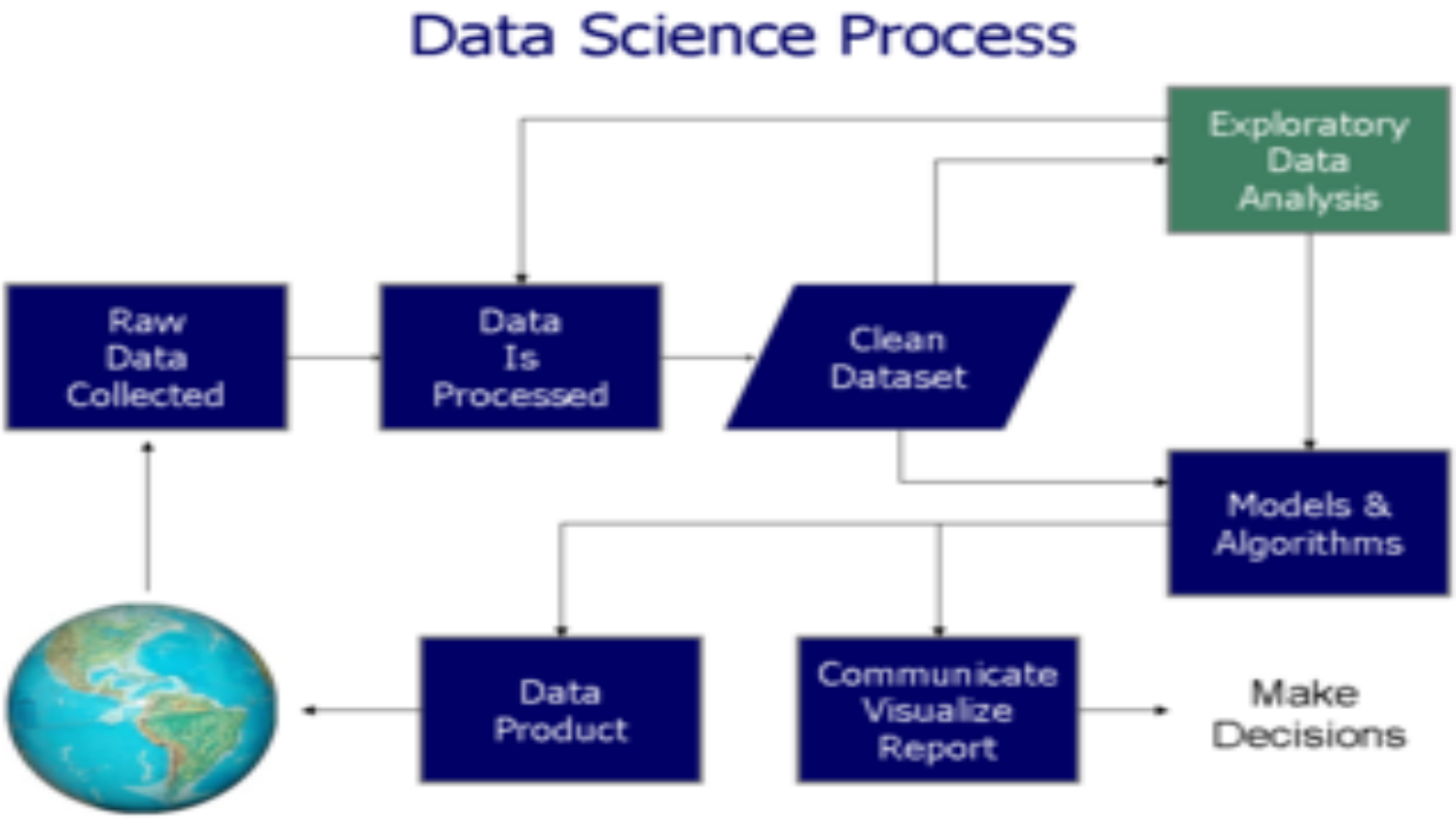
Phương pháp chỉ số dẫn báo và ứng dụng trong phân tích dữ liệu

PSG.TS.Đỗ Văn Thành
Đại học Nguyễn Tất Thành

NỘI DUNG TRÌNH BẦY

1. Sơ lược về Phân tích dữ liệu và kỹ thuật phân tích dữ liệu
2. Mục đích của bài giảng
3. Một số khái niệm cần thiết
4. Chỉ số dẫn báo – Quan hệ nhân quả
5. Chỉ số đồng thời – Quan hệ đồng tích hợp
6. Hồi quy với biến chuỗi thời gian
7. Case Study: Dự báo chỉ số kinh tế vĩ mô theo quý bằng sử dụng các chỉ số báo trước và báo đồng thời
8. Case Study: Dự báo chỉ số VNINDEX bằng ứng dụng phương pháp chỉ số báo trước (hay dẫn báo)

Khoa học dữ liệu



Phân tích dữ liệu (Data Analysis)

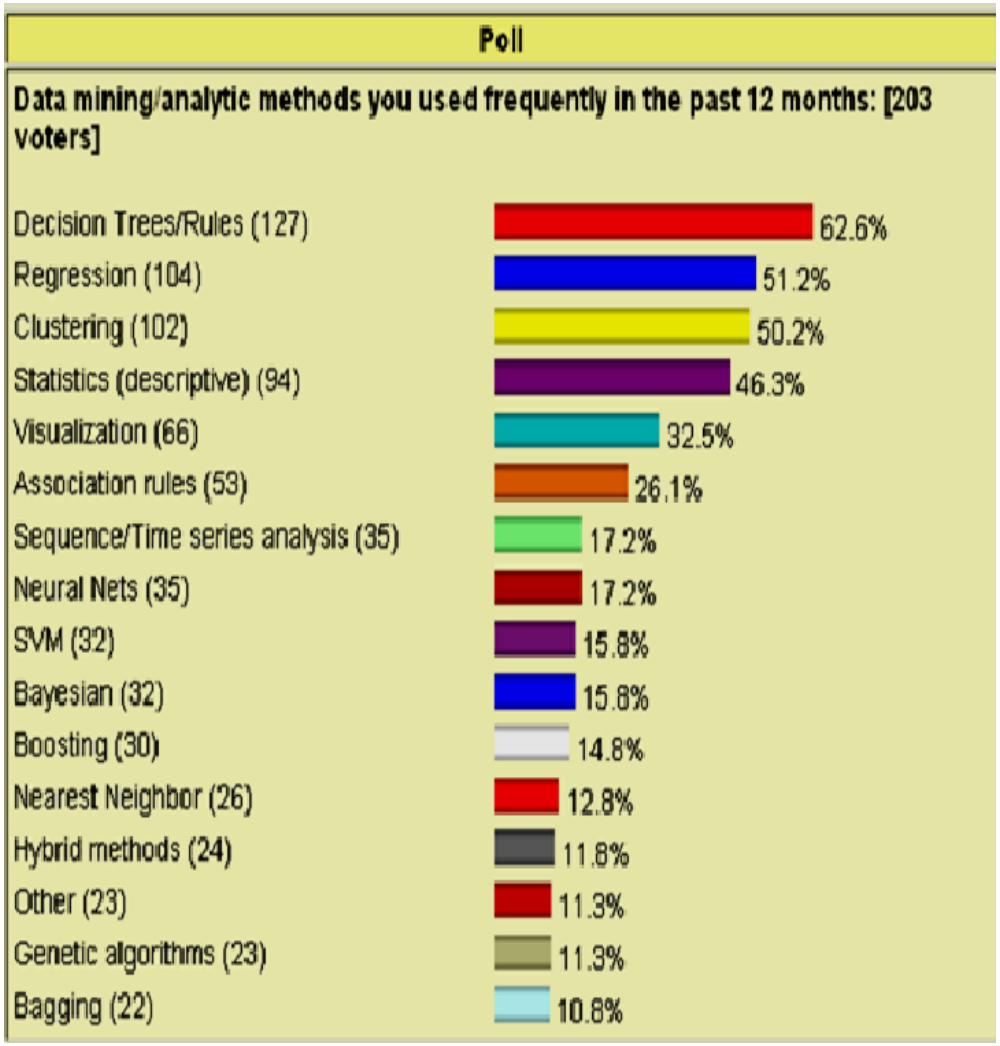
- **Data Analysis (DA):** là quá trình kiểm tra, làm sạch, chuyển đổi, và mô hình hóa dữ liệu với mục đích tìm được những thông tin hữu ích, gợi ý kết luận và hỗ trợ ra quyết định.
- **Data Analysis vs Data mining (DM)**: DM là kỹ thuật DA cụ thể, nó tập trung vào việc mô hình hóa và phát hiện tri thức để dự báo, dự đoán chứ không phải tập trung vào mục đích mô tả.
- **DA vs Business Intelligence (BI)**: bao gồm DA, BI tập trung vào các thông tin kinh doanh, nó chủ yếu dựa vào tích hợp thông tin.

Các kỹ thuật phân tích dữ liệu

1. Các kỹ thuật khai phá dữ liệu (DM),
2. Các kỹ thuật phân tích thông kê: Phân tích thống kê mô tả, phân tích thông kế nhiều chiều, phân tích dữ liệu dạng hàm (bao gồm phân tích thành phần chính dạng hàm), phân tích dữ liệu chuỗi thời gian dạng hàm và chuỗi thời gian mờ, …
3. Phân tích text, BI, trực quan hóa dữ liệu.
4. Các kỹ thuật học máy khác, …

Tình hình ứng dụng các kỹ thuật khai phá dữ liệu (DM)

1. Sở dĩ cây quyết định được sử dụng nhiều nhất bởi: mọi hoạt động của con người đều liên quan đến ra quyết định và cây quyết định là kỹ thuật không quá phức tạp
2. Phương pháp hồi quy có lịch sử nghiên cứu và ứng dụng rất lâu dài và có rất nhiều kiểu hồi quy, ký thuật đã đạt mức độ tinh tế
3. Phương pháp hồi quy đặc biệt được ứng dụng nhiều trong các lĩnh vực KT-XH.



II. MỤC ĐÍCH BÀI GiẢNG

- Bài giảng sẽ giới thiệu:

1. Chỉ số báo trước, chỉ số báo đồng thời và phương pháp xác định các chỉ số này
2. Mô hình dự báo được xây dựng dựa trên các chỉ số báo trước, chỉ số báo đồng thời;
3. Phân tích thông tin rút ra từ mô hình
4. Case Study: Ứng dụng chỉ số báo trước, báo đồng thời trong việc dự báo chỉ số kinh tế vĩ mô/cảnh báo kinh tê theo quý cho Việt Nam
5. Case Study: Ứng dụng chỉ số báo trước trong việc xây dựng mô hình dự báo không điều kiện chỉ số VNINDEX;

III. MỘT SỐ KHÁI NIỆM CẦN THIẾT

1. Khái niệm biến trễ
2. Khái niệm sai phân (thường, mùa vụ)
3. Chuỗi dừng: hồi quy với các biến chuỗi thời gian không dừng có thể là hồi quy sai.
4. Phương pháp đưa chuỗi thời gian không dừng về chuỗi dừng
5. Kiểm định chuỗi dừng trong thực tế ứng dụng

Hiểu biến trễ thế nào ?

Có 3 biến X, W, Z;

Biến **X**: có các quan sát X_t, t = 1,..., T;

Biến **W**: $W_t = X_t$, t = 2,...,T;

Biến **Z**: $Z_t = X_{t-1}$, t = 2,...,T.

Như vậy: W và Z đều có T-1 quan sát.

Nhưng: Z là X trễ, W không là trễ của X.

X	Z	W
X_1	X_2	.
X_2	X_2	X_2
X_3	X_4	X_3
...	...	...
...	...	...
...	...	...
X_{T-1}	X_{T-1}	X_{T-1}
X_T		X_T

Hiểu biến trễ như thế nào ?

- **Ví dụ**: Có 10 quan sát đối với biến Y và X (t =1,…,10) và để chạy mô hình hồi qui Y theo biến X, X trễ, X trễ 2 thời kỳ, X trễ 3 thời kỳ thì làm thể nào ? Thực chất phải ước lượng Phương trình:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + e_t$$

Tạo ra các biến trễ

	Cột A	Cột B	Cột C	Cột D	Cột E
	Y	X	X trễ 1 thời kỳ	X trễ 2 thời kỳ	X trễ 3 thời kỳ
Dòng 1	Y_4	X_4	X_3	X_2	X_1
Dòng 2	Y_5	X_5	X_4	X_3	X_2
Dòng 3	Y_6	X_6	X_5	X_4	X_3
Dòng 4	Y_7	X_7	X_6	X_5	X_4
Dòng 5	Y_8	X_8	X_7	X_6	X_5
Dòng 6	Y_9	X_9	X_8	X_7	X_6
Dòng 7	Y_{10}	X_{10}	X_9	X_8	X_7

Khái niệm sai phân

- Sai phân thường

Toán tử sai phân thường được sử dụng để biến đổi chuỗi thời gian không có tính mùa vụ **không dừng** thành **chuỗi dừng** và được định nghĩa như sau:

- Sai phân bậc 1: $\Delta = z_t - z_{t-1} = (1-B)z_t$

- Sai phân bậc 2: $\Delta^2 z_t = \Delta(\Delta z_t) = z_t - 2z_{t-1} + z_{t-2} = (1\text{-}B)^2 z_t$

- Sai phân bậc d: $\Delta^d z_t = \Delta(\Delta^{d-1} z_t) = (1\text{-}B)^d z_t$

Sai phân mùa vụ

Toán tử sai phân theo mùa vụ được sử dụng để khử tính không dừng theo mùa vụ trong quá trình biến đổi chuỗi thời gian **có tính mùa vụ không dừng** thành thành chuỗi **không có tính mùa vụ và dừng**, nó được định nghĩa như sau:

- Sai phân theo trễ mùa vụ bậc 1: $\Delta_s z_t = z_t - z_{t-s} = (1 - B^s) z_t$

- Sai phân theo trễ mùa vụ bậc D: $\Delta_s^D z_t = (1 - B^s)^D z_t$

ở đây s là khoảng thời gian của một chu kỳ mùa vụ).

Khái niệm chuỗi dừng

- **Chuỗi thời gian dừng:** z_t (t = 1, 2...n) được gọi là chuỗi dừng nếu:

  - - Kỳ vọng: $E(z_t) = \mu = const \quad \forall t$ (1.1.)
  - - Phương sai: $\mathbf{var}(z_t) = \sigma_z^2 \quad \forall t$
  - Tự hiệp phương sai: $\gamma_z(k) = \mathrm{cov}(z_t, z_{t-k}) = \mathrm{cov}(z_q, z_{q-k}) \quad \forall t,q | t \neq q$

với $\gamma_z(k) = \mathrm{cov}\left[z_t, z_{t+k}\right] = E\left[(z_t - \mu)(z_{t+k} - \mu)\right]$

Công thức tổng quát chuyển chuỗi không có xu thế, không dừng thành dừng

Để biến đổi chuỗi mùa vụ thành chuỗi dừng cần thực hiện tuần tự các phép sai phân theo mức trễ mùa vụ và sai phân thường, công thức tổng quát như sau:

$$\Delta^d \Delta^D z_t = (1-B)^d (1-B^s)^D z_t$$

Các phương pháp biến đổi chuỗi thời gian thành chuỗi dừng

1. **Dùng hàm biến đổi**: log, căn bậc 2, …
2. **Dùng phép sai phân**
3. **Khử xu thế: cho chuỗi** Yt, Hồi quy Y theo t
   - Đặt Zt = Yt –(a+bt): thực chất là chuỗi phần dư trong phép hồi quy nêu trên

Nhìn chung phải kết hợp đồng thời cả 3 phép biến đổi trê mới chuyển được một chuỗi không dừng thành chuỗi dừng

Triển khai ứng dụng trong thực tế
(kiểm định nghiệm đơn vị: ADF)

Hồi quy: $Y_t = a^* Y_{t-1} + b$

1) Y_t: không có nghiệm đơn vị (a<1): Y_t dừng không có hệ số chặn;

2) Y_t - a: không có nghiệm đơn vị ; Y_t dừng có hệ số chặn;

3) Y_t – (a + b*t) không có nghiệm đơn vị: Y_t dừng xu thế.

Trong các phần mềm thống kê thường ký hiệu: **d(X)/dX** là sai phần bậc 1 của X

Y_t: không dừng nhưng:

1) $d(Y_t)$ dừng, thì Y_t được gọi là dừng sai phân bậc 1 không có hệ số chặn;

2) $d(Y_t)$ - a: dừng thì Y_t đgl dừng sai phân bậc 1 có hệ số chặn;

3) $d(Y_t)$ – (a + b*t) dừng, thì Y_t đgl dừng sai phân bậc 1 có xu thế.

4) Tương tự cho sai phân bậc 2, 3….

Một số lưu ý

Trong lĩnh vực KT-XH: thay vì làm việc với biến kinh tế X người ta thường làm việc với lô ga cơ số tự nhiên của X (cơ số e) và thường được ký hiệu log(X), vì 2 lý do chính sau:

1). Log(X) là phép biến đổi thường được sử dụng để chuyển một chuỗi không dừng thành chuỗi dừng;

2) (Quan trọng hơn): dlog(X) xấp xỉ bằng tốc độ thay đổi của X,

Thông lệ này phổ biến đến mức người ta nói về biến X nhưng thực ra đang làm việc với log(X) mà không cần bất kỳ giải thích gì thêm.

IV. CHỈ SỐ BÁO TRƯỚC - QUAN HỆ NHÂN QUẢ

1. **Chỉ số tác động**: là biến mà sự biến động của nó có quan hệ ổn định đến biến động của một số biến (chỉ số) khác.
2. Chỉ số tác động có 3 loại (2 loại đầu là quan trọng nhất)
    1) Chỉ số báo trước (Leading Indicator): sự biến động của nó báo trước cho sự biến động của một số chỉ số khác;
    2) Chỉ số báo đồng thời (Coincident Indicator): nó và chỉ số mà nó có quan hệ ổn định: xẩy ra đồng thời, nó có tác động đồng thời;
    3) Chỉ số báo sau (Lag Indicator) cung cấp thông tin biến động trước đó của chỉ số mà nó có quan hệ.
3. PT&DB KT-XH: cần xem xét đồng thời cả 3 chỉ số này.
4. Grange-Engle (Nobel 2004): Là người đặt nền móng về nghiên cứu và ứng dụng của chỉ số báo trước và báo đồng thời.

CHỈ SỐ BÁO TRƯỚC & QUAN HỆ NHÂN QUẢ

Giả sử X, Y đều là chuỗi dừng, Xét 2 phương trình:

$$Y = \sum_{i=1}^{n} a_i X(-i) + \sum_{j=1}^{m} b_j Y(-j) + u_{1t} \qquad (1)$$

$$X_t = \sum_{i=1}^{p} c_i X(-i) + \sum_{j=1}^{q} d_j Y(-j) + u_{2t} \qquad (2)$$

a_i, c_i, b_j, d_j, là các tham số cần được ước lượng; X(-i), Y(-j) là X trễ i và Y trễ j; n, m, p, q là số các biến trễ; u_{it} (i =1, 2) là sai số được giả định là nhiễu trắng.

- ✓ Tồn tại quan hệ nhân quả (QHNQ) duy nhất từ X đến Y nếu $\sum_{i=1}^{n} a_i^2 \neq 0$ và $\sum_{j=1}^{m} d_j^2 = 0$, **X là chỉ số báo trước (dẫn báo) của Y.**
- ✓ Tồn tại QHNQ quả duy nhất từ Y đến X nếu $\sum_{i=1}^{n} a_i^2 = 0$ và $\sum_{j=1}^{m} d_j^2 \neq 0$.
- ✓ Nếu $\sum_{i=1}^{n} a_i^2 \neq 0$ và $\sum_{j=1}^{m} d_j^2 \neq 0$ thì hai biến X và Y: có QHNQ 2 chiều.
- ✓ Các biến X và Y là độc lập nhau nếu $\sum_{i-1}^{n} a_i^2 = 0$ và $\sum_{i-1}^{m} d_j^2 = 0$.

Kiểm định quan hệ nhân quả

1) Kiểm định quan hệ nhân quả (QHNQ) được thực hiện dựa theo 2 công thức (1) và (2) đã nêu;
2) Kiểm định quan hệ nhân quả Granger: xem n=m=p=q với giả thuyết H_0: "biến X không là quan hệ nhân quả của biến Y" .

   Sau đó kiểm tra giá trị xác suất của phân phối T (Student) để bác bỏ hay chấp nhận giả thuyết.
3) Kiểm định QHNQ Granger là có sẵn trong phần mềm R hoặc Eview.

V. QUAN HỆ ĐỒNG TÍCH HỢP VÀ CHỈ SỐ BÁO ĐỒNG THỜI

Các chuỗi I(1) $y_1, y_2, \ldots y_m$ được gọi là đồng tích hợp nếu tồn tại các tham số $\lambda_1, \lambda_2, \ldots, \lambda_m$ không đồng thời bằng 0 sao cho:$(\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_m y_m)$ là chuỗi I(0).

Như vậy: nếu $(\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_m y_m)$ là I(0) thì $(\lambda'_1 y_1 + \lambda'_2 y_2 + \ldots + \lambda'_m y_m)$ cũng là I(0) với $\lambda_i' = a\lambda_i$. Người ta thường chuẩn hóa véc tơ đồng tích hợp bằng cách cho một trong các λ_i nhận giá trị 1, và khi đó mỗi biểu thức I(0): $(\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_m y_m)$ được gọi là *một quan hệ đồng tích hợp, và véc tơ* $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ là *véc tơ đồng tích hợp*

Kiểm định quan hệ đồng tích hợp

Kiểm định quan hệ đồng tích hợp theo lý thuyết

Giả sử Y, X1, X2, …, Xn là I(1), việc kiểm định quan hệ đồng tích hợp giữa biến Y và X1, X2, …, Xn được thực hiện theo các bước dưới đây (còn được gọi là kiểm định Granger – Engle):

1) Hồi qui của biến Y theo X1, X2, …, Xn và lưu phần dư vào một biến khác;
2) Thực hiện kiểm định nghiệm đơn vị của chuỗi phần dư;
3) Nếu giả thuyết nghiệm đơn vị bị loại bỏ thì Y và X1, X2, …, Xn là đồng tích hợp.

Quan hệ đồng tích hợp (tổng quát)

Các chuỗi thời gian $y_1, y_2, \ldots y_m$ được gọi là đồng tích hợp, CI(p,b) nếu:

1. Chúng dừng sai phân bậc p (hay tích hợp bậc p: I(p))

2. Tồn tại tổ hợp tuyến tính của chúng mà tổ hợp này là I(p-b) trong đó b>0.

Kiểm định Quan hệ đồng tích hợp

Kiểm định đồng tích hợp trong Eviews – kiểm định Johansen

Kiểm định vết (trace test):

H_0: số quan hệ đồng tích hợp $\leq r$; H_1: số quan hệ đồng tích hợp $> r$

Thống kê sử dụng là: $\lambda_{trace}(r) = -T\sum_{r+1}^{n}\ln(1-\hat{\lambda}_i)$

Kiểm định dựa trên giá trị riêng lớn nhất

H_0: số quan hệ đồng tích hợp $= r$; H_1: số quan hệ đồng tích hợp $= r+1$

Thống kê sử dụng là $\lambda_{max}(r, r+1) = -T\ln(1-\hat{\lambda}_i)$

Trong đó các ước lượng của các giá trị riêng được xếp theo thứ tự từ lớn đến bé.

Các kiểm định này đều mang tính tuần tự và được thực hiện từ trên xuống và dừng khi nào giả thiết H_0 bị bác bỏ.

Mô hình hiệu chỉnh sai số: ECM

3. Mô hình hiệu chỉnh sai số

Định lý biểu diễn Granger: khi Y và X là đồng tích hợp thì quan hệ giữa chúng được biểu diễn bởi mô hình ECM.

Xét trường hợp mô hình ECM đơn giản:

$$\Delta Y_t = \varphi + \lambda e_{t-1} + \omega_0 \Delta X_t + \varepsilon_t,$$

ở đây $e_{t-1} = Y_{t-1} - \alpha - \beta X_{t-1}$, và ε_t là sai số trong mô hình ECM.

Mô hình ECM có cả tính chất dài hạn lẫn ngắn hạn. Các tính chất dài hạn được tích trữ trong e_{t-1}.

Hành vi ngắn hạn được nắm bắt một phần bởi e_{t-1}, cụ thể nó nói rằng nếu Y nằm ngoài trạng thái cân bằng, Y sẽ được kéo lại ở giai đoạn tiếp theo. Hành vi ngắn hạn còn được nắm giữ bởi việc bao gồm ΔX_t như là biến giải thích. Điều đó ngầm ý rằng nếu X thay đổi, giá trị cân bằng của Y cũng thay đổi và khi đó Y cũng thay đổi.

Ước lượng mô hình hiệu chỉnh sai số

5. Ước lượng mô hình ECM:

1) Hồi qui Y theo X và lưu phần dư vào biến khác;

2) Hồi qui ΔY theo ΔX và theo phần dư ở bước 1 được trễ 1 giai đoạn.

Cần lưu ý là trước khi thực hiện thủ tục 2 bước ước lượng mô hình ECM, cần phải kiểm tra rằng Y và X có nghiệm đơn vị và đồng tích hợp.

Mô hình hiệu chỉnh sai số tổng quát

4. Mô hình ECM tổng quát

Mô hình ADL(p,q) có trễ của biến độc lập và biến giải thích. Mô hình ECM cũng vậy, nó cũng có các trễ và nó cũng có thể có xu thế, bởi vậy mô hình ECM sai số tổng quát đối với hai biến Y, X có dạng:

$$\Delta Y_t = \alpha + \delta t + \rho e_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta Y_{t-i} + \sum_{m-0}^{q-1} \beta_m \Delta X_{t-m} + \varepsilon_t ,$$

ở đây ε_t là phần dư của mô hình ECM, e_t là phần dư trong phép hồi qui biến chuỗi thời gian Y theo biến X.

VI. HỒI QUY VỚI BIẾN CHUỖI THỜI GIAN

- Trước khi chạy hồi qui bất kỳ chuỗi thời gian nào, cần phải kiểm tra các tính chất đơn biến của các biến, trong đó đặc biệt là kiểm định nghiệm đơn vị.
- **Dưới đây giả thiết các biến Y và X cùng kiểu dừng**

Hồi qui chuỗi thời gian khi X, Y là dừng

Mô hình trễ phân bố tự hồi quy ADL(p, q) giữa biến Y và X có thể được viết như sau:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta Y_{t-i} + \theta_1 X_t + \sum_{m=0}^{q-1} \beta_m \Delta X_{t-m} + e_t$$

ở đây: α là hệ số chặn và e_t là sai số, được giả thiết là nhiễu trắng; p, q là độ dài các trễ và được xác định bằng phân tích tương quan đồ của hàm tự tương quan từng phần hoặc bằng phương pháp đã được trình bầy.

- **Nhân tử dài hạn đánh giá tác động X đến Y:** $-\frac{\theta_1}{\rho}$

Hồi qui chuỗi thời gian khi X, Y không dừng nhưng có quan hệ đồng tích hợp

1. Khái niệm đồng tích hợp

Y và X được gọi là đồng tích hợp nếu X, Y có nghiệm đơn vị nhưng sai số của phép hồi qui Y theo X (hoặc X theo Y) là chuỗi dừng.

Khi đó tránh được hồi qui sai.

2. Kiểm định quan hệ đồng tích hợp

Việc kiểm định quan hệ đồng tích hợp giữa biến Y và X được thực hiện theo các bước dưới đây (còn được gọi là kiểm định Granger – Engle):

1) Hồi qui của biến Y theo X và lưu phần dư vào một biến khác;
2) Thực hiện kiểm định nghiệm đơn vị của chuỗi phần dư;
3) Nếu giả thuyết nghiệm đơn vị bị loại bỏ thì Y và X là đồng tích hợp.

3. Y và X có quan hệ đồng tích hợp: Hồi quy Y theo X dựa vào mô hình hiệu chỉnh sai số

Hồi qui chuỗi thời gian khi X, Y có nghiệm đơn vị nhưng không là đồng tích hợp

Trường hợp hai biến chuỗi thời gian Y, X có nghiệm đơn vị nhưng không đồng tích hợp, ta không thể hồi qui Y theo X. Khi đó mô hình ước lượng có dạng:

$$\Delta^2 Y_t = \alpha + \delta t + \rho \Delta Y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta^2 Y_{t-i} + \theta_1 \Delta X_t + \sum_{m-0}^{q-1} \beta_m \Delta^2 X_{t-m} + e_t$$

ở đây $\Delta^2 X_t = \Delta(\Delta X_t) = X_t - 2X_{t-1} + X_{t-2}$ là sai phân bậc hai của X_t.

Dự báo kiểm định – đánh giá độ chính xác dự báo

Bảng 5: *So sánh VNINDEX thực tế và VNINDEXF được dự báo bằng mô hình*

Thứ	Ngày	VNINDEX	VNINDEXF	% sai số
Thứ 2	25/04/2016	596.6	587.1673	-1.58
Thứ 3	26/04/2016	594.35	600.2165	0.99
Thứ 4	27/04/2016	596.6	594.815	-0.30
Thứ 5	28/04/2016	593.65	597.9075	0.72
Thứ 6	29/04/2016	595.45	591.7211	-0.63
Thứ 4	04/05/2016	597.75	596.4889	-0.21
Thứ 5	05/05/2016	602.15	597.9781	-0.69

VII. DỰ BÁO CHỈ SỐ KINH TẾ VĨ MÔ THEO QUÝ - SỬ CÁC CHỈ SỐ DẪN BÁO

- **Tập số liệu các biến**: từ 1995Q1 đến 2012Q3, Nguồn: TCTK

1) GDP94: Tổng sản phẩm trong nước quý (GDP quý) theo giá cố định 1994;
2) NN94: GDP quý theo giá cố định 1994 ngành Nông, Lâm, thủy sản;
3) CN94: GDP quý theo giá cố định 1994 ngành Công nghiệp-Xây dựng;
4) DV94: GDP quý theo giá cố định 1994 ngành Dịch vụ;
5) CPI: Chỉ số giá tiêu dùng quý so với tháng 12 năm trước;
6) CPI_CK: Chỉ số giá tiêu dùng so với cùng kỳ năm trước;
7) GDPHH: Tổng sản phẩm trong nước quý (GDP quý) theo giá hiện hành;
8) DTNS: Đầu tư xây dựng cơ bản theo quý từ nguồn ngân sách nhà nước giá hiện hành;
9) TDTN: tiêu dùng tư nhân quý giá hiện hành;
10) TMBL: Tổng mức bán lẻ hàng hóa quý giá hiện hành;
11) XK : Kim ngạch xuất khẩu hàng hóa theo quý giá hiện hành;
12) NK Kim ngạch xuất khẩu hàng hóa theo quý giá hiện hành;
13) ER là tỷ giá hối đoái trung bình theo quý của đồng USD so với Việt Nam đồng.

Kiểm định tính dừng của các chỉ số KTVM theo quý

- Log(GDP94) là chuỗi dừng sai phân bậc 2;
- Log(CPI) là chuỗi dừng;
- Log(GDPHH), Log(DTNS), Log(TDTN), Log(TMBL), Log(XK) Log(NK là các chuỗi dừng sai phân bậc 1;
- Log(ER): là chuỗi dừng sai phân bậc 2;

Phát hiện quan hệ nhân quả giữa các biến kinh tế vĩ mô quý (trễ 4)

Biến kết Quả	Các biến nguyên nhân											
	D(GDP94, 2)	D(NN94)	D(CN94)	D(DV94)	D(GDPhh)	D(CPI)	D(DTNS)	D(TDTN)	D(TMBL)	D(XK)	D(NK)	D(ER)
D(GDP94, 2)		X	X	X	X	0	X	X	X	X	X	0
D(NN94)	X		0	0	X	0	0	X	0	X	X	X
D(CN94)	X	0		X	X	X	X	X	X	X	X	0
D(DV94)	X	X	X		X	0	0	X	0	0	X	0
D(GDPhh)	X	X	X	X		0	0	X	0	X	X	0
D(CPI)	X	X	X	X	0		X	0	0	X	X	0
D(DTNS)	X	X	X	X	X	X		X	X	X	X	0
D(TDTN)	X	X	X	X	X	0	0		0	X	X	0
D(TMBL)	X	0	X	X	X	X	0	X		X	X	0
D(XK)	X	X	X	X	X	0	X	X	X		X	X
D(NK)	X	X	X	X	X	X	X	X	X	X		X
D(ER)	X	X	X	X	X	0	X	X	0	0	X	

Dấu x: là ký hiệu có quan hệ nhân quả mức ý nghĩa thống kê <10%; đồng nhất log(Y) với biến Y.

Mô hình dự báo XK theo quý sử dụng chỉ số dẫn báo

Mô hình dự báo xuất khẩu sử dụng chỉ số dẫn báo

Dependent Variable: D(LOG(XK))
Sample (adjusted): 1995Q1 2012Q2
Included observations: 64 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOG(XK(-1)))	0.495459	0.074860	6.618506	0.0000
D(LOG(XK(-3)))	0.196949	0.068248	2.885777	0.0057
D(LOG(TDTN(-3)))	0.562655	0.043452	12.94893	0.0000
D(LOG(TDTN(-4)))	0.215937	0.043349	4.981314	0.0000
D(LOG(GDPHH(-2))/LOG(GDP94(-2)))	-6.816157	1.235754	-5.515788	0.0000
D(LOG(GDPHH(-4))/LOG(GDP94(-4)))	-12.17916	1.388070	-8.774165	0.0000
D(LOG(ER(-4)),2)	-0.451549	0.234952	-1.921875	0.0601
D(LOG(DTNS(-1)))	-0.043962	0.016883	-2.603987	0.0120
D(LOG(DTNS(-3)))	-0.039227	0.018774	-2.089457	0.0416
D(LOG(TMBL(-2)))	0.454620	0.078305	5.805766	0.0000
AR(1)	-0.338577	0.137543	-2.461619	0.0172
R-squared	0.747679	Adjusted R-squared		0.694303
Log likelihood	106.7198	Durbin-Watson stat		1.978628

Phát hiện quan hệ đồng tích hợp và chỉ số báo đồng thời

Chỉ có các biến log(GDPhh), log(DTNS), log(TDTN), log(TMBL), log(XK) và log(NK) là cùng kiểu dừng (dừng sai phân bậc 1).

Quan hệ đồng tích hợp được phát hiện trên các biến này.

Dependent Variable: LOG(XK)
Method: Least Squares
Date: 12/08/12 Time: 17:36
Sample (adjusted): 1996Q1 2012Q3
Included observations: 67 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(NK)	0.584411	0.077686	7.522709	0.0000
LOG(TMBL)	0.168272	0.088617	1.898861	0.0622
LOG(GDPHH)	0.313752	0.114478	2.740728	0.0080
C	-2.226334	0.417487	-5.332701	0.0000

R-squared	0.983983	Mean dependent var	8.796932
Adjusted R-squared	0.983220	S.D. dependent var	0.845133
S.E. of regression	0.109477	Akaike info criterion	-1.528368
Sum squared resid	0.755062	Schwarz criterion	-1.396744
Log likelihood	55.20031	F-statistic	1290.083
Durbin-Watson stat	1.102027	Prob(F-statistic)	0.000000

Mô hình dự báo XK theo quý sử dụng Quan hệ đồng tích hợp

Mô hình dự báo xuất khẩu sử dụng Quan hệ đồng tích hợp

Biến phụ thuộc: D(LOG(XK))
Mẫu (được hiệu chỉnh): 1996Q4 2012Q2
Số quan sát: 64 sau khi hiệu chỉnh

Biến	Hệ số	Sai số chuẩn	GT thống kê T	Xác suất
LOG(XK(-1))-0.58*LOG(NK(-1))-0.17*LOG(TMBL(-1))-0.31*LOG(GDPHH(-1))+2.23	-0.237450	0.071205	-3.334733	0.0015
D(LOG(NK))	0.247772	0.075366	3.287577	0.0017
D(LOG(NK(-1)))	0.112443	0.057986	1.939134	0.0574
D(LOG(NK(-2)))	0.147641	0.069212	2.133157	0.0372
D(LOG(TMBL(-2)))	0.336077	0.117648	2.856625	0.0060
D(LOG(GDPHH))	0.383450	0.060735	6.313485	0.0000
D(LOG(GDPHH(-3)))	0.309856	0.064854	4.777713	0.0000

R2	0.660979	Thống kê DW	2.006597
R2 được hiệu chỉnh	0.625292		

Mô hình hiệu chỉnh sai số (ECM): dự báo XK

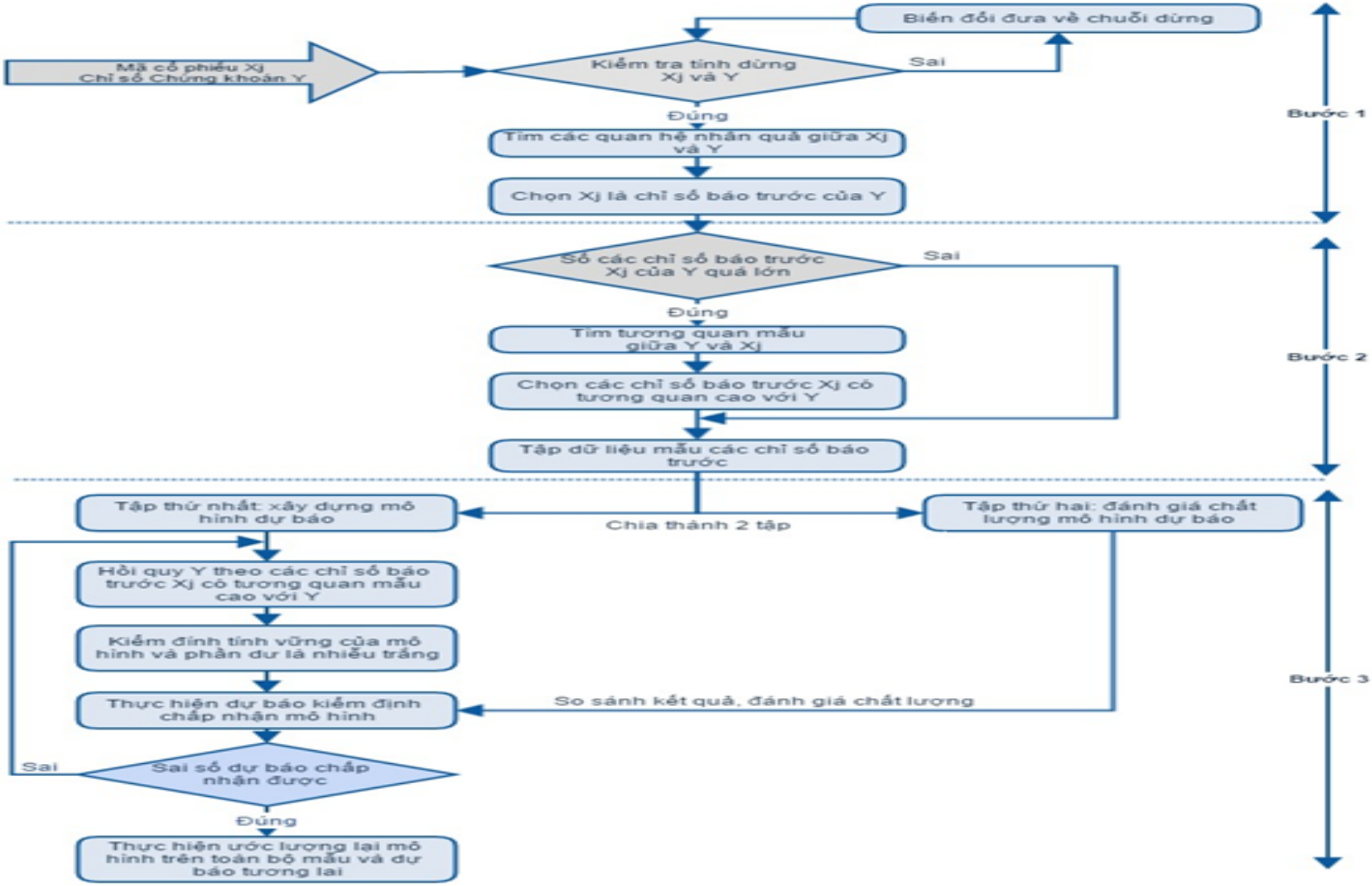
$$D(LOG(XK)) = -0.237*[LOG(XK(-1))-0.58*LOG(NK(-1))-0.17*LOG(TMBL(-1))-0.31*LOG(GDPHH(-1))+2.23] + 0.248*D(LOG(NK)) + 0.112*D(LOG(NK(-1))) + 0.148*D(LOG(NK(-2))) + 0.336*D(LOG(TMBL(-2))) + 0.383*D(LOG(GDPHH)) + 0.31*D(LOG(GDPHH(-3)));$$

R^2: 0.66; DW: 2.006

Đánh giá chất lượng dự báo bằng mô hình

	Mô hình dử dụng chỉ số báo trước	Mô hình sử dụng Quan hệ đồng tích hợp (báo đồng thời)	Giá trị thực
Quý 2/2012	28976.31	29397.56	28604
Quý 3/2012	30821.3	30849.63	30662
% Sai sô tuyệt đối	1.3	2.8	
% Sai sô tuyệt đối	0.5	0.6	

VIII. DỰ BÁO VNINDEX BẰNG ỨNG DỤNG PHƯƠNG PHÁP CHỈ BÁO TRƯỚC



ÁP DỤNG PHƯƠNG PHÁP: DỰ BÁO VNINDEX

Tập dữ liệu để dự báo: 277 biến dữ liệu giao dịch CP và chỉ số VNINDEX, từ 4/1/2010 đến 5/5/2016 (1574 quan sát).

Bước 1: Tìm chỉ số báo trước của dVNINDEX

- Tính chất tập dữ liệu: tất cả các biến dữ liệu giao dịch CP đều dừng; VNINDEX không dừng nhưng dVNINDEX dừng;

$$Y = \sum_{i=1}^{n} a_i X(-i) + \sum_{j=1}^{m} b_j Y(-j) + u_{1t}$$

$$X_t = \sum_{i=1}^{p} c_i X(-i) + \sum_{j=1}^{q} d_j Y(-j) + u_{2t}$$

ÁP DỤNG PHƯƠNG PHÁP: DỰ BÁO VNINDEX

Bước 2: Lựa chọn chỉ số báo trước làm biến giải thích

- Thực hiện tính hệ số tương quan mẫu giữa các chỉ số dẫn báo với dVNINDEX
- Xác định ngưỡng của hệ số tương quan mẫu: có trị tuyệt đối $\geq$ 0.0399. Có 6 biến dữ liệu như vậy.

Các biến dữ liệu là nguyên nhân Gr, hệ số tương quan với DVNINDEX

Số thứ tự	Mã cổ phiếu (1)	Mức ý nghĩa của giả thuyết H_0 (2)	Tương quan mẫu với ΔVNINDEX (3)	Mã cổ phiếu được lựa chọn	Số thứ tự	Mã cổ phiếu (1)	Mức ý nghĩa của giả thuyết H_0 (2)	Tương quan mẫu với ΔVNINDEX (3)	Mã cổ phiếu được lựa chọn
1	AGF	*	0.0239		20	PVD	*	-0.0054	
2	AGR	*	0.0204		21	PXL	*	0.0399	PXL
3	ANV	*	-0.0629	ANV	22	RIC	**	-0.0049	
4	ASM	***	0.0022		23	SAM	*	-0.0016	
5	BGM	*	0.0132		24	SFI	*	-0.0048	
6	D2D	*	-0.0151		25	SRF	**	-0.0258	
7	DPR	**	-0.0191		26	TCL	*	0.0099	
8	EIB	**	0.0436	EIB	27	TDC	**	-0.0043	
9	FMC	**	-0.0361		28	TIC	**	-0.0666	TIC
10	HU1	*	0.0175		29	TLH	*	0.0122	
11	ITC	***	-0.0534	ITC	30	TMP	***	-0.0284	
12	KSA	*	0.0017		31	TNC	*	-0.0132	
13	LSS	*	0.0238		32	TRA	*	0.0316	
14	NBB	*	0.0032		33	TYA	**	0.0349	
15	NTL	**	0.0048		34	VCF	***	-0.0876	VCF
16	PJT	*	-0.0096		35	VFG	*	0.0047	
17	POM	*	-0.012		36	VID	*	-0.0178	
18	PPI	***	-0.0206		37	VNA	**	0.002	
19	PTC	**	-0.0256		38	VRC	*	0.0164	

ÁP DỤNG PHƯƠNG PHÁP: DỰ BÁO VNINDEX

Bước 3: Xây dựng mô hình dự báo

- Chia Tập dữ liệu thành 2: tập thứ nhất từ 4/1/2010 đến 22/4/2016; tập thứ 2: còn lại;
- Thực hiện hồi quy d(log(VNINDEX)) theo các chỉ số dẫn báo được chọn theo công thức:

$$Y_t = \sum_{i=1}^{n} a_i X_1(-i) + \sum_{j=1}^{p} b_j X_2(-j) + ... + \sum_{h=1}^{k} c_h X_n(-h) + \sum_{q=1}^{m} d_q Y(-q) + u_t$$

Y = d(log(VNINDEX)); $X_1, X_2, \ldots, X_n$ (n=6) là các chỉ số báo trước của biến dVNINDEX.

- Quá trinh hồi quy phải kiểm định: mô hình cần ổn định, phần dư phải nhiễu trắng,

Mô hình dự báo chỉ số VNINDEX

Biến phụ thuộc: dlog(VNINDEX)

Số quan sát: 1560 sau khi điều chỉnh;

Biến	Hệ số	Sai số	Mức ý nghĩa
DLOG(VNINDEX(-1))	0.251563	0.026508	***
DLOG(VNINDEX(-2))	-0.070045	0.025714	***
DLOG(VNINDEX(-3))	0.116759	0.031631	***
DLOG(VNINDEX(-4))	-0.070953	0.025003	***
ITC(-1)	-5.84E-11	2.28E-11	**
ITC(-4)	1.02E-10	2.27E-11	***
ITC(-5)	-5.55E-11	2.21E-11	**
TIC(-1)	-4.12E-09	1.22E-09	***
VCF(-2)	-1.09E-09	3.06E-10	***
C	0.000946	0.000341	***
~~DLOG(VNINDEX(-1))*VCF(-4)~~	~~4.63E-08~~	~~1.24E-08~~	~~***~~
~~DLOG(VNINDEX(-3))*DLOG(VNINDEX(-4))~~	~~6.940.499~~	~~1.831.745~~	~~***~~
~~DLOG(VNINDEX(-3))*EIB(-2)~~	~~-2.63E-09~~	~~9.16E-10~~	~~***~~
~~DLOG(VNINDEX(-3))*VCF(-5)~~	~~3.20E-08~~	~~1.43E-08~~	~~**~~
~~ANV(-5)*PXL(-3)~~	~~-7.47E-17~~	~~2.67E-17~~	~~***~~
~~EIB(-2)*VCF(-5)~~	~~-6.57E-17~~	~~1.64E-17~~	~~***~~
~~EIB(-4)*PXL(-1)~~	~~-1.21E-17~~	~~4.19E-18~~	~~***~~
~~EIB(-4)*PXL(-2)~~	~~1.48E-17~~	~~3.91E-18~~	~~***~~
~~EIB(-4)*VCF(-5)~~	~~4.96E-17~~	~~1.46E-17~~	~~***~~
~~ITC(-1)*VCF(-4)~~	~~-6.25E-17~~	~~3.15E-17~~	~~**~~
~~ITC(-4)*VCF(-5)~~	~~-1.49E-16~~	~~4.73E-17~~	~~***~~
~~PXL(-1)*VCF(-2)~~	~~2.37E-16~~	~~8.37E-17~~	~~***~~
~~VCF(-4)*VCF(-5)~~	~~4.39E-16~~	~~8.11E-17~~	~~***~~

$R^2 = 0.164$; Thống kê DW = 1.994797.

Một vài kết luận

- **Ưu**: PP được áp dụng cho nhiều chỉ số KT-XH, kể cả giá, giá trị dao dịch của một số cổ phiếu nào đó; Độ chính xác dự báo là khá cao; Là mô hình dự báo không điều kiện; thích hợp với dự báo ngắn hạn;
- **Nhược**: chỉ đưa được 6/38 chỉ số dẫn báo vào mô hình, thiếu các biến kinh tế - tài chính khác (rất khó có theo ngày, chỉ có được khi điều tra); quan hệ nhân quả thường thay đổi nên phải cập nhật; Để dự báo trung và dài hạn VNINDEX: cần sử dụng mô hình dự báo có điều kiện;

HỎI VÀ ĐÁP