Mathematical Formulations of ARIMAX

Jerome Chou

January 31, 2025

What is ARIMAX?

- ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) extends ARIMA by incorporating exogenous predictors.
- It is used for time series forecasting when external factors influence the target variable.
- ARIMAX is represented as:

$$ARIMAX(p, d, q) + X_t$$

where X_t represents exogenous variables.

Mathematical Definition of ARIMA

- An ARIMA(p, d, q) model is given by:

$$\Phi_p(B)(1 - B)^d y_t = \Theta_q(B)\epsilon_t$$

where:

- y_t is the observed time series.
- B is the backshift operator ($By_t = y_{t-1}$).
- d is the differencing order.
- ϵ_t is white noise.

Mathematical Definition of ARIMAX

- The ARIMAX(p, d, q) model extends ARIMA by adding exogenous variables:

$$\Phi_p(B)(1 - B)^d y_t = \Theta_q(B)\epsilon_t + \beta X_t$$

  where:
  - X_t represents the exogenous variable(s).
  - β is the coefficient for the exogenous input.
- This allows ARIMAX to capture the effect of external variables on the time series.

Autoregressive (AR) Process

- The AR component models past values:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$$

where:

- ϕ_i are the autoregressive coefficients.
- c is a constant term.
- ϵ_t is white noise.

Moving Average (MA) Process

- The MA component models past error terms:

$$y_t = \mu + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

where:

- θ_j are the moving average coefficients.
- μ is the mean of the series.
- ϵ_t is white noise.

Exogenous Variables in ARIMAX

- Exogenous variables allow ARIMAX to account for external influences:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{m} \beta_k X_{t-k} + \epsilon_t$$

where:

- X_{t-k} represents the exogenous variables.
- β_k are the regression coefficients for exogenous variables.

Loss Function for ARIMAX

- **Mean Squared Error (MSE):**

$$J(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2$$

- **Log-Likelihood Function for MLE:**

$$L(\theta) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2$$

Model Selection Criteria

- **Akaike Information Criterion (AIC):**

$$AIC = 2k - 2\log L$$

  where k is the number of parameters.
- **Bayesian Information Criterion (BIC):**

$$BIC = k\log N - 2\log L$$

  where N is the number of observations.
- Lower AIC/BIC values indicate a better model.

Conclusion

- ARIMAX extends ARIMA by incorporating exogenous predictors.
- The model captures both autoregressive and moving average dynamics, along with external influences.
- Parameters are estimated using MLE or Least Squares.
- Model selection relies on AIC and BIC to balance complexity and fit.