



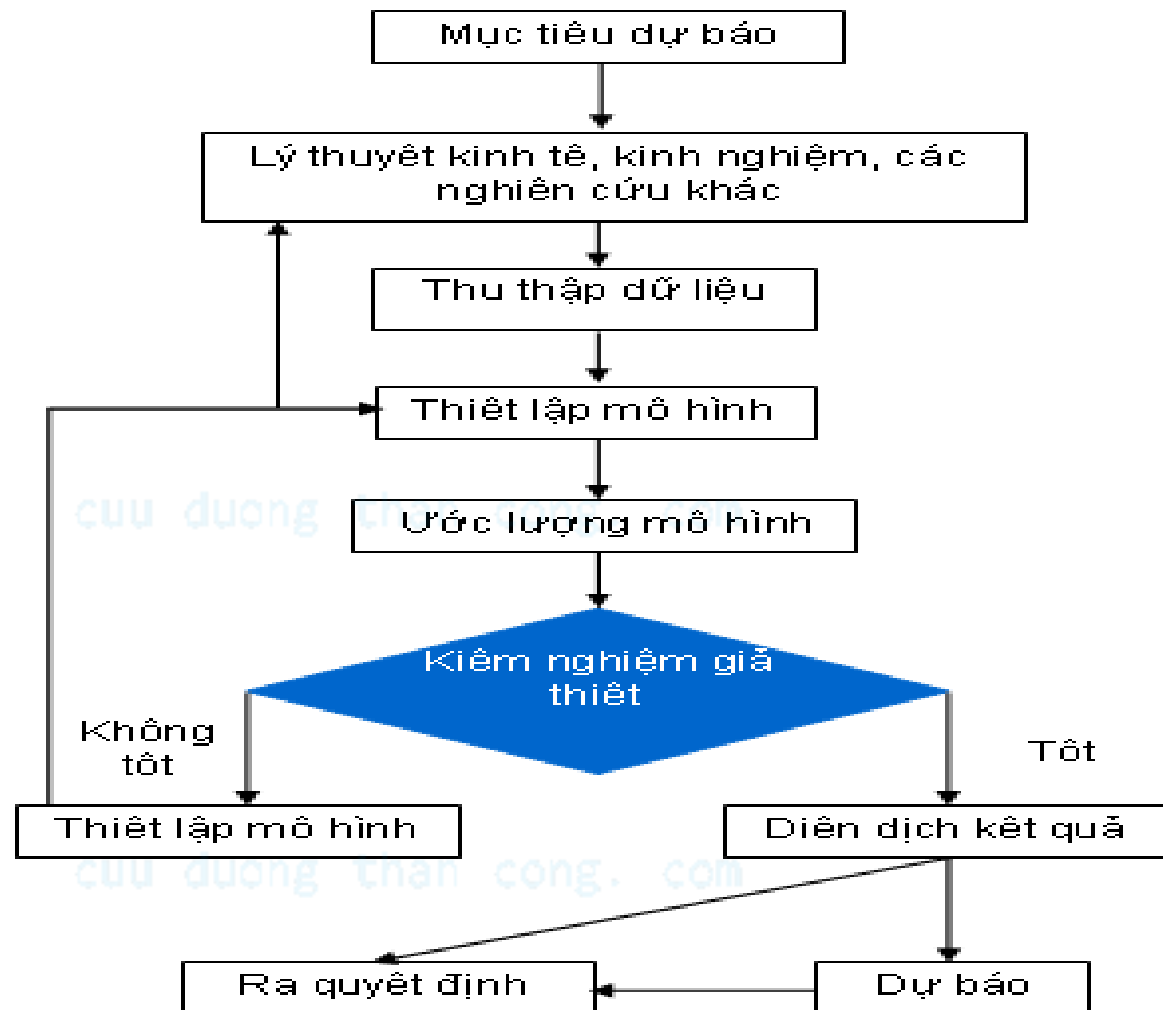
Chương 6

DỰ BÁO BẰNG MÔ HÌNH

NHÂN TỐ

cuu duong than cong. com

Phương pháp luận dự báo nhân quả



Nguồn: Ramanathan, 1998

1. TƯ TƯỞNG CƠ SỞ CỦA PHƯƠNG PHÁP:

Phương pháp này dựa trên quan hệ giữa nguyên nhân và kết quả để dự báo trạng thái của đối tượng dự báo trong tương lai.

Mối quan hệ nhân quả được xác lập trên cơ sở mô hình hồi quy của biến phụ thuộc (biến được giải thích) theo các biến độc lập (hay biến giải thích).

Mô hình nhân tố được sử dụng phổ biến trong dự báo với tư cách là dự báo có điều kiện.

Mô hình nhân tố được tiến hành trên cơ sở kỹ thuật kinh tế lượng.

Nếu ký hiệu y_t là biến phụ thuộc (hay còn gọi là biến được giải thích, biến nội sinh);

x_{it} là biến độc lập (hay còn gọi là biến giải thích, biến ngoại sinh)

Với $i = 1, 2, \dots, m$ (m là số biến giải thích trong mô hình nhân tố) thì mô hình nhân tố có dạng tổng quát như sau:

$$y_t = f(x_{1t}, x_{2t}, \dots, x_{mt}) + u_t \quad t = 1, 2, \dots, n$$

Tùy theo kỹ thuật xây dựng mô hình, có thể chia mô hình nhân tố thành: mô hình một nhân tố và mô hình đa nhân tố.

2. Dự báo theo MÔ HÌNH MỘT NHÂN TỐ:

Là mô hình chỉ có 1 biến giải thích:

$$y = f(x) + u$$


Trong đó: y là chỉ tiêu cần dự báo

x là nhân tố ảnh hưởng

➤ Quy trình thực hiện:

c 1: Xác định nhân tố ảnh hưởng, với tiêu chuẩn:

- Đảm bảo mối quan hệ kinh tế - kỹ thuật
- Tác động mạnh nhất đến hiện tượng dự báo
- Định lượng được và có đủ thông tin dữ liệu.



c 2: Xác định dạng hàm phản ánh mối quan hệ
(có thể bằng biểu diễn đồ thị, phân tích thống kê...)

Bước 3: Thu thập và xử lý sơ bộ thông tin, số liệu.

c 4: Xây dựng mô hình dự báo bằng kỹ thuật
hồi quy OLS.

Bước 5: Kiểm tra mô hình

Bước 6: Dự báo và khoảng dự báo.

Mô hình một nhân tố

- Dạng mô hình:

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

- Hệ phương trình chuẩn:

$$\begin{cases} n \hat{a}_0 + \hat{a}_1 \sum x = \sum y \\ \hat{a}_0 \sum x + \hat{a}_1 \sum x^2 = \sum y \cdot x \end{cases}$$

- Ý nghĩa các hệ số hồi quy.

Kiểm tra mô hình

- Kiểm định các hệ số hồi quy:
- Kiểm định t-test: Hệ số có ý nghĩa thống kê:

— *2 phía* : $|t| > t_{\alpha/2; (n-2)}$

— *Bên phải* : $t > t_{\alpha; (n-2)}$

— *Bên trái* : $t < -t_{\alpha; (n-2)}$

- Tính P-value = $P(|t_{\alpha/2}| > |t|)$; Hệ số có ý nghĩa khi P-value $< \alpha$ hoặc: P-value/2 $< \alpha$ (kiểm định một bên).
- Tính t: =TINV(xác suất, bậc tự do);
- P-value: =TDIST($|t|$, bậc tự do, đuôi 1 hoặc 2)

c 4: Đánh giá mô hình

- Tính độ xác định:

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS} = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

$R \in [-1, 1]$: mức độ giải thích được của biến độc lập đối với sự thay đổi của biến phụ thuộc.

- Kiểm định sự phù hợp của mô hình:

Tính: $F = \frac{R^2}{1 - R^2} * (n - 2)$ Tra bảng: $F_{(\alpha; 1, n - 2)}$

Nếu: $F > F_{(\alpha; 1, n - 2)}$: Mô hình phù hợp.

Sai số dự báo

- Sai số dự báo và khoảng dự báo trung bình:

$$\Delta_1 = t_{\alpha/2; (n-2)} \sqrt{\frac{S_u^2}{n-2} \left(\frac{1}{n} + \frac{\bar{x}_{tp} - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}; \quad S_u^2 = RSS = \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

$$\hat{y}_{x_p} - \Delta_1 \leq Y \leq \hat{y}_{x_p} + \Delta_1$$

- Sai số dự báo và khoảng dự báo cá biệt:

$$\Delta_2 = t_{\alpha/2; (n-2)} \sqrt{\frac{S_u^2}{n-2} \left(1 + \frac{1}{n} + \frac{\bar{x}_{tp} - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}; \quad S_u^2 = RSS = \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

$$\hat{y}_{x_p} - \Delta_2 \leq Y \leq \hat{y}_{x_p} + \Delta_2$$

Mô hình tổng quát

- Phương trình và Hệ phương trình chuẩn:

$$Y = a_0 + a_1 x + a_2 x^2 + \dots + x^p$$

$$na_0 + a_1 \sum x + a_2 \sum x^2 + \dots + a_p \sum x^p = \sum Y$$

.....

$$a \sum x^p + a_1 \sum x^{p+1} + a_2 \sum x^{p+2} + \dots + a_p \sum x^{2p} = \sum Y . x^p$$

Các dạng mô hình một nhân tố khác

- Mô hình Logarit kép (log-log) $Y = a_0 x^{a_1}$
- Mô hình Log – Lin (linear) như: tăng trưởng dân số, GNP, Mức cung tiền tệ, thất nghiệp,...
- Mô hình tuyến tính- logarit: $Y_t = Y_0 (1 + r)^t$
(Lượng cung tiền ảnh hưởng tới GDP, Diện tích trồng trọt ảnh hưởng đến sản lượng, diện tích căn nhà tác động tới giá nhà...)
- Mô hình nghịch đảo: $Y_t = a_0 + a_2 \frac{1}{X_t}$
(Sản lượng đến chi phí cố định trung bình, Tỷ lệ thất nghiệp đến tiền lương, thu nhập lên chi tiêu một loại mặt hàng,...)

$$Y_t = a_0 + a_1 \ln X_t$$

3. MÔ HÌNH ĐA NHÂN TỐ:

- Mô hình tập trung vào 2 dạng cơ bản:

$$y_t = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m + u_t$$

$$\text{hoac} : y_t = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} + u_t$$

- Mô hình này sử dụng trong dự báo với ý nghĩa:
 - Để ngoại suy giá trị Y theo giá trị X_0 . Giá trị X_0 cho trước hoặc phải dự báo.
 - Để mô phỏng theo các kịch bản khác nhau thông qua mô hình (tức là đưa ra các điều kiện để đạt được kết quả cho trước). Thực chất là thay các giá trị biến đầu vào X để kiểm tra Biến phụ thuộc đầu ra.

Quy trình thực hiện

Bước 1: Lựa chọn nhân tố, thu thập số liệu.

c 2: Xác định dạng hàm phản ánh mối quan hệ giữa đối tượng dự báo và nhân tố ảnh hưởng

c 3: Xây dựng mô hình dự báo:

- dạng tuyến tính: $\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \dots + \hat{a}_p x_p$

Các tham số \hat{a}_i ($i = 0, 1, 2, \dots, p$) được ước lượng bằng phương pháp OLS.

Bước 4: Đánh giá mô hình

Bước 5: Dự báo và khoảng dự báo

Ước lượng tham số của mô hình

- Hệ phương trình chuẩn:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 + \dots + a_p \sum x_p = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + \dots + a_p \sum x_p x_1 = \sum Y \cdot x_1 \\ \dots \dots \dots \\ a_0 \sum x_p + a_1 \sum x_p x_1 + \dots + a_p \sum x_p^2 = \sum Y \cdot x_p \end{cases}$$

cuu duong than cong. com

Phân tích và đánh giá mô hình

- Hệ số xác định:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{\sum (y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2}; TSS = \sum (Y_i - \bar{Y})^2$$

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2; RSS = TSS - ESS$$

Ý nghĩa: R^2 đo tính phù hợp của mô hình, các biến độc lập giải thích được bao nhiêu % sự thay đổi của biến phụ thuộc.

- Hệ số xác định hiệu chỉnh: $R^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$

Ý nghĩa: Căn cứ để xác định việc đưa thêm biến độc lập vào mô hình.

Đánh giá mô hình

- ✓ Kiểm định hệ số hồi quy:

$$t = \frac{\hat{a}_j}{se(\hat{a}_j)} \sim t_{\frac{\alpha}{2}; (n-k)}$$

Ý nghĩa: Hệ số hồi quy có ý nghĩa thống kê.

- ✓ Kiểm định giả thiết đồng thời bằng 0: $F = \frac{R^2(n-k)}{2(1-R^2)} \sim F_{\alpha; (2, n-k)}$

Ý nghĩa: Các hệ số HQ không đồng thời = 0

- ✓ Khoảng tin cậy của hệ số HQ: $\hat{a}_j \pm t_{\frac{\alpha}{2}; (n-k)} \cdot se(\hat{a}_j)$

Kiểm định cấu trúc(ổn định)của Mô hình (Chow test)

Chuỗi thời gian có n quan sát, chia thành 2 mẫu nhỏ có n_1 và n_2 quan sát ($n = n_1 + n_2$) .

- ✓ Bước 1: Ước lượng mô hình n quan sát, có RSS_R
- ✓ Bước 2: Ước lượng mô hình có n_1 QS, có RSS_1
- ✓ Bước 3: Ước lượng mô hình n_2 QS, có RSS_2 .
- ✓ Bước 4: Tính $RSS_u = RSS_1 + RSS_2$
- ✓ Bước 5: Tính:
$$F = \frac{(RSS_R - RSS_u) / k}{(RSS_u) / (n - 2k)}$$

- Nếu: $F > F_{\alpha; (k-1; n-k)}$ thì mô hình ở hai giai đoạn khác nhau.
- Tính đa cộng tuyến: Hệ số tương quan giữa 2 biến $> 0,8$ hoặc R^2 cao $> 0,8$ nhưng t-test thấp. Khắc phục là bỏ đi một biến .

Dự báo

- Dự báo giá trị trung bình:

$$\hat{y}_0 = \hat{a}_0 + \hat{a}_1 x_1^0 + \hat{a}_2 x_2^0 + \dots + \hat{a}_p x_p^0$$

$$\hat{y}_0 - t_{\frac{\alpha}{2}; (n-k)} \cdot \widehat{se}(Y_0) < Y_0 < \hat{y}_0 + t_{\frac{\alpha}{2}; (n-k)} \cdot \widehat{se}(Y_0)$$

$$\widehat{Var}(\hat{y}_0) = \sigma^2 (X^0)^T (X^T X)^{-1} X^0; \widehat{se}(Y_0) = \sqrt{\widehat{Var}(\hat{y}_0)}$$

- Dự báo giá trị cá biệt:

$$\hat{y}_0 - t_{\frac{\alpha}{2}; (n-k)} \cdot \widehat{se}(Y_0 - \hat{Y}_0) < Y_0 < \hat{y}_0 + t_{\frac{\alpha}{2}; (n-k)} \cdot \widehat{se}(Y_0 - \hat{Y}_0)$$

$$\widehat{Var}(Y_0 - \hat{Y}_0) = \widehat{Var}(\hat{Y}_0) + \sigma^2$$

$$\widehat{se}(Y_0 - \hat{Y}_0) = \sqrt{\widehat{Var}(Y_0 - \hat{Y}_0)}$$

c 4: Đánh giá mô hình

- Tính hệ số tương quan bội

Thiết lập ma trận hệ số tương quan: nếu có m biến thì có ma trận cấp (m+1)

$$H = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1m} & r_{10} \\ r_{21} & r_{22} & \dots & r_{2m} & r_{20} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mm} & r_{m0} \\ r_{10} & r_{20} & \dots & r_{m0} & 0 \end{pmatrix}$$

$$H' = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mm} \end{pmatrix}$$

$$r_{i0} = \frac{\overline{y x_i} - \overline{y} \overline{x_i}}{\delta_y \delta_{x_i}}$$

r_{i0} thể hiện mối quan hệ giữa y và x_i

$$r_{ij} = \frac{\overline{x_i x_j} - \overline{x_i} \overline{x_j}}{\delta_{x_i} \delta_{x_j}}$$

r_{ij} thể hiện mối quan hệ giữa x_i và x_j

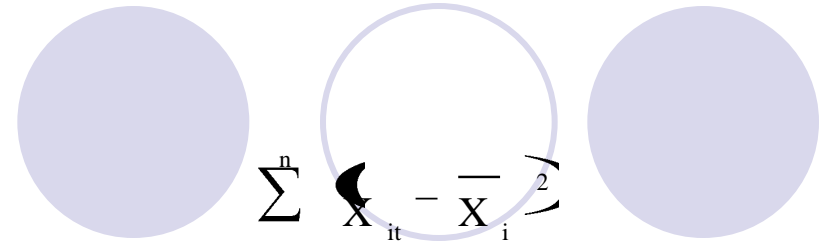
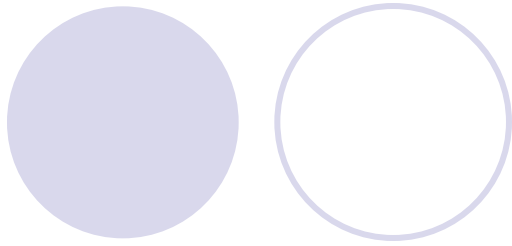
$$\delta_y^2 = \frac{\sum (y_t - \bar{y})^2}{n}$$

$$\delta_{x_i}^2 = \frac{\sum (x_{it} - \bar{x}_i)^2}{n}$$

$$\longrightarrow R = \sqrt{\frac{\det H}{\det L}}$$

- Sai số dự báo:

$$\Delta = t_{\alpha}(n) \sqrt{\frac{S_u^2}{n-m-1} \left[1 + \sum_{q,u} L_{qu} \frac{(x_{qtp} - \bar{x}_q)(x_{utp} - \bar{x}_u)}{\det L} \right]}$$



Trong đó:

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \dots & \dots & \dots & \dots \\ l_{m1} & l_{m2} & \dots & l_{mm} \end{pmatrix}$$

$$l_{ii} = \frac{\sum_{t=1}^n (X_{it} - \bar{X}_i)^2}{n}$$

$$l_{ij} = \frac{\sum_t (X_{it} - \bar{X}_i)(X_{jt} - \bar{X}_j)}{n}$$

$X_{u(t,p)}$ $x_{qt(p)}$ là giá trị biến u, q tại thời điểm dự báo
 L_{qu} là phần bù đại số của phần tử l_{qu} trong ma trận L
 m là số nhân tố đưa vào mô hình
 $n-m-1$ là số bậc tự do

- Khoảng dự báo:

$$\hat{y}_{x_{tp}} - \Delta \leq y^* \leq \hat{y}_{x_{tp}} + \Delta$$