SỬ DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO CHUỖI THỜI GIAN

NỘI DUNG

- **Giới thiệu xây dựng Mô Hình ARIMA**

  (Auto-**R**egressive **I**ntegrated **M**oving **A**verage**)**

  Tự Hồi Qui Kết Hợp Trung Bình Trượt

- **Ứng dụng dự báo giá cá sông tại Tp. HCM**

GIỚI THIỆU

Hai loại mô hình dự báo chính:

- **Mô hình nhân quả**
- **Mô hình chuỗi thời gian**

- Đối với các chuỗi thời gian

  → ARIMA thường được sử dụng để dự báo

- Theo mô hình ARIMA, giá trị dự báo sẽ phụ thuộc vào các giá trị quá khứ và tổng có trọng số các nhiễu ngẫu nhiên hiện hành và các nhiễu ngẫu nhiên có độ trễ

MÔ HÌNH ARIMA

- Tính dừng (Stationary)
- Tính mùa vụ (Seasonality)
- Nguyên lý Box-Jenkin
- Nhận dạng mô hình ARIMA
- Xác định thông số mô hình ARIMA
- Kiểm định về mô hình ARIMA

TÍNH DỪNG

Một quá trình ngẫu nhiên Y_t được xem là dừng nếu

- **Trung bình:** $E(Y_t) = const$
- **Phương sai:** $Var(Y_t) = \sigma^2 = const$
- **Đồng phương sai:** $Covar(Y_t, Y_{t-k}) = const$

Nhận biết:

- Đồ thị $Y_t = f(t)$
- Hàm tự tương quan mẫu

  (SAC – Sample Auto Correllation)

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

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

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

→ Nếu SAC = f(t) giảm nhanh và tắt dần về 0 thì chuỗi có tính dừng

- **Kiểm định Dickey-Fuller**

xác định xem chuỗi thời gian có phải là Bước Ngẫu Nhiên (Random Walk); nghĩa là

$$Y_t = 1 * Y_{t-1} + \varepsilon_t$$

→ Nếu chuỗi là Bước Ngẫu Nhiên thì không có tính dừng

BIẾN ĐỔI CHUỖI KHÔNG DỪNG THÀNH CHUỖI DỪNG:

→ Lấy sai phân bậc 1 hoặc bậc 2 thì sẽ được một chuỗi kết quả có tính dừng

- **Chuỗi gốc: Y_t**
- **Chuỗi sai phân bậc 1: $W_t = Y_t - Y_{t-1}$**
- **Chuỗi sai phân bậc 2: $V_t = W_t - W_{t-1}$**

TÍNH MÙA VỤ

Tính mùa vụ là hành vi có tính chu kỳ của chuỗi thời gian trên cơ sở năm lịch

Tính mùa vụ có thể được nhận ra dựa vào đồ thị SAC = f(t). Nếu cứ sau m thời đoạn thì SAC lại có giá trị cao thì đây là dấu hiệu của tính mùa vụ

Chuỗi thời gian có tồn tại tính mùa vụ sẽ không có tính dừng

Phương pháp đơn giản nhất để khử tính mùa vụ là lấy sai phân thứ m

$$Z_t = Y_t - Y_{t-m}$$

MÔ HÌNH ARIMA

Theo Box- Jenkin mọi quá trình ngẫu nhiên *có tính dừng* đều có thể biểu diễn bằng mô hình ARIMA

- **Mô Hình AR(p)**

Quá trình phụ thuộc vào tổng có trọng số của các giá trị quá khứ và số hạng nhiễu ngẫu nhiên

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

- **Mô Hình MA(q)**

Quá trình được mô tả bằng tổng có trọng số của các ngẫu nhiên hiện hành có độ trễ

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

- **Mô Hình ARIMA(p,d,q)**

Phương trình tổng quát của ARIMA

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

NHẬN DẠNG MÔ HÌNH

Tìm các giá trị thích hợp của p, d, q. Với

- d là bậc sai phân của chuỗi được khảo sát
- p và q sẽ phụ thuộc vào

  SPAC = f(t) và SAC = f(t)

→ Chọn mô hình AR(p) nếu SPAC có giá trị cao tại độ trễ 1, 2, ..., p và giảm nhiều sau p và dạng 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

Mô hình	SAC = f(t)	SPAC = f(t)
AR (p)	Giảm dần	Có đỉnh ở p
MA(q)	Có đỉnh ở q	Giảm dần
ARMA(p,q)	Giảm dần	Giảm dần

THÔNG SỐ CỦA ARIMA (p,d, q)

Các thông số ϕ_i và θ_j của ARIMA sẽ được xác định theo phương pháp bình phương tối thiểu (OLS-Ordinary Least Square) sao cho:

$$\sum (Y_t - \hat{Y}_t)^2 \rightarrow Min$$

Với

$$\varepsilon_t = (Y_t - \hat{Y}_t)$$

KIỂM TRA CHẨN ĐOÁN MÔ HÌNH

Kiểm định xem số hạng ε_t của mô hình có phải là một nhiễu trắng (white noise, nhiễu ngẫu nhiên thuần túy) hay không.

ε_t được tạo ra bởi quá trình nhiều trắng nếu:

$$+\varepsilon_t \sim N(0,\sigma_\varepsilon^2) \quad \begin{cases} E(\varepsilon_t) = 0 \\ Var(\varepsilon_t) = \sigma_\varepsilon^2 = const \end{cases}$$

$$+\gamma_k = Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$$

Việc kiểm định tính nhiễu trắng sẽ dựa trên đồ thị SAC của chuỗi ε_t.

DỰ BÁO

- Dự báo điểm

$$\hat{Y}_t$$

- Khoảng tin cậy

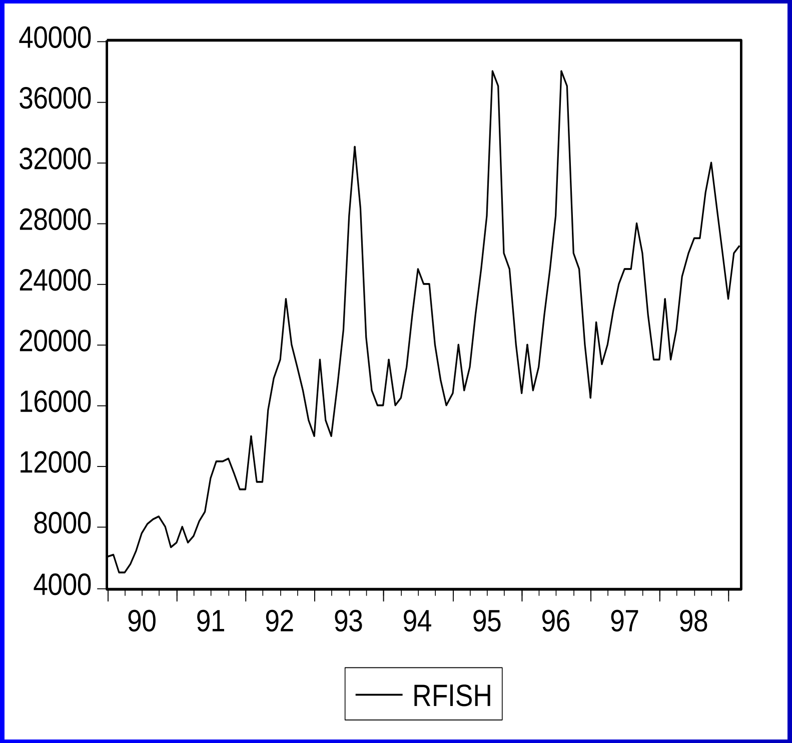
$$\hat{Y}_t - k\sigma(\varepsilon_t) < Y_t < \hat{Y}_t + k\sigma(\varepsilon_t)$$

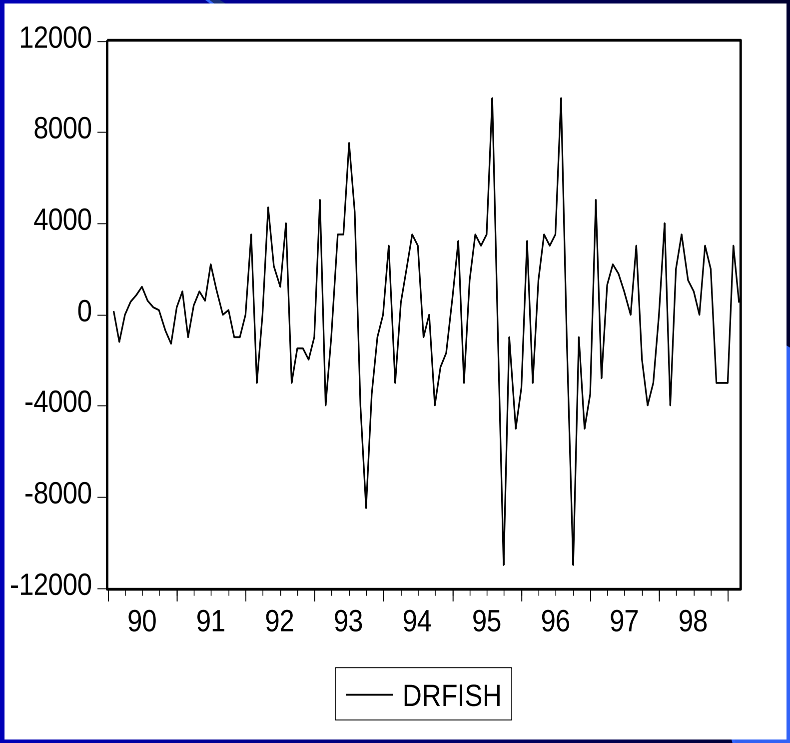
SỬ DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO GIÁ

Chuỗi giá cá sông tại Tp.HCM gồm 111 dữ liệu tháng từ 1/1990 đến 3/1999 và phần mềm EVIEWS để dự báo giá trị tháng 4/1999

Các dữ liệu quá khứ của giá cá sông được đặt tên là RFISH và chuỗi sai phân bậc 1 được đặt tên là DRFISH.

SỬ DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO GIÁ





Chuỗi RFISH và DRFISH không có tính dừng do dữ liệu có tính mùa vụ

SỬ DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO GIÁ

Sử dụng phần mềm EVIEW để khử tính mùa vụ và tiến hành thử nghiệm cho nhiều mô hình ARIMA

Mô hình tối ưu có dạng ARIMA(2,1,2) với thời đoạn khử tính mùa vụ là m = 12

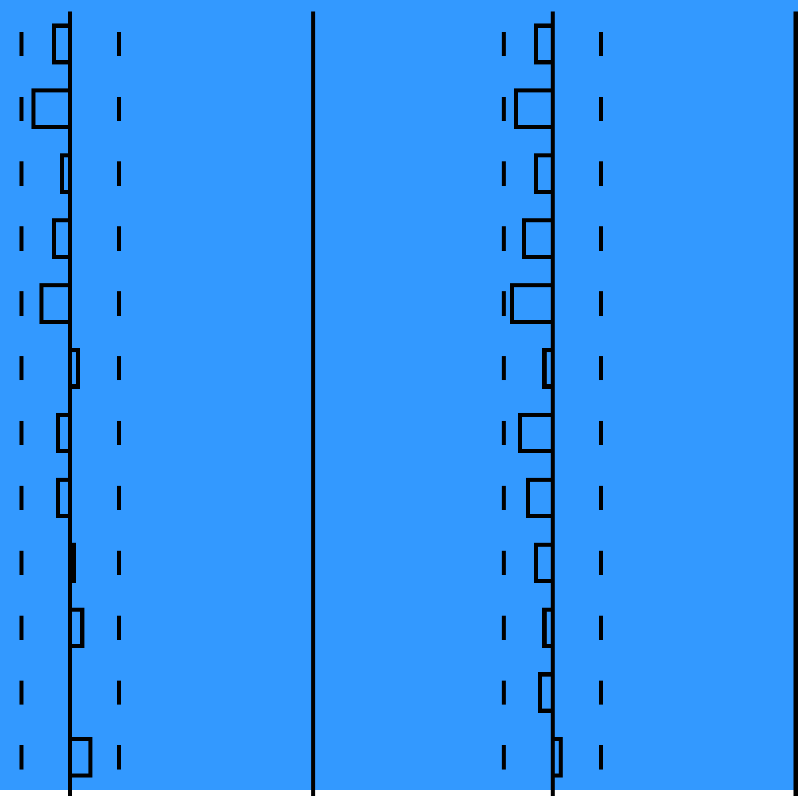
Kết quả về các thông số ϕ_i và θ_j được trình bày trong bảng sau:

Dependent Variable: D(RFISH)
Method: Least Squares
Date: 2/3/2002 Time: 18:17
Sample(adjusted): 1991:04 1999:03
Included observations: 96 after adjusting endpoints
Convergence achieved after 50 iterations
Backcast: 1990:02 1991:03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-283.3601	1010.997	-0.280278	0.7799
AR(2)	0.413278	0.135466	3.050799	0.0030
SAR(12)	0.963121	0.044544	21.62164	0.0000
MA(2)	-0.846851	0.118603	-7.140218	0.0000
SMA(12)	-0.781433	0.078476	-9.957634	0.0000
R-squared	0.614807	Mean dependent var		203.1250
Adjusted R-squared	0.597875	S.D. dependent var		3545.923
S.E. of regression	2248.588	Akaike info criterion		18.32467
Sum squared resid	4.60E+08	Schwarz criterion		18.45823
Log likelihood	-874.5842	F-statistic		36.31124
Durbin-Watson stat	1.718345	Prob(F-statistic)		0.000000

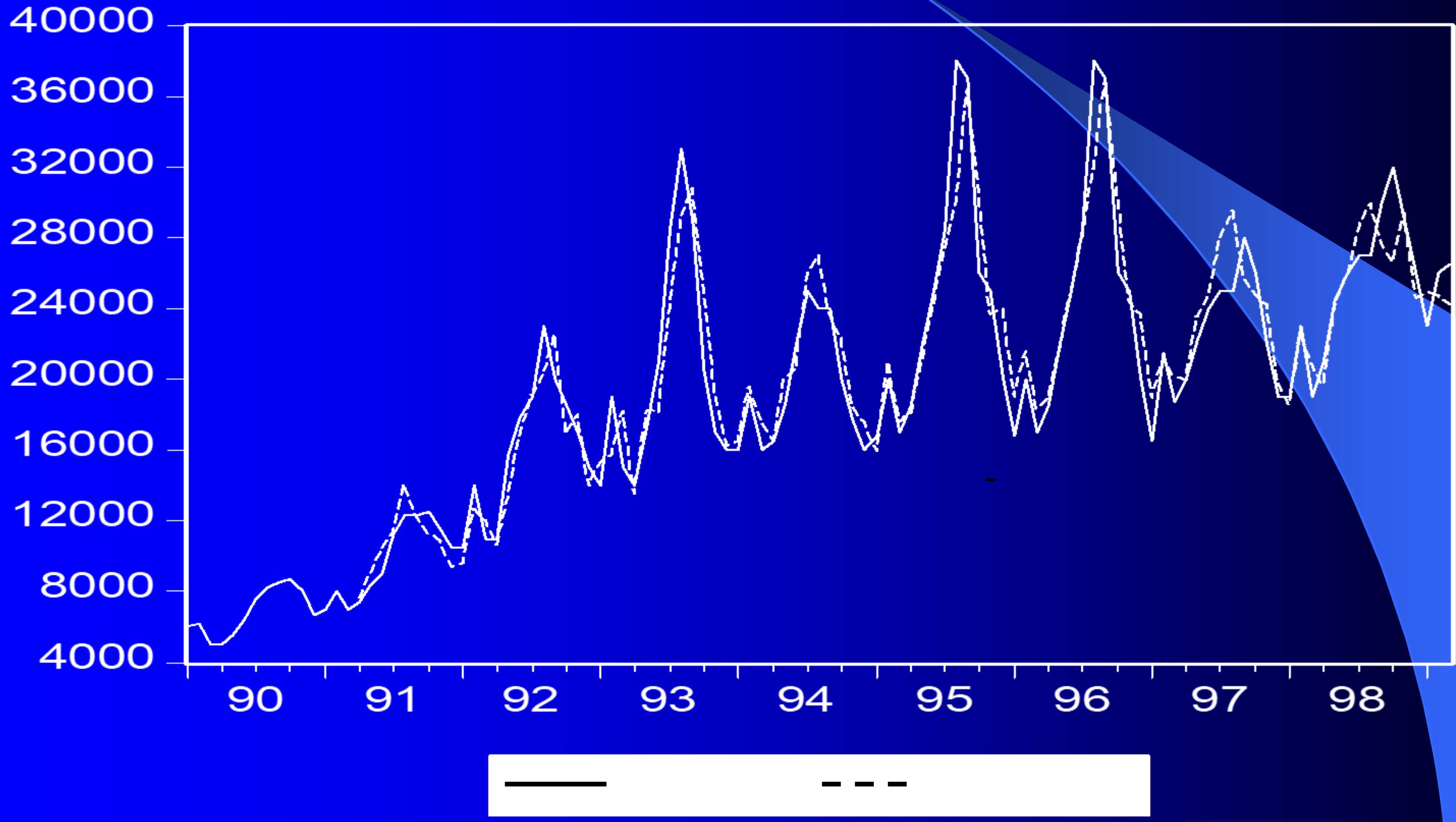
THẨM ĐỊNH TÍNH NHIỄU TRẮNG CỦA ε_t

Đồ thị SAC của chuỗi e_t. cho thấy e_t có tính nhiễu trắng và được trình bày như sau:



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.078	-0.078	0.6208	
		2	-0.154	-0.161	3.1000	
		3	-0.033	-0.061	3.2128	
		4	-0.078	-0.116	3.8608	
		5	-0.130	-0.174	5.6875	0.017
		6	0.038	-0.034	5.8479	0.054
		7	-0.057	-0.132	6.2061	0.102
		8	-0.048	-0.111	6.4620	0.167
		9	0.017	-0.080	6.4944	0.261
		10	0.056	-0.026	6.8463	0.335
		11	-0.007	-0.054	6.8519	0.444
		12	0.084	0.037	7.6631	0.467

ĐỒ THỊ CỦA RFISH VÀ RFISHF



KẾT QUẢ

Dự báo điểm là $\hat{Y}_t$ = 26267 Đ

- Khoảng tin cậy 95% là [21742 Đ, 30792 Đ]
- Giá trị thực tháng 4/1999 là Y_t = 26000 Đ
- Giá trị này nằm trong khoảng tin cậy 95% và xấp xỉ với giá trị dự báo điểm
- Sai số dự báo là ($\hat{Y}_t$ -Y_t)/ Y_t *100 = 1,03%

KẾT LUẬN

- Đồ thị RFISHF bám rất sát đồ thị RFISH
- Giá trị dự báo xấp xỉ với giá trị trên thực tế (sai số dự báo nhỏ) và khoảng tin cậy 95% cũng chứa giá trị thực → độ tin cậy của mô hình dự báo
- Đã áp dụng mô hình ARIMA để dự báo cho hơn 20 loại mặt hàng tại Tp.HCM theo qui trình tương tự và cũng đạt được các kết quả dự báo với độ tin cậy cao

→ TÓM LẠI, MÔ HÌNH ARIMA LÀ MỘT MÔ HÌNH ĐÁNG TIN CẬY ĐỐI VỚI DỰ BÁO NGẮN HẠN

TÀI LIỆU THAM KHẢO

Bowerman B.L., and O'Connell R.T., 1993. *Forecasting and Time Series*. 3rd ed., Wadsworth, Inc.

Cao Hào Thi và Các Cộng Sự 1998. *Bản Dịch Kinh Tế Lượng Cơ Sở (Basic Econometrics của Gujarati D.N.)*. Chương Trình Fulbright về Giảng Dạy Kinh Tế tại Việt Nam.

EVIEWS, 2000. Quantitative Micro Software.

TÀI LIỆU THAM KHẢO

Pindyck R.S., and Rubinfeld D.L., 1991. *Econometric Models and Economic Forecast*. 3rd ed., McGraw-Hill.

Ramanathan R., 2001. *Introductory Econometrics with Applications*. 5th ed., Harcourt College Publishers