



Chương 3: Mô hình dự báo đa biến

Thời gian: 5LT+5TH

GVHD: TS. Dương Thị Kim Chi





Nội dung

3.1.

- Giới thiệu

3.2.

- Giới thiệu mô hình đa biến

3.3.

- Các giả định cơ bản của mô hình VAR

3.4.

- Các bước triển khai VAR trên mô hình chuỗi thời gian
- Ứng dụng



1. Giới thiệu

3.1. Giới thiệu mô hình đa biến

Mô hình **Vector Autoregressive (VAR)** là một công cụ mạnh mẽ để hiểu và dự báo các mối quan hệ động giữa nhiều biến

Các mô hình VAR đã trở nên phổ biến đáng kể do khả năng nắm bắt các tương tác phức tạp giữa các biến, cung cấp thông tin chi tiết về hành vi của các hệ thống kinh tế, thị trường tài chính và các quá trình động khác

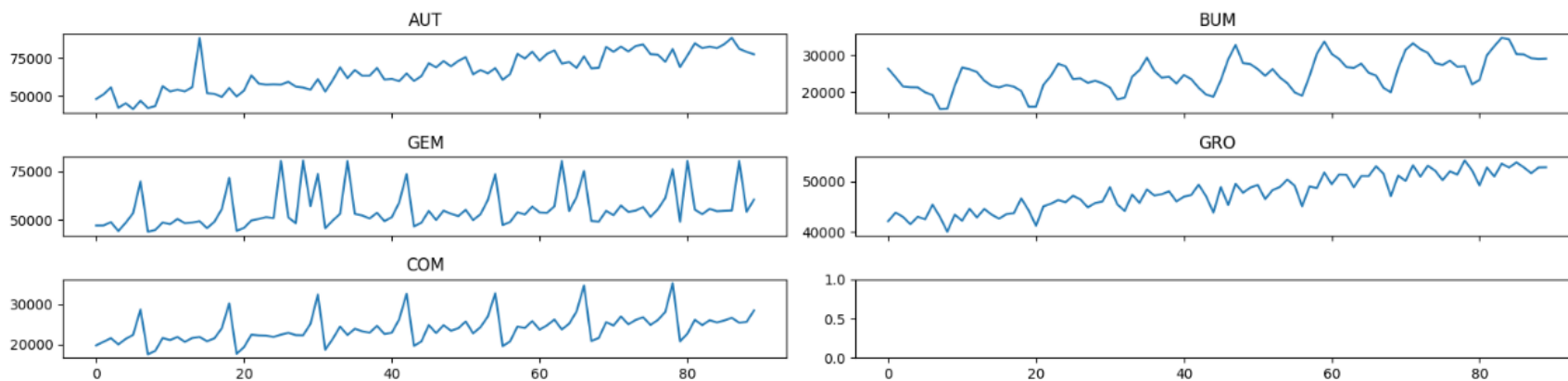
→ **Mô hình VAR** là một mô hình thống kê mô tả hành vi chung của nhiều biến chuỗi thời gian dựa trên các giá trị trễ của chúng. VAR cho phép phân tích đồng thời các mối quan hệ phụ thuộc giữa nhiều biến.



1. Giới thiệu

Xét dữ liệu

| Period | AUT | BUM | GEM | GRO | COM |
|--------------|-------|-------|-------|-------|-------|
| June 09 | 48078 | 26350 | 47089 | 42146 | 19822 |
| July 09 | 51093 | 24003 | 47097 | 43838 | 20704 |
| August 09 | 55677 | 21522 | 48826 | 42994 | 21609 |
| September 09 | 42299 | 21304 | 44202 | 41537 | 20050 |
| October 09 | 45268 | 21263 | 48509 | 43020 | 21425 |
| November 09 | 41448 | 19915 | 53489 | 42508 | 22439 |
| December 09 | 46936 | 19182 | 69970 | 45395 | 28782 |
| January 10 | 42143 | 15349 | 43854 | 42986 | 17570 |
| February 10 | 43434 | 15468 | 44756 | 40061 | 18443 |





Giới thiệu

Mục tiêu: xác định phương trình hồi quy đa biến từ bộ dữ liệu chuỗi thời gian thực tế

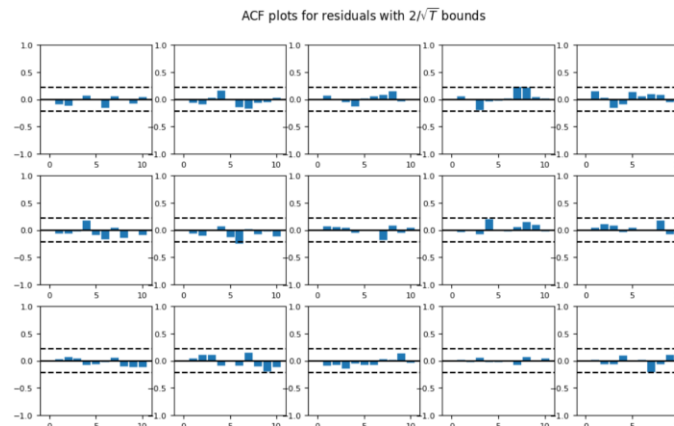
```
Summary of Regression Results
=====
Model:                VAR
Method:               OLS
Date:                 Wed, 31, Jul, 2024
Time:                  23:05:37
-----
No. of Equations:      5.00000      BIC:                81.2990
Nobs:                  84.0000      HQIC:               79.0493
Log likelihood:        -3722.51     FPE:                 5.24177e+33
AIC:                   77.5370      Det(Omega_mle):     1.36116e+33
-----
Results for equation AUT
=====
```

| | coefficient | std. error | t-stat | prob |
|--------|-------------|------------|--------|-------|
| const | 2332.699534 | 774.171432 | 3.013 | 0.003 |
| L1.AUT | -0.653030 | 0.131165 | -4.979 | 0.000 |
| L1.BUM | -0.002831 | 0.535472 | -0.005 | 0.996 |
| L1.GEM | -0.061081 | 0.062048 | -0.984 | 0.325 |
| L1.GRO | 0.269939 | 0.860782 | 0.314 | 0.754 |
| L1.COM | -1.340356 | 0.341155 | -3.929 | 0.000 |

```
res_df = pd.DataFrame(results, columns=['p', 'q', 'AIC'])
res_df.sort_values(by=['AIC']).head()
```

Out[111]:

| | p | q | AIC |
|----|---|---|-------------|
| 12 | 4 | 1 | 8268.239889 |
| 13 | 4 | 2 | 8292.722869 |
| 14 | 4 | 3 | 8331.090871 |
| 15 | 4 | 4 | 8357.682655 |
| 8 | 3 | 1 | 8360.731798 |





1. Giới thiệu (tt)

3.1.2 Chuỗi thời gian đa biến:

Chuỗi thời gian đa biến là dữ liệu đa chiều liên quan đến thời gian. Chúng ta có thể định nghĩa dữ liệu chuỗi thời gian đa biến trong công thức toán học như sau:

$$\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{m,t}]^T$$

$$\text{dim:} \quad (m \times t) \quad (1 \times t) \quad (1 \times t) \quad (1 \times t)$$

Ký hiệu chuỗi thời gian đa biến

trong đó $Z_{i,t}$ là biến thành phần thứ i tại thời điểm t và lưu ý rằng nó là biến ngẫu nhiên cho mỗi i và t . Z_t có chiều dim (m, t) . Khi chúng ta phân tích chuỗi thời gian đa biến, chúng ta không thể áp dụng lý thuyết thống kê chuẩn. ➔ **Hồi quy tuyến tính bội**



3.2. Cấu trúc và Giả định của Mô hình VAR

Mô hình VAR bao gồm một tập hợp các phương trình, trong đó mỗi phương trình biểu diễn một biến như một hàm của các giá trị trễ của chính nó và các **giá trị trễ** của các biến khác trong hệ thống

Các giả định chính của mô hình VAR là **tính dừng, tính tuyến tính và ma trận hiệp phương sai hằng số của các điều khoản lỗi**. Ngoài ra, các mô hình VAR giả định rằng các biến trong hệ thống có tác động đồng thời lên nhau, nắm bắt các tương tác động trong hệ thống.



3.2. Cấu trúc và Giả định của Mô hình VAR

1/Hồi quy tuyến tính bội

Khi tính toán khả năng của hồi quy tuyến tính bội (1), chúng ta giả định rằng mỗi quan sát ($= x_i$) từ mẫu là độc lập với các quan sát khác.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_m x_{i,m} \quad (1)$$

Công thức hồi quy tuyến tính bội

Giả định rằng mỗi quan sát ($= x_i$) từ mẫu là độc lập với các quan sát khác. Do đó, chúng ta có thể dễ dàng tính khả năng bằng tích của mật độ xác suất của các quan sát đơn lẻ. Nhìn chung, chúng ta giả định rằng các quan sát tuân theo phân phối chuẩn với các tham số sau

$$y_i \sim \mathcal{N}(x_i \beta, \sigma^2) \quad (2)$$

$$\mathcal{L}(\beta, \sigma^2; y, X) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_i - x_i \beta)^2}{2\sigma^2} \right) \quad (3)$$

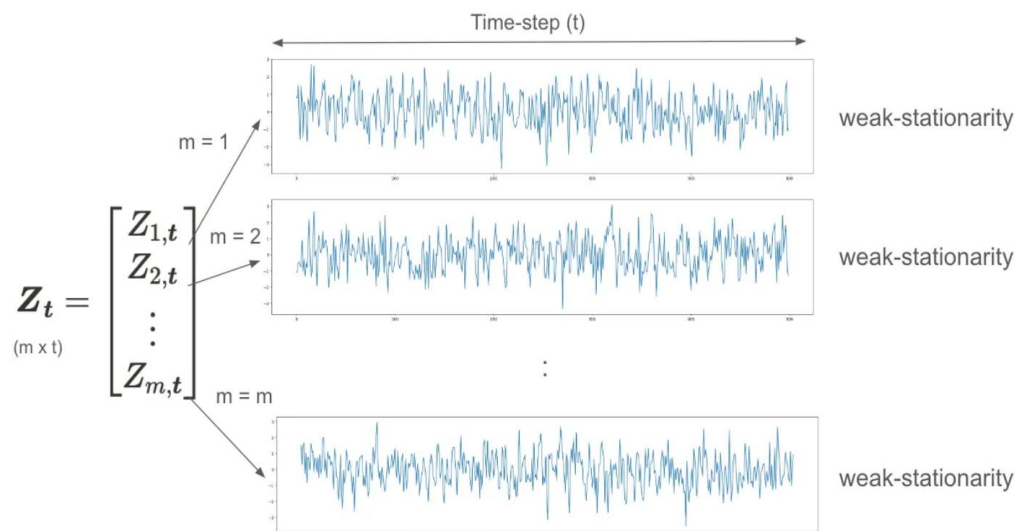


1. Giới thiệu (tt)

Chuỗi thời gian đa biến **m chiều** cũng có tính dừng nếu mỗi chuỗi thành phần là dừng yếu và giá trị trung bình và phương sai của nó không thay đổi theo thời gian

Tính dừng (Stationarity):

Trong chuỗi thời gian đơn biến, khi chuỗi thời gian có cùng giá trị trung bình và phương sai theo thời gian, và hiệp phương sai phụ thuộc vào độ trễ thời gian, thì nó có **tính dừng yếu**. Tương tự, chuỗi thời gian đa biến **m chiều** cũng có tính dừng nếu mỗi chuỗi thành phần là dừng yếu và giá trị trung bình và phương sai của nó không thay đổi theo thời gian.



Mình họa về tính dừng của chuỗi thời gian đa biến



3.2. Cấu trúc và Giả định của Mô hình VAR

2/Ma trận hiệp phương sai và tương quan

Giá trị trung bình của quá trình chuỗi thời gian đa biến m chiều có thể được viết như sau:

$$E[\mathbf{Z}_t] = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}$$

Giá trị kỳ vọng của chuỗi thời gian đa biến m chiều

Vectơ trung bình có chiều $(m, 1)$. Mặt khác, ma trận hiệp phương sai trễ k sẽ như sau:

$$\begin{aligned} \Gamma(k) &= \text{Cov}[\mathbf{Z}_t, \mathbf{Z}_{t+k}] \\ &= \begin{bmatrix} \gamma_{1,1}(k) & \gamma_{1,2}(k) & \cdots & \gamma_{1,m}(k) \\ \gamma_{2,1}(k) & \gamma_{2,2}(k) & \cdots & \gamma_{2,m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,1}(k) & \gamma_{m,2}(k) & \cdots & \gamma_{m,m}(k) \end{bmatrix} \end{aligned}$$

where:

$$\gamma_{i,j}(k) = E[(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j)]$$

Ma trận hiệp phương sai của chuỗi thời gian đa biến m chiều

$k = 0$, ma trận $\Gamma(0)$ có thể dễ dàng được coi là ma trận phương sai-hiệp phương sai giữa các biến khác. Khi $k > 0$, ma trận $\Gamma(k)$ là sự mở rộng của tự hiệp phương sai cho chuỗi thời gian đơn biến



3.2. Cấu trúc và Giả định của Mô hình VAR

Σ là **ma trận hiệp phương sai đối xứng** ($m \times m$). Đây là sự mở rộng của nhiễu trắng trong chuỗi thời gian đơn biến. Các thành phần của quá trình nhiễu trắng không tương quan tại các thời điểm khác nhau, như trường hợp của quá trình nhiễu trắng đơn biến.

Do đó, hiệp phương sai giữa một vector và vector trễ thời gian của nó trở thành 0. Lưu ý rằng nó có thể tương quan đồng thời giữa các thành phần của quá trình nhiễu trắng.



3.4. Các bước triển khai VAR trên mô hình chuỗi thời gian và Ứng dụng

Quy trình triển khai:

Quy trình của họ có thể được tóm tắt như sau:

1. Xác định Mục tiêu.
2. Lấy Dữ liệu.
3. Khám phá và trực quan Chuỗi.
4. Tiền xử lý Dữ liệu.
5. Phân vùng
6. Áp dụng Phương pháp/các Phương pháp Dự báo.
7. Đánh giá và So sánh Hiệu suất.
8. Triển khai Dự báo/Hệ thống.



3.4. Các bước triển khai VAR trên mô hình chuỗi thời gian và Ứng dụng

Bài tập:

Ôn tập các khái niệm có trong VAR

Trình bày cách xây dựng mô hình VAR

Minh họa code



3.4. Các bước triển khai VAR trên mô hình chuỗi thời gian và Ứng dụng



3.4. Các bước triển khai VAR trên mô hình chuỗi thời gian và Ứng dụng



3.4. Các bước triển khai VAR trên mô hình chuỗi thời gian và Ứng dụng



Bài tập (đánh giá định kì)

Tìm hiểu Các mô hình thống kê