] \\ \gamma_j &= \gamma_{-j} = F(j) \end{aligned}$$

Using either definition, we can easily show that the auto-covariance function does not vary with respect to time (t), but is rather determined solely by the lag order j. In sum, the moving average process is a weak-sense stationary process.

## Correlogram

What do the moving average correlogram plots look like? Can we identify a moving average process (and its order) from correlogram plots (i.e. ACF and PACF)?

The auto-correlation function (ACF) is defined as the ratio of auto-covariance and unconditional variance:

$$\text{ACF}_j = \rho_j = \frac{\gamma_j}{\gamma_0}$$

By definition;  $\rho_0 = 1$  and the next q ACF:

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_4 + \dots + \theta_{q-1}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 \dots + \theta_q^2}$$

$$\rho_2 = \frac{\theta_2 + \theta_1\theta_3 + \theta_2\theta_4 + \theta_3\theta_5 + \dots + \theta_{q-2}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 \dots + \theta_q^2}$$

$$\rho_3 = \frac{\theta_3 + \theta_1\theta_4 + \theta_2\theta_5 + \theta_3\theta_6 + \dots + \theta_{q-3}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 \dots + \theta_q^2}$$

...

$$\rho_q = \frac{\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 \dots + \theta_q^2}$$

$$\rho_{k>q} = 0$$

What does this mean? If the ACF plot exhibits significant values for the first j-lags and then drops to zero, we can probably model the process with a moving average model of order j.

Using only the ACF plot, I should be able to construct an MA model for any process, right? **Yes.**

Can two or more different MA processes have the same ACF function values? **Yes.**

Is there a special case (restricted) of MA processes that have a unique ACF function? Yes, they are called invertible MA processes.

In theory, no two invertible MA processes have the same ACF function.

## MA Invertibility

An invertible MA model MA(q) is one that can be represented as a converging infinite order autoregressive AR( $\infty$ ) model. By converging, we mean that the AR coefficient decreases to zero as we go back in time.

Using the lag-operator ( $L$ ) notation, the moving average process can be represented as follows:

$$x_t - \mu = y_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t$$

$$\frac{y_t}{(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)} = a_t$$

Using partial-fraction decomposition:

$$\frac{y_t}{(1 + \lambda_1 L)(1 + \lambda_2 L) \dots (1 + \lambda_q L)} = a_t$$

$$\left[ \frac{c_1}{1 + \lambda_1 L} + \frac{c_2}{1 + \lambda_2 L} + \dots + \frac{c_q}{1 + \lambda_q L} \right] y_t = a_t$$

Where  $\{\lambda_i\}$  is the set of characteristic roots for the MA process.  $\lambda_i$  can take real or complex values.

Now, assuming that  $\lambda_i$  falls inside the unit circle (i.e.  $\|\lambda_i\| < 1$ ) for  $\forall i \leq q$ , each fraction can be expressed as a converging geometric series:

$$\frac{1}{1 + \lambda_i L} = 1 - \lambda_i L + \lambda_i^2 L^2 - \lambda_i^3 L^3 + \lambda_i^4 L^4 - \dots$$

Applying this transformation to all partial-fractions, you can express the MA process as an infinite order AR process follows:

$$\left[ (c_1 + c_2 + \dots + c_q) - (c_1 \lambda_1 + c_2 \lambda_2 + \dots + c_q \lambda_q) L + (c_1 \lambda_1^2 + c_2 \lambda_2^2 + \dots + c_q \lambda_q^2) L^2 \dots \right] y_t = a_t$$

$$[\phi_0 - \phi_1 L + \phi_2 L^2 - \phi_3 L^3 + \dots + (-1)^k \phi_k L^k + \dots] y_t = a_t$$

Where

- $\phi_k = c_1 \lambda_1^k + c_2 \lambda_2^k + \dots + c_q \lambda_q^k$
- $\lim_{k \rightarrow \infty} \phi_k = 0$

To check for invertibility, it is sufficient to find the characteristics roots  $\lambda_i$  and verify that their values fall inside the unit circle ( $\|\lambda_i\| < 1$ ). Going forward, we will only consider the invertible MA process.

**Example: MA(1)**

$$x_t - \mu = (1 + \theta L) a_t$$

The characteristic root for the MA(1) process is  $\theta$ . Assuming  $\|\theta\| < 1$ , the algebraic AR representation of the MA(1) is expressed as follows:

$$\frac{x_t - \mu}{1 + \theta L} = (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \dots)(x_t - \mu) = a_t$$

The AR representation is converging as  $\lim_{k \rightarrow \infty} \theta^k = 0$  for  $\|\theta\| < 1$

We are not quite done yet. There are a couple of pivotal mathematical tricks that we still need to cover: (1) Impulse response function (IRF) and (2) Integrated moving average process.

### 1. Impulse Response Function (IRF)

The impulse response function describes the model output triggered by a single shock at time T.

$$a_t = \begin{cases} 0 & t \neq 1 \\ 1 & t = 1 \end{cases}$$

Applying it to a moving average model of order q, the process value is as follows:

$$\begin{aligned} x_1 &= 1 \\ x_2 &= \theta_1 \\ x_3 &= \theta_2 \\ &\dots \\ x_q &= \theta_{q-1} \\ x_{q+1} &= \theta_q \end{aligned}$$

**Note:** the IRF for an MA process is finite ( $q + 1$ ) and its values are equal to the coefficient.

If we have an IRF of some unknown process, can we model it as an MA process? **You bet!**

Finding the values of the MA coefficients algebraically can be tedious and complex, but using the IRF approach can vastly simplify the task.

#### Example: ARMA(p,q)

Consider the general ARMA(p,q) process:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

OR

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) x_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t$$

We wish to derive the MA representation of the process. We can divide the two components (i.e. AR and MA) into polynomials:

$$x_t = \frac{(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)} a_t$$

Or we can use the IRF of the ARMA process to derive the values of the MA coefficients:

$$\begin{aligned} x_t &= a_t = 1 \\ x_{t+1} &= \phi_1 x_t + \theta_1 a_t = \phi_1 + \theta_1 \\ x_{t+2} &= \phi_1 x_{t+1} + \phi_2 x_t + \theta_2 a_t = \phi_1(\phi_1 + \theta_1) + \phi_2 + \theta_2 \\ x_{t+3} &= \phi_1 x_{t+2} + \phi_2 x_{t+1} + \phi_3 x_t + \theta_3 a_t = \phi_1 x_{t+2} + \phi_2 x_{t+1} + \phi_3 + \theta_3 \\ &\dots \\ x_{t+q+1} &= \phi_1 x_{t+q} + \phi_2 x_{t+q-1} + \dots + \phi_{q+1} x_t \end{aligned}$$

$p > q$

Deriving the MA coefficient values is an iterative and straightforward process that will save us from carrying out complex polynomial division.

By now, you may be wondering why we would wish to convert a finite-order ARMA process to an infinite-order MA representation. For starters, forecasting (mean and error) using an MA representation is much easier than using the original higher order ARMA representation.

## 2. Integration

Integration (i.e. unit root) often arises in time series (e.g. random walk, ARIMA, etc.). In these situations, we model the differenced time series with an ARMA class model:

$$\begin{aligned} (1-L)^d (1-L^s)^D x_t &= y_t \\ (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t &= (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t \end{aligned}$$

But how do we take the ARMA outputs back to the un-differenced scale?

Example 1: Consider a first order integration of the MA(q) process:

$$(1-L)x_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t$$

To calculate the out-of-sample (i.e. forecast) values:

$$\begin{aligned}
 x_{T+1} - x_T &= (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_{T+1} \\
 x_{T+1} &= x_T + a_{T+1} + \theta_1 a_T + \theta_2 a_{T-1} + \dots + \theta_q a_{T+1-q} \\
 x_{T+2} &= x_{T+1} + \theta_1 a_{T+1} + \theta_2 a_T + \theta_3 a_{T-1} + \dots + \theta_q a_{T+2-q} = \\
 x_{T+2} &= x_T + a_{T+2} + (1 + \theta_1) a_{T+1} + (\theta_1 + \theta_2) a_T + (\theta_2 + \theta_3) a_{T-1} + \dots + (\theta_{q-1} + \theta_q) a_{T+2-q} + \theta_q a_{T+1-q} \\
 x_{T+3} &= x_T + a_{T+3} + (1 + \theta_1) a_{T+2} + (1 + \theta_1 + \theta_2) a_{T+1} \\
 &\quad + (\theta_1 + \theta_2 + \theta_3) a_T + \dots + (\theta_{q-2} + \theta_{q-1} + \theta_q) a_{T+3-q} + (\theta_{q-1} + \theta_q) a_{T+2-q} + \theta_q a_{T+1-q} \\
 x_{T+k} &= x_T + a_{T+k} + (1 + \theta_1) a_{T+k-1} + (1 + \theta_1 + \theta_2) a_{T+k-2} + \dots + (1 + \theta_1 + \theta_2 + \dots + \theta_q) a_{T+k-q} \\
 &\quad + (\theta_1 + \theta_2 + \dots + \theta_q) a_T + (\theta_2 + \dots + \theta_q) a_{T-1} + \dots + \theta_q a_{T+1-q}
 \end{aligned}$$

Alternatively,

$$\begin{aligned}
 x_{T+k} &= (1 + (1 + \theta_1)L + (1 + \theta_1 + \theta_2)L^2 + (1 + \theta_1 + \theta_2 + \theta_3)L^3 + \dots + (1 + \theta_1 + \theta_2 + \dots + \theta_q)L^q) a_{T+k} + M \\
 M &= x_T + (\theta_1 + \theta_2 + \dots + \theta_q) a_T + (\theta_2 + \dots + \theta_q) a_{T-1} + (\theta_3 + \dots + \theta_q) a_{T-2} + \dots + \theta_q a_{T+1-q}
 \end{aligned}$$

Furthermore, the variance of the forecast value is constant after q steps:

$$\text{Var}[x_{T+k}] = (1 + (1 + \theta_1)^2 + (1 + \theta_1 + \theta_2)^2 + (1 + \theta_1 + \theta_2 + \theta_3)^2 + \dots + (1 + \theta_1 + \theta_2 + \dots + \theta_q)^2) \sigma^2$$

In sum, the integrated zero-mean MA process yielded another MA process, but with a mean.

What if the differenced time series has a non-zero mean?

$$(1-L)x_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t$$

In this case, we have a time trend of rate equal to  $\mu$

$$\begin{aligned}
 x_{T+k} &= k \times \mu + (1 + (1 + \theta_1)L + (1 + \theta_1 + \theta_2)L^2 + (1 + \theta_1 + \theta_2 + \theta_3)L^3 + \dots + (1 + \theta_1 + \theta_2 + \dots + \theta_q)L^q) a_{T+k} + M \\
 M &= x_T + (\theta_1 + \theta_2 + \dots + \theta_q) a_T + (\theta_2 + \dots + \theta_q) a_{T-1} + (\theta_3 + \dots + \theta_q) a_{T-2} + \dots + \theta_q a_{T+1-q}
 \end{aligned}$$

For higher order integration  $d$  ( $d > 1$ ), we repeat the procedure above  $d$  times, but we would need more realized observations than the last one. For example:

$$\begin{aligned}
 (1-L)^2 x_t &= (1-L)y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) a_t \\
 (1-L)x_t &= y_t
 \end{aligned}$$



To solve for  $y_t$ , we'd need  $y_T$ , which is calculated by  $x_T - x_{T-1}$ . Next, to solve for  $x_t$  using the  $y_t$  process, we'd need  $x_T$ . In sum,  $(x_T, x_{T-1})$  are required as initial conditions for the integration.



## Conclusion

The moving average process, despite its simplicity, is a rather useful model to work with, especially when it comes to forecasting.

Armed with a couple of mathematical tricks (IRF and Integration), we can tackle many more complex processes by representing them first by an MA.

In future technical notes, we will discuss advanced models, but often refer to the MA process and material presented here.

## References

- Hamilton, J.D.; *Time Series Analysis*, Princeton University Press (1994), ISBN 0-691-04289-6
- D. S.G. Pollock; *Handbook of Time Series Analysis, Signal Processing, and Dynamics* , Academic Press (1999), ISBN: 0125609906
- Box, Jenkins and Reisel; *Time Series Analysis: Forecasting and Control* , John Wiley & SONS. (2008) 4th edition, ISBN:0470272848