ARIMA Model and Box-Jenkins Methodology

The ARIMA (AutoRegressive Integrated Moving Average) model is a popular time series forecasting method that captures the relationship between an observation and a number of lagged observations, as well as the lagged forecast errors. The model is defined as:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where:

- Y_t is the actual value at time t.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.
- ϵ_t is the error term (white noise).
- p is the order of the autoregressive part.
- q is the order of the moving average part.
- d is the number of differencing required to make the series stationary.

The Box-Jenkins methodology involves the following steps to identify the best ARIMA model:

1. **Model Identification**: Determine the values of p, d, and q using tools like the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.
2. **Parameter Estimation**: Estimate the coefficients ϕ and θ using methods like Maximum Likelihood Estimation (MLE).
3. **Model Checking**: Validate the model by analyzing residuals and ensuring they behave like white noise.
4. **Forecasting**: Use the fitted model to forecast future values.

Matrices and Equations

For AR(1) process:
$Y_t = \phi_1 Y_{t-1} + \epsilon_t$
Matrix form:

$$\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_1 \end{pmatrix} = \begin{pmatrix} \phi_1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_0 \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$