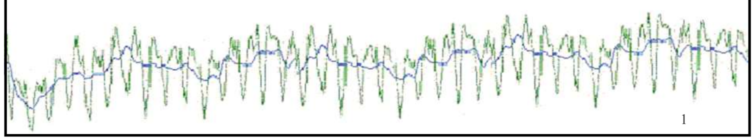
CHƯƠNG 1.
KINH TẾ LƯỢNG TRONG TÀI CHÍNH – NGÂN HÀNG

Kinh tế lượng ứng dụng

Khoa Kinh tế quốc tế - Đại học Ngân Hàng Tp.HCM

Ths. Đỗ Hoàng Oanh



Mô tả môn học

Môn học cung cấp cho sinh viên kiến thức về các phương pháp và mô hình kinh tế lượng được ứng dụng trong nghiên cứu thực nghiệm chuyên ngành tài chính ngân hàng, gồm: mô hình chuỗi thời gian đơn biến, mô hình phương sai có điều kiện thay đổi, các phương pháp cho dữ liệu bảng.

Môn học được xây dựng nhằm giới thiệu những mô hình cụ thể, các kỹ thuật cần thiết để thực hiện ước lượng, kiểm định, dự báo, cách thực hiện các kỹ thuật ước lượng trên các nghiên cứu điển hình.

Các bài thực hành trên bộ dữ liệu thực tế được cấu trúc đi kèm bài giảng nhằm đảm bảo sinh viên có thể hiểu và ứng dụng vào nghiên cứu.

Tài liệu tham khảo cho nội dung

- **Phạm Thị Tuyết Trinh, *Kinh tế lượng ứng dụng trong kinh tế và tài chính,* Khoa Kinh tế quốc tế, 2016**
- Brooks, Chris, *Introductory econometrics for finance*, 2nd edition, Cambridge University Press, 2008.
- Nguyễn Trọng Hoài và các tác giả, *Dự báo và phân tích dữ liệu trong kinh tế và tài chính*, NXB Thống Kê. 2009.
- Nguyễn Quang Dong, *Kinh tế lượng*, NXB Đại học Kinh tế quốc dân, 2012.

Sinh viên đọc kỹ nội dung:

1. Giảng viên gửi đến lớp: bài giảng + ebook + giới thiệu TLTK + data vào tuần đầu tiên.

2. Theo quy chế tín chỉ:

Học trên lớp 1 tiết + Tự học 3 tiết tại nhà

Những nội dung giảng viên không giảng thì sinh viên tự học, tự nghiên cứu và có quyền trao đổi với giảng viên trong buổi giảng tiếp theo.

Sv mang theo laptop đã cài Eviews mỗi buổi học

Sinh viên đọc kỹ nội dung:

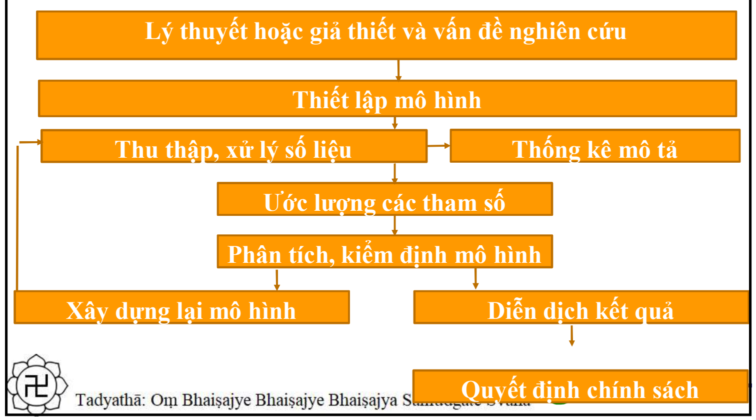
3. Thi giữa kỳ 40% không sử dụng tài liệu. Lớp có quyền chọn 1 trong 2 hình thức thi giữa kỳ:

- 1 bài thi viết trên lớp 40%.
- 20% thi viết (1 bài trên lớp) + 20% (1 tiểu luận tại nhà)

Lưu ý: Bài thi viết giữa kỳ trên lớp **chỉ tổ chức duy nhất 1 lần**. Lớp quyết định hình thức nào thì sau đó **không** được thay đổi vì bất kỳ lý do gì. Suy nghĩ thật kỹ !

- Cuối kỳ (60%): trắc nghiệm, không sử dụng tài liệu

Quy trình Nghiên cứu kinh tế lượng ứng dụng trong Tài chính – Ngân hàng



Dữ liệu chéo, thời gian và bảng

1) Số liệu chuỗi thời gian

Là tập hợp giá trị quan sát được từ **một thực thể** ở các khoảng **thời gian khác nhau** (theo định kỳ) như hàng ngày (ví dụ giá cổ phiếu), hàng tuần (ví dụ số liệu về lượng phát hành của một đĩa nhạc), hàng tháng (chỉ số giá tiêu dùng)

2) Số liệu chéo

Số liệu chéo là tập hợp giá trị của một hoặc **nhiều thực thể** từ nhiều đối tượng ở **cùng một thời điểm**

(3) Số liệu bảng

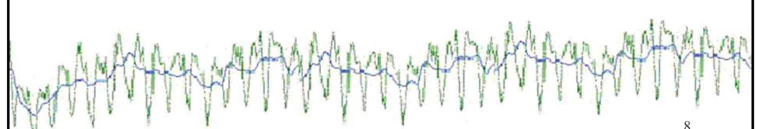
Số liệu bảng là sự kết hợp của cả hai chiều thời gian (số liệu chuỗi thời gian) và không gian (số liệu chéo).

CHƯƠNG 2. MÔ HÌNH CHUỖI THỜI GIAN ĐƠN BIẾN

Kinh tế lượng ứng dụng

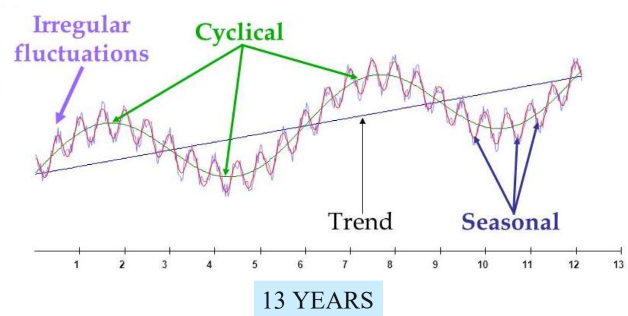
Khoa Kinh tế quốc tế Đại học Ngân Hàng Tp.HCM

Đỗ Hoàng Oanh



Nội dung

2.1. Ôn tập chung về chuỗi thời gian:



2.1. Ôn tập chung về chuỗi thời gian (tt):

Khái niệm chuỗi thời gian: Chuỗi các quan sát được thu thập trên cùng một đối tượng tại các mốc thời gian khác nhau được gọi là chuỗi thời gian.

Các chuỗi thời gian thường được mô hình hóa dạng tổng (hay tích) của 4 thành phần: (1) xu hướng (2) chu kỳ (3) mùa và (4) thành phần ngẫu nhiên. Do đó ta có :

$$Y_t = T_t C_t S_t u_t \qquad Y_t = T_t + C_t + S_t + u_t$$

Với Tr (Trend - Xu thế), Cl (Cyclical - Chu kỳ), Sn (Seasonal - Mùa), Ir (Irregular - Ngẫu nhiên/bất thường)

2.1. Ôn tập chung về chuỗi thời gian (tt):

- **Xu thế (Trend)** là thành phần thể hiện sự tăng (hoặc giảm) ẩn bên trong của một chuỗi thời gian. Thường được ký hiệu là **T**.
- **Chu kỳ (Cyclical)** là thành phần thể hiện sự tăng (hoặc giảm) ẩn bên trong của thời kỳ. Thông thường là trên 1 năm. Người ta thường ký hiệu thành phần chu kỳ là **C**.

2.1. Ôn tập chung về chuỗi thời gian (tt):

- **Mùa (Seasonal) :** Những dao động mùa vụ rất thường được tìm thấy với dữ liệu theo quý, theo tháng, hoặc thậm chí theo tuần. Nếu chỉ có dữ liệu theo năm thì không có biến động mùa. Sự dao động mùa vụ liên quan đến kiểu thay đổi khá ổn định xuất hiện hàng năm và kiểu thay đổi đó được lặp lại ở năm sau và các năm sau nữa. Thường được ký hiệu là **S**.
- **Ngẫu nhiên/bất thường (Irregular)** : Gồm những thay đổi ngẫu nhiên hay không dự đoán trước được. Những sự thay đổi bất thường là kết quả của vô số những sự kiện mà nếu xét riêng lẻ thì không quan trọng gì, còn nếu kết hợp các sự kiện riêng lẻ đó lại thì có thể tạo ra một ảnh hưởng lớn. Thành phần này được ký hiệu là **u** hay **I**.

2.1. Ôn tập chung về chuỗi thời gian (tt):

Khái niệm chuỗi thời gian: Chuỗi các quan sát được thu thập trên cùng một đối tượng tại các mốc thời gian khác nhau được gọi là chuỗi thời gian.

Các chuỗi thời gian thường được mô hình hóa dạng tổng (hay tích) của 4 thành phần: (1) xu hướng (2) chu kỳ (3) mùa và (4) thành phần ngẫu nhiên. Do đó ta có :

$$Y_t = T_t C_t S_t u_t \qquad Y_t = T_t + C_t + S_t + u_t$$

Với Tr (Trend - Xu thế), Cl (Cyclical - Chu kỳ), Sn (Seasonal - Mùa), Ir (Irregular - Ngẫu nhiên/bất thường)

Khi nghiên cứu các thành phần của một chuỗi thời gian. Nhà phân tích phải xem xét các thành phần này liên quan như thế nào với chuỗi dữ liệu gốc (biến Y).

Có 2 mô hình thể hiện mối quan hệ này:

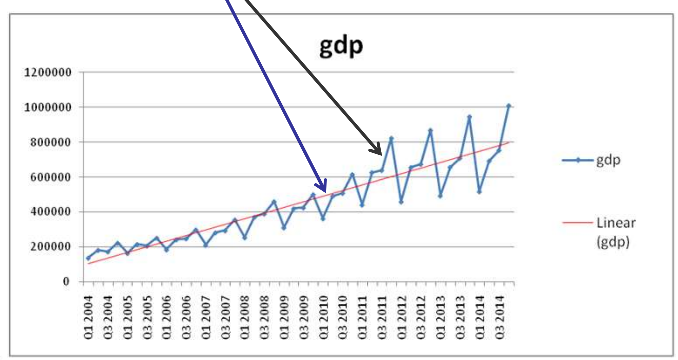
- **Mô hình nhân tính (Multiplicative Components Model)** xem các giá trị của một chuỗi thời gian (biến Y) được tạo thành bởi tích số của từng thành phần Tr, Cl, Sn, Ir.

  Mô hình nhân tính: $Y_t = Tr_t . Cl_t . Sn_t . Ir_t$

- **Mô hình cộng tính (Additive Components Model)** xem các giá trị của một chuỗi thời gian (biến Y) được tạo thành bởi tổng các giá trị của các thành phần Tr, Cl, Sn, Ir.

  Mô hình cộng tính: $Y_t = Tr_t + Cl_t + Sn_t + Ir_t$

Phân biệt tính mùa và tính xu hướng trong hình



2.2. Giới thiệu chung về các mô hình chuỗi thời gian đơn biến:

Được biết rộng rãi dưới cái tên phương pháp Box-Jenkins (BJ), nhưng về mặt kỹ thuật được gọi là phương pháp ARIMA.

Không giống như các mô hình hồi quy trong đó Yt được giải thích bởi k biến làm hồi quy X_1, X_2,..., X_k, trong các mô hình chuỗi thời gian kiểu BJ, Y_t có thể được giải thích bởi các giá trị trong quá khứ hay giá trị trễ của bản thân biến Y và các sai số ngẫu nhiên.

Trọng tâm của phương pháp dự báo này là phân tích các tính chất xác suất hay ngẫu nhiên của bản thân các chuỗi thời gian kinh tế theo cách "tự mình giải thích cho chính mình" (bản thân chuỗi thời gian này dự báo cho chính chuỗi thời gian này).

2.2. Giới thiệu chung về các mô hình chuỗi thời gian đơn biến:

Mô hình chuỗi thời gian đơn biến (univariate time series model) được sử dụng khi hành vi của biến số cần giải thích được quyết định bởi những thông tin về giá trị của chính nó trong quá khứ và/hoặc giá trị hiện tại và quá khứ của hạng nhiễu. Ý tưởng của mô hình này dựa trên giả định những yếu tố xảy ra trong quá khứ sẽ tiếp tục lặp lại cho hiện tại và tương lai. Theo đó, các nhà nghiên cứu sẽ dựa trên các vận động và biến động của chính yếu tố cần dự báo để giải thích cho chính yếu tố này.

Phương pháp trung bình trượt kết hợp tự hồi quy (Autoregressive integrated moving average - ARIMA)

Để dự báo theo phương pháp ARIMA thì chuỗi thời gian phải là một chuỗi dừng. Vì vậy, phần này sẽ đi tìm hiểu:

(1) Làm thế nào để lập mô hình một chuỗi thời gian dừng?

(2) Làm thế nào sử dụng mô hình thích hợp cho mục đích dự báo?

2.2.1. Chuỗi dừng (Stationarity)

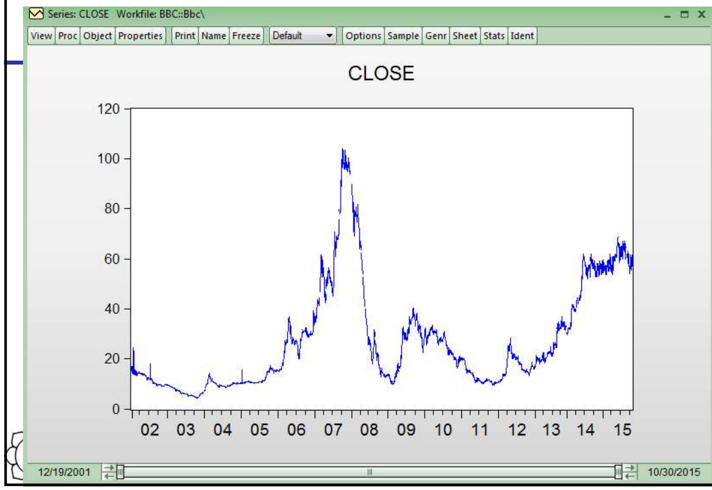
Một chuỗi dữ liệu thời gian được xem là dừng nếu như trung bình và phương sai của phương trình không thay đổi theo thời gian và giá trị của đồng phương sai giữa hai đoạn chỉ phụ thuộc vào khoảng cách hay độ trễ về thời gian giữa hai thời đoạn này chứ không phụ thuộc vào thời điểm thực tế mà đồng phương sai được tính (Ramanathan, 2002)

$$E(Y_t) = \mu = const$$

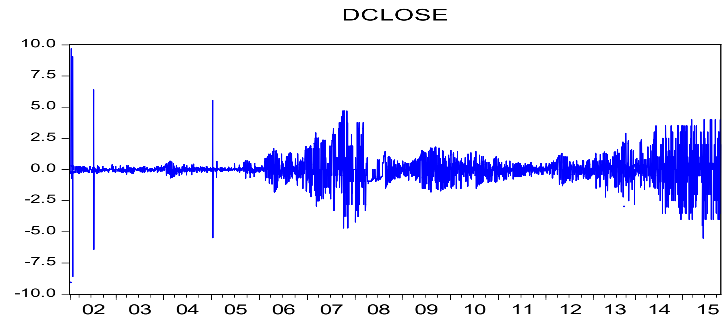
$$Var(Y_t) = \sigma^2 = const$$

$$Cov(Y_t, Y_{t-k}) = \gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)]$$

Giá đóng cửa của cổ phiếu Bibica 19/12/2001 – 30/10/2015
Chuỗi thời gian không dừng



Giá đóng cửa của cổ phiếu Bibica 19/12/2001 – 30/10/2015
Chuỗi thời gian đã được thực hiện thành chuỗi dừng



Nhiễu trắng (White noise)

Chuỗi thời gian có kỳ vọng bằng 0, phương sai không đổi và hiệp phương sai bằng 0 được gọi là nhiễu trắng (White noise) thường được ký hiệu là ε_t

(1) $E(\varepsilon_t) = 0$

(2) $Var(\varepsilon_t) = \sigma^2$

(3) $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0, \; t \neq s$

Kiểm định tính dừng

Để kiểm định tính dừng của chuỗi thời gian, người ta thường dùng hai cách:

- Kiểm định tính dừng dựa vào biểu đồ tương quan
- Kiểm định nghiệm đơn vị đối với tính dừng

Kiểm định nghiệm đơn vị (Unit Root Test)

Kiểm định nghiệm đơn vị là một kiểm định được sử dụng khá phổ biến để kiểm định một chuỗi thời gian dừng hay không dừng. Giả sử ta có phương trình tự hồi quy như sau:

Auto regressive function:

AR(1): $Y_t = \rho Y_{t-1} + u_t$ với u_t ~iid $N(0; \sigma^2)$ u_t là nhiễu trắng

Nếu $\rho < 1$: chuỗi dừng

Nếu $\rho = 1$: chuỗi có nghiệm đơn vị (chuỗi không dừng)

Nếu $\rho > 1$: chuỗi bị bùng nổ (explosive series)

Kiểm định nghiệm đơn vị (Unit Root Test) (tt)

AR(1): $Y_t = \rho Y_{t-1} + u_t$ với u_t ~iid $N(0; \sigma^2)$ u_t là nhiễu trắng

Giả thiết: $H_0: \rho = 1$ Y_t (là chuỗi không dừng)

$H_1: \rho < 1$ Y_t (là chuỗi dừng)

Phương trình: $Y_t = \rho Y_{t-1} + u_t$

Tương đương với : $Y_t - Y_{t-1} = \rho Y_{t-1} + u_t - Y_{t-1} = (\rho - 1) Y_{t-1} + u_t$

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

Kiểm định nghiệm đơn vị (Unit Root Test) (tt)

Như vậy, các giả thiết ở trên có thể được viết lại như sau:

$H_0: \delta = 0$ Y_t là chuỗi có nghiệm đơn vị, chuỗi không dừng

$H_1: \delta < 0$ Y_t là chuỗi dừng

Dickey and Fuller cho rằng giá trị t ước lượng của hệ số Y_{t-1} sẽ theo phân phối xác suất τ (τ= giá trị δ ước lượng / sai số của hệ số δ). Kiểm định thống kê τ còn gọi kiểm định Dickey – Fuler (DF)

Kiểm định DF được ước lượng với ba hình thức :

Khi là một bước ngẫu nhiên không hằng số (Without Constant and trend)

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

Khi là một bước ngẫu nhiên có hằng số (Without Constant)

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$$

Khi Y_t là một bước ngẫu nhiên có hằng số xoay quanh một đường xu thế ngẫu nhiên.

$$\Delta Y_t = \beta_1 + \beta_2 @ trend + \delta Y_{t-1} + u_t$$

Để kiểm định H_0, so sánh giá trị thống kê τ tính toán với giá trị thống kê τ tra bảng DF

Nếu số hạng sai số u_t là tự tương quan, ta sẽ biến đổi phương trình trên thành:

$$\Delta Y_t = \beta_1 + \beta_2 @ \, trend + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + \varepsilon_t \quad (*)$$

Giả thuyết không vẫn là $H_0: \delta = 0$ hoặc $H_0: \rho = 1$ có nghĩa là Y có nghiệm đơn vị, (Y là không dừng).

Khi kiểm định DF được áp dụng cho các mô hình như (*) nó được gọi là kiểm định Dickey-Fuller mở rộng (Augmented Dickey-Fuller (ADF) test). Trị thống kê của kiểm định ADF có cùng một phân bổ tiệm cận giống như của trị thống kê DF, do vậy có thể sử dụng cùng các giá trị tới hạn giống nhau (Cao Hào Thi,2011).

2.2.3 Kiểm định nghiệm đơn vị (Unit Root Test) (tt)

Khi đó :

Nếu $|\tau_\alpha|$ tính toán > $|\tau|$ giá trị ADF (ADF test statistic) suy ra bác bỏ giả thiết H_0 (tồn tại nghiệm đơn vị) => chuỗi dữ liệu không dừng

Nếu $|\tau_\alpha|$ tính toán < $|\tau|$ giá trị ADF (ADF test statistic) suy ra không có cơ sở bác bỏ giả thiết H_0, hay không tồn tại nghiệm đơn vị => chuỗi dữ liệu là chuỗi dừng

Hàm tự tương quan và hàm tự tương quan riêng phần

- Hàm tự tương quan (autocorrection function, ACF) hoặc biểu đồ tương quan (correlogram) là biểu đồ có được khi vẽ các hệ số tự tương quan (τ_s) tại các $s = 0, 1, 2$...
- Hàm tự tương quan riêng phần (Partial autocorrelation, PACF) là biểu đồ có được khi vẽ các hệ số tự tương quan riêng phần, τ_{kk}, lần lượt tại $kk = 1, 2, 3$...

Do tồn tại tương quan trực tiếp giữa Y_t và Y_{t-s} ($s \le p$) và không tồn tại tương quan trực tiếp giữa Y_t và Y_{t-s} ($s > p$), PACF thường có hệ số tự tương quan riêng phần khác 0 đối với các bậc trễ nhỏ hơn hoặc bằng bậc trễ của mô hình ($\tau_{kk} \ne 0$, $kk \le p$) và có hệ số tương quan riêng phần bằng 0 đối với các bậc trễ lớn hơn bậc trễ của mô hình ($\tau_{kk} = 0$, $kk > p$).

Sai phân

- **Để thực hiện được ước lượng và dự báo với chuỗi thời gian, thì chuỗi thời gian phải là một chuỗi dừng.**
- Do đó, nếu chuỗi không dừng thì ta lấy sai phân cho đến khi chuỗi dừng

	$GDPR_t$	$GDPR_{t-1}$	$\Delta GDPR = DGDPR = GDP_t - GDP_{t-1}$
2004	67.8912		
2005	73.6225	67.8912	5.7312
2006	78.7598	73.6225	5.1373
2007	84.3745	78.7598	5.6152
2008	89.1521	84.3745	4.7771
2009	93.9644	89.1521	4.8123
2010	100	93.9644	6.0356
2011	106.2403	100	6.2403
2012	111.8151	106.2403	5.5748

Tóm tắt về kiểm định chuỗi dừng

- **Nên sử dụng kiểm định nghiệm đơn vị** để quyết định chuỗi có dừng hay không
- **Theo kiểm định nghiệm đơn vị, chuỗi dừng phải đạt được 2 yếu tố:**
  - **δ < 0** tức là ρ phải nhỏ hơn 1
  - Nếu $|\tau_\alpha|$ tính toán < $|\tau|$ giá trị ADF (ADF test statistic) suy ra **không có cơ sở bác bỏ giả thiết H_0** với H_0 tồn tại nghiệm đơn vị => **chuỗi dữ liệu là chuỗi dừng**
- **Nếu như chuỗi không dừng thì sai phân cho thành chuỗi dừng. Sau khi chuỗi dừng thì tiến hành chạy mô hình ARIMA**

Mô hình ARIMA

- *Mục đích của ARIMA là gì?*
- *Mô Hình Tự Hồi Quy Bậc p-AR(p)*

Trong mô hình tự hồi qui quá trình phụ thuộc vào tổng có trọng số của các giá trị quá khứ và số hạng nhiễu ngẫu nhiên.

$$AR(1): Y_t = \phi_0 + \phi_1 Y_{t-1} + u_t$$

$$AR(p): Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + u_t$$

Trong đó u_t là nhiễu trắng

Mô hình ARIMA

- Dạng toán tử lùi, mô hình Tự Hồi Quy Bậc p-AR(p)

$$AR(p): Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + u_t$$

Điều kiện cần để AR(p) dừng thì

$$\left|\sum_{i=1}^{p} \phi_i\right| < 1, \text{gọi } AR(p) \text{khả nghịch}$$

Trong đó u_t là nhiễu trắng

- Mô Hình Trung Bình Trượt Bậc q – MA(q)

Trong Mô hình trung bình trượt, quá trình được mô tả hoàn toàn bằng tổng có trọng số của các ngẫu nhiên hiện hành có độ trễ.

$$MA(q): Y_t = \mu + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q} + u_t, t = 1, 2, \ldots, n.$$

$$MA(1): Y_t = \mu + \theta_1 u_{t-1} + u_t$$

Trong đó u_t là nhiễu trắng

- Mô Hình Trung Bình Trượt Bậc q – MA(q)

Toán tử lùi của:

$$MA(q): Y_t = \mu + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q} + u_t, t = 1, 2, \ldots, n.$$

Điều kiện để MA(q) khả nghịch:

$$\left|\sum_{i=1}^{p} \theta_i\right| < 1, \text{gọi } MA(q) \text{khả nghịch}$$

Trong đó u_t là nhiễu trắng

- Mô Hình Hồi Quy và Trung Bình Trượt -ARMA (p,q)

Tổng quát, Y_t là quá trình ARMA(p,q) nếu Y có thể biểu diễn dưới dạng :

$$Y_t = \theta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \theta_0 u_t + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q}$$

Trong đó u_t là nhiễu trắng.

- Mô Hình Tự Hồi Quy và Trung Bình Trượt -ARMA (p,q)

ARMA(1,1), nếu Y có thể biểu diễn dưới dạng

$$Y_t = \theta + \phi_1 Y_{t-1} + \theta_0 u_t + \theta_1 u_{t-1}$$

$$ARMA(2,2): Y_t = \theta + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \theta_0 u_t + \theta_1 u_{t-1} + \theta_1 u_{t-2}$$

$$ARMA(p,q):$$

$$Y_t = \theta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \theta_0 u_t + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q}$$

Trong đó u_t là nhiễu trắng.

Mô Hình Hồi Quy Kết Hợp Trung Bình Trượt -ARIMA (p,d,q)

Theo phương pháp Box-Jenkins, nếu ta muốn dự báo, ta phải giả thiết rằng mô hình không thay đổi theo thời gian. Vì vậy, mô hình ta tiến hành dự báo phải có tính dừng.
Nếu chuỗi gốc Y_t không dừng, ta tiến hành sai phân ΔY_t
Lúc này ta nói $Y_t \sim I(1)$
Nếu như chuỗi vẫn không dừng, ta lấy thêm một lần sai phân nữa trên sai phân này, và nếu chuỗi sai phân hai lần này dừng.
Ta gọi là chuỗi này là sai phân bậc 2: $Y_t \sim I(2)$
Vậy nếu chuỗi dừng bậc d, ta có: $Y_t \sim I(d)$
Lúc này, ta có mô hình ARIMA(p,d,q) từ mô hình ARMA(p,q)

AR(p) là trường hợp đặc biệt của ARIMA(p,d,q), khi d=0 và q=0.

MA(q) là trường hợp đặc biệt của ARIMA(p,d,q), khi d=0 và p=0.

ARIMA(2,1,2)- nghĩa là chuỗi Y_t có sai phân bậc 1 là chuỗi dừng, chuỗi sai phân dừng này có thể biểu diễn dưới dạng

ARIMA(2,1,2):

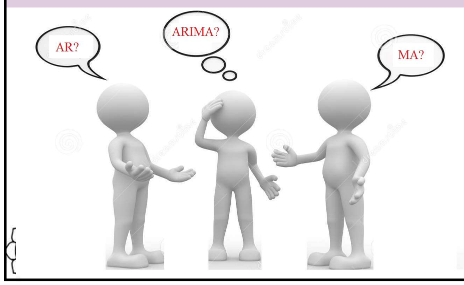
$$\Delta Y_t = \theta + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \theta_0 u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$$

Trong đó u_t là nhiễu trắng.

Như vậy nếu biết các tham số p,q,d, khi đó ta có thể mô hình hóa được chuỗi. Vấn đề đặt ra là xác định d,p,q và các tham số θ, ϕ

Câu hỏi đặt ra là:

- Làm cách nào để biết chuỗi thời gian mình nghiên cứu thuộc mô hình nào?
- Làm cách nào để dự báo?



QUY TRÌNH LỰA CHỌN VÀ THỰC HIỆN MÔ HÌNH ARIMA(p,d,q)

Bước 1 : Nhận dạng d mô hình ARIMA(p,d,q)

Tức là tìm các giá trị thích hợp của *p,d và q*.

Ta sẽ trình bày ngắn gọn biểu đồ **tương quan** (correlogram) và biểu đồ **tương quan riêng phần** (partial correlogram) hỗ trợ cho công việc này như thế nào.

QUY TRÌNH LỰA CHỌN VÀ THỰC HIỆN MÔ HÌNH ARIMA(p,d,q)

Bước 2a : Xác định SACF sample autocorrelation function:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_o} = SAC$$

$$\hat{\gamma}_k = E[(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})] = \frac{\sum (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{n} = Cov(Y_t, Y_{t-k})$$

$$\hat{\gamma}_o = E[(Y_t - \bar{Y})^2] = \frac{\sum (Y_t - \bar{Y})^2}{n} = Var(Y_t)$$

SACF có xu hướng giảm dần và tắt nhanh về 0 => chuỗi dừng

QUY TRÌNH LỰA CHỌN VÀ THỰC HIỆN MÔ HÌNH ARIMA(p,d,q)

Bước 2b : Xác định PACF:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + u_t$$

Với u_t là nhiễu trắng

ACF và PACF có tương quan với nhau

$PACF_1 = \rho_1 = \rho_{11} = corr(Y_t, Y_{t-1}) = ACF_1$

$PACF_2 = corr(Y_t, Y_{t-2}) = \rho_{22} = (\rho_2 - \rho_1^2)/(1-\rho_1^2)$

- Chọn mô hình AR(p) nếu đồ thị SPAC có giá trị cao tại độ trễ 1,2,…, p và giảm nhiều sau p và dạn hàm SAC giảm dần.

- Chọn mô hình MA(q) nếu đồ thị SAC có giá trị cao tại độ trễ 1,2,…,q và giảm nhiều sau q và dạng hàm SPAC giảm dần. Tóm lại:

Loại mô hình	**Dạng đồ thị SACF**	**Dạng đồ thị SPACF**
AR(p)	Giảm dần	Có đỉnh ở p
MA(q)	Có đỉnh ở q	Giảm dần
ARMA(p, q)	Giảm dần	Giảm dần

Nhận dạng mô hình trên lược đồ tương quan và tự tương quan

Dùng lược đồ tương quan và tự tương quan riêng là phương pháp hiệu quả để xác định p va q. Lược đồ vẽ ACF va PACF theo độ dài của độ trễ.

Đồng thời cũng vẽ đường phân giải chỉ khoảng tin cậy 95% cho giá trị bằng 0 của hệ số tự tương quan (Bartlett,1946) và hệ số tự tương quan riêng $(\pm 1{,}96 / \sqrt{n})$

Qua đó, nhận diện các hệ số tương quan và tự tương quan riêng khác không với mức ý nghĩa 5%.

Từ đó có thể đưa ra các đoán nhận chuỗi dừng, các giá trị p và q của các quá trình AR(p), MA(q).

Bước 3 : Ước lượng mô hình ARIMA

Sau khi đã nhận dạng các giá trị thích hợp của p, d và q.

Bước tiếp theo là ước lượng các tham số của các số hạng tự hồi quy và trung bình trượt trong mô hình.

Ta sử dụng phương pháp bình phương tối thiểu OLS

Bước 4 : Lựa chọn ARIMA phù hợp và Kiểm định chẩn đoán

Sau khi đã lựa chọn mô hình ARIMA cụ thể và ước lượng các tham số của nó, ta tìm hiểu xem mô hình lựa chọn có phù hợp với dữ liệu ở mức chấp nhận hay không bởi vì có thể một mô hình ARIMA khác cũng phù hợp với dữ liệu.

Một kiểm định đơn giản về mô hình lựa chọn là xem các phần dư ước lượng từ mô hình này có tính ngẫu nhiên thuần túy hay không (tức là có nhiễu trắng hay không).

Bước 4 : Lựa chọn ARIMA phù hợp và Kiểm định chẩn đoán (tt)

Ba tiêu chuẩn cơ bản:

1) **Mô hình ARIMA (p,d,q) phải vượt qua tất cả các kiểm định**
2) **Adjusted R^2 cao**
3) **Các chuẩn thông tin thấp (AIC, SBIC, HQIC nhỏ)**

B4.1. Các chuẩn thông tin:

- Tiêu chuẩn AIC (Akaike info criterion)

$$AIC = \left(\frac{RSS}{n}\right).e^{2k/n} \quad (13)$$

Trong đó k là số tham số của mô hình hồi quy. Giá trị AIC càng nhỏ chứng tỏ mô hình hồi quy càng phù hợp. Phần mềm EViews ước lượng giá trị AIC bằng biểu thức sau:

$$AIC = -\frac{2L}{n} + \frac{2k}{n} \quad (14)$$

- Tiêu chuẩn Schwarz (Schwarz criterion)

$$SC = \left(\frac{RSS}{n}\right)n^{k/n} \quad (15)$$

Giá trị SC càng nhỏ chứng tỏ mô hình hồi quy càng phù hợp. Phần mềm EViews ước lượng SC bằng công thức:

$$SC = -\frac{2L}{n} + \frac{k \log n}{n} \quad (16)$$

B4.2. Kiểm định giá trị nghịch đảo của nghiệm đặc trưng AR(p) và MA(q)

Các nghiệm đặc trưng đều nằm trong vòng tròn đơn vị. Hay nói cách khác, AR và MA đều khả nghịch

B4.2. Kiểm định tính thích hợp của mô hình

Ý nghĩa việc kiểm định u_t là nhiễu trắng là để mô hình đã đưa đầy đủ các thông tin đã biết trong quá khứ Y_{t-1}, Y_{t-2},…và u_{t-1}, u_{t-2},… và thật sự có ảnh hưởng đến Y_t vào mô hình.

$$E(u_t) = 0, E(u_s, u_t) = \begin{cases} \sigma^2; s = t \\ 0; s \neq t \end{cases}$$

Vẽ lược đồ tương quan SACF và SPACF cho phần dư.

Nếu SACF và SPACF không có thành phần có ý nghĩa thống kê thì u_t tương tự với nhiễu trắng. Khi đó mô hình có thể là chấp nhận được. Nếu không, ta phải lặp lại từ đầu

Như vậy, quá trình ước lượng mô hình ARIMA(p,d,q) là quá trình lặp lại nhiều lần.

Cách khác, dựa hệ số tự tương quan của các phần dư.

Kiểm định khi bình phương dựa trên thống kê Q của Box-Pierce:

$$H_0: \rho_{\varepsilon,1} = \rho_{\varepsilon,2} = ... = \rho_{\varepsilon,k} = 0;$$

H_1 có ít nhất $\rho_{\varepsilon,k} \neq 0$ $\qquad Q = n\sum_{i=1}^{k} \hat{\rho}^2_{\varepsilon,1}$

$\hat{\rho}^2_{\varepsilon,k}$ Là hệ số tương quan mẫu bậc k của phần dư;
n là kích thước mẫu.

Nếu u_t là nhiễu trắng thì Q có phân bố tiệm cận χ^2 với bậc tự do (k-p-q).Tính toán giá trị Q, so sánh với giá trị tới hạn. Nếu giá trị của Q lớn hơn giá trị tới hạn thì bác bỏ giả thiết H_0 về phần dư không tương quan.

Bước 4 : Dự báo

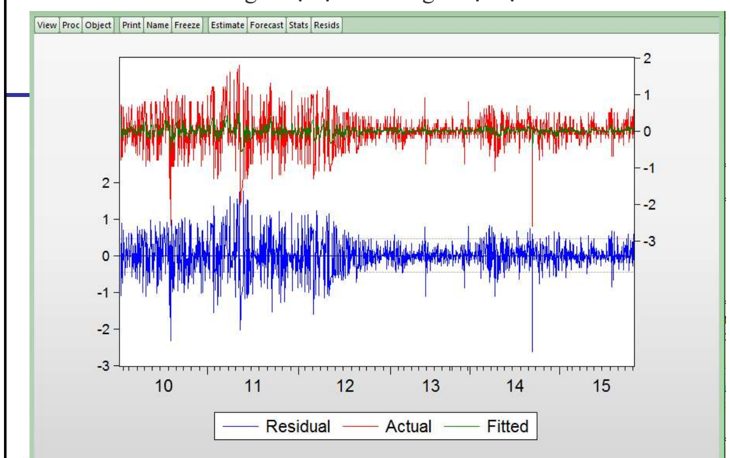
Một trong số các lý do về tính phổ biến của phương pháp này là ở tính thành công trong mô hình dự báo.

Có rất nhiều mô hình dự báo, nhưng ưu điểm của kỹ thuật này là dùng chính chuỗi thời gian này dự báo cho chính nó.

Hay nói cách khác, nó đã mô hình hóa gần như tất cả dao động của chuỗi thời gian ban đầu.

Do đó, với những chuỗi dữ liệu mang tính nhạy cảm như dự báo biến động giá dầu, giá vàng, giá chứng khoán.. Mô hình ARIMA được xem là một mô hình phù hợp

So sánh giá trị dự báo với giá trị thực tế



TÓM TẮT

Để kiểm định tính dừng của chuỗi thời gian, người ta thường dùng hai cách:

- Kiểm định tính dừng dựa vào biểu đồ tương quan
  + Dựa vào khoảng tin cậy 95%
  + Dựa vào kiểm định Box-Pierce và Ljung - Box
- Kiểm định nghiệm đơn vị đối với tính dừng

TÓM TẮT (tt)

- **Nên sử dụng kiểm định nghiệm đơn vị** để quyết định chuỗi có dừng hay không
- **Theo kiểm định nghiệm đơn vị, chuỗi dừng phải đạt được 2 yếu tố:**
  - **$\delta < 0$** tức là ρ phải nhỏ hơn 1
  - Nếu $|\tau_\alpha|$ tính toán < $|\tau|$ giá trị ADF (ADF test statistic) suy ra không có cơ sở **bác bỏ giả thiết H_0** với H_0 tồn tại nghiệm đơn vị => **chuỗi dữ liệu là chuỗi dừng**
- **Nếu như chuỗi không dừng thì sai phân cho thành chuỗi dừng. Sau khi chuỗi dừng thì tiến hành chạy mô hình ARIMA**

TÓM TẮT (tt)

AR(p) là trường hợp đặc biệt của ARIMA(p,d,q), khi d=0 và q=0.

MA(q) là trường hợp đặc biệt của ARIMA(p,d,q), khi d=0 và p=0.

ARIMA(2,1,2):

$$\Delta Y_t = \theta + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \theta_0 u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$$

Trong đó u_t là nhiễu trắng.

Như vậy nếu biết các tham số p,q,d, khi đó ta có thể mô hình hóa được chuỗi. Vấn đề đặt ra là xác định d,p,q và các tham số θ, ϕ

TÓM TẮT (tt)

Các bước thực hiện mô hình ARIMA(p,d,q):

Bước 1: nhận dạng mô hình

Bước 2: Ước lượng mô hình

Bước 3: Kiểm định chẩn đoán

Bước 4: Dự báo

TÓM TẮT (tt)

Vai trò quan trọng nhất của mô hình ARIMA chính là khả năng dự báo.

Ưu điểm của kỹ thuật này là dùng chính chuỗi thời gian này dự báo cho chính nó.

Hay nói cách khác, nó đã mô hình hóa gần như tất cả dao động của chuỗi thời gian ban đầu.

Do đó, với những chuỗi dữ liệu mang tính nhạy cảm như dự báo biến động giá dầu, giá vàng, giá chứng khoán.. Mô hình ARIMA được xem là một mô hình phù hợp

Hướng dẫn lấy data:

Bước 1. Trang lấy mã chứng khoán:
http://ivt.ssi.com.vn/CorporateInformation.aspx

B2. Trang lấy dữ liệu data chứng khoán:
http://www.bvsc.com.vn/DownloadMSData.aspx