**Vector Autoregression (VAR)**

**1. Khái niệm**

Vector Autoregression (VAR) là một mô hình thống kê được dùng để mô hình hóa và dự báo các chuỗi thời gian đa biến. Không giống như mô hình autoregression đơn (AR), VAR cho phép nhiều biến phụ thuộc lẫn nhau, trong đó mỗi biến được mô hình hóa như một hàm tuyến tính của chính nó và các độ trễ (lags) của tất cả các biến khác trong hệ thống.

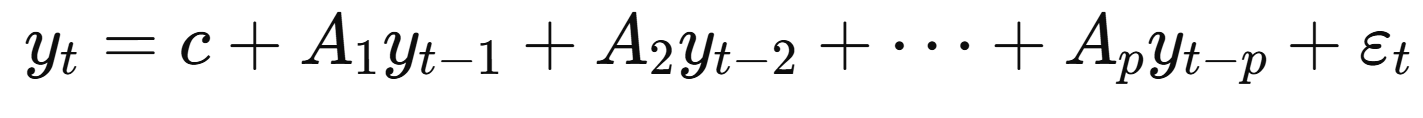
VAR (Vector Autoregression) là một mô hình thống kê dùng để mô hình hóa nhiều biến thời gian phụ thuộc lẫn nhau, trong đó mỗi biến được dự đoán dựa trên:

* Giá trị quá khứ của chính nó
* Giá trị quá khứ của các biến khác trong hệ thống.

### **2. Cấu trúc cơ bản**

**Công thức tổng quát của mô hình VAR**

Giả sử có k biến thời gian và ​ là vector có kích thước k×1, thì mô hình VAR bậc  được viết như sau:



Trong đó:

* ​: vector gồm k biến tại thời điểm t (ví dụ: Open, High, Low, Volume, Close).
* c: vector hằng số
* ​: ma trận hệ số k×k cho từng độ trễ i
* ​: nhiễu trắng (white noise) tại thời điểm t, là phần mà mô hình không thể giải thích được, có kỳ vọng bằng 0 và phương sai không đổi

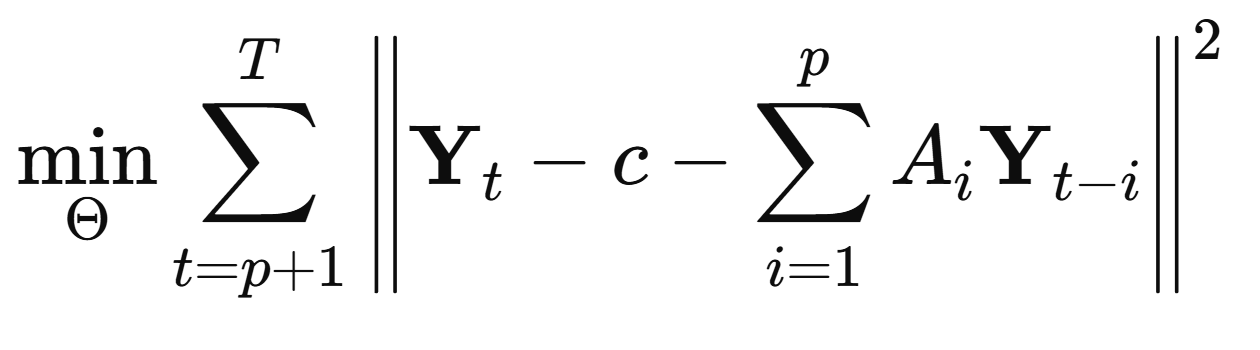
(Nguồn tham khảo công thức trên:

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Link sách: https://link.springer.com/book/10.1007/978-3-540-27752-1)

**Hàm tổn thất và ước lượng bằng phương pháp OLS**

Ước lượng các tham số A1,A2,...,Ap,c bằng hồi quy tuyến tính riêng biệt cho từng biến sử dụng Ordinary Least Squares (OLS):

****

Trong đó:

* 
* Mỗi phương trình hồi quy trong VAR đều có thể ước lượng riêng bằng OLS.

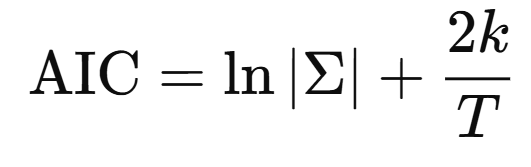
(**Nguồn tham khảo**:

* Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

Link sách: https://press.princeton.edu/books/hardcover/9780691042893/time-series-analysis

)

**Tiêu chí chọn độ trễ: Akaike Information Criterion (AIC)**



* : định thức ma trận hiệp phương sai phần dư (residuals),
* k: số lượng tham số ước lượng,
* T: số quan sát.

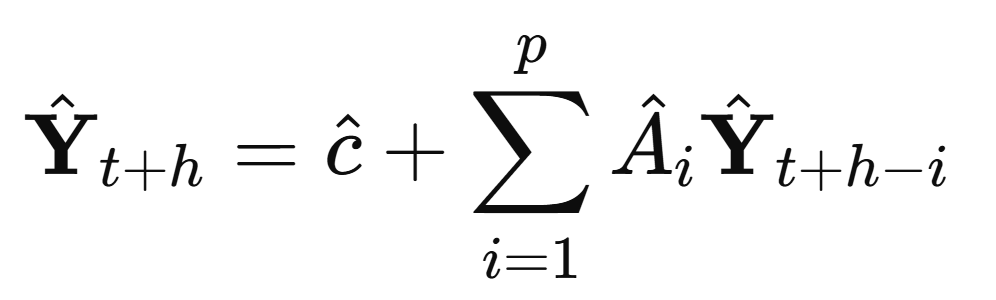
**(Nguồn tham khảo**:

* Akaike, H. (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control*.  
  + DOI: https://ieeexplore.ieee.org/document/1100705

)

**Dự báo tương lai (Forecasting from VAR)**

Giả sử đã ước lượng xong mô hình VAR(p), ta có thể dự báo h bước tiếp theo bằng cách thay thế các giá trị bằng dữ liệu thực (nếu có) hoặc giá trị dự báo:



Dự báo lặp (recursive forecasting): Nếu h>1h > 1h>1, ta dùng giá trị dự báo trước làm đầu vào cho dự báo tiếp theo.

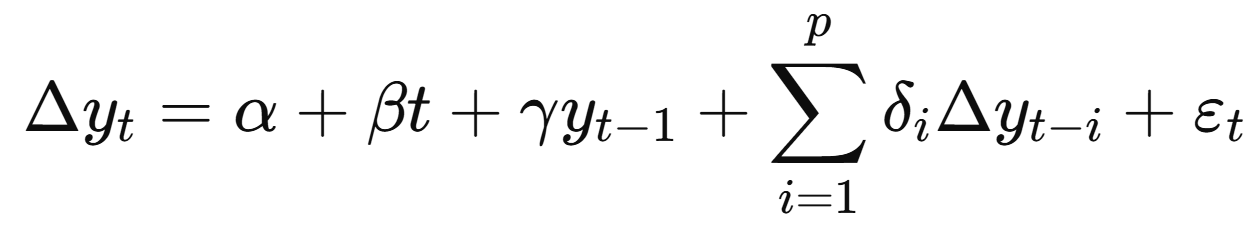
(Nguồn tham khảo công thức trên ***(cùng nguồn với công thức tổng quát)***:

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Link sách: https://link.springer.com/book/10.1007/978-3-540-27752-1)

**Kiểm định tính dừng: Augmented Dickey-Fuller (ADF) Test**

VAR yêu cầu chuỗi phải dừng (stationary), vì vậy cần dùng ADF test để kiểm định:

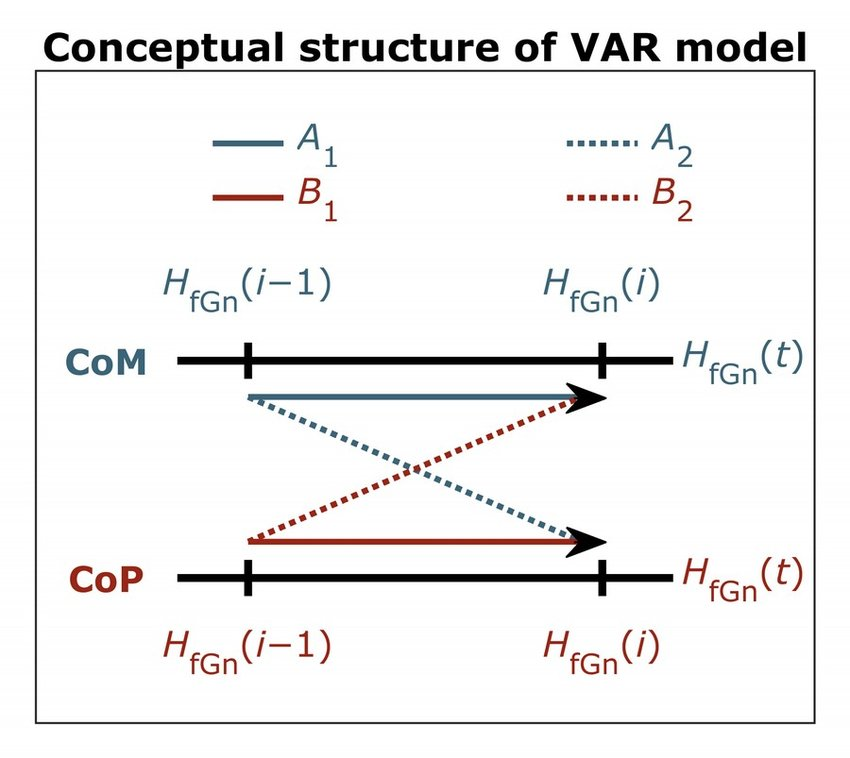


Nếu giá trị p-value < 0.05 ⇒ bác bỏ H0 ⇒ chuỗi dừng.

(**Nguồn tham khảo**:

* Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association*.  
  + JSTOR Link: https://www.jstor.org/stable/2286348

)



Ảnh: Conceptual structure of VAR model

Bảng so sánh Testing Metrics (VAR):

| **Tỷ lệ Train/Test/Val** | **MAE** | **MSE** | **RMSE** | **MAPE** | **R²** |
| --- | --- | --- | --- | --- | --- |
| **75 / 10 / 15** | 1004.94 | 1,579,015.22 | 1256.59 | 9.76% | -0.3769 |
| **70 / 10 / 20** | 1397.44 | 2,777,246.50 | 1,666.51 | 13.23% | -1.2127 |
| **65 / 10 / 25** | **946.54** | **1,233,433.18** | **1,110.60** | **8.40%** | **0.4822** |

### **Nhận xét & Đánh giá:**

### **Tỉ lệ tốt nhất: 65 / 10 / 25**