

ỨNG DỤNG MÔ HÌNH ARIMA ĐỂ DỰ BÁO VNINDEX APPLICATION OF ARIMA MODEL TO FORECAST VNINDEX

**SVTH: Bùi Quang Trung, Nguyễn Quang Minh Nhi,
Lê Văn Hiếu, Nguyễn Hồ Diệu Uyên**

Lớp 33K15, Khoa Tài Chính – Ngân hàng, Trường Đại học Kinh tế

GVHD: TS. Võ Thị Thúy Anh

Khoa Tài Chính – Ngân hàng, Trường Đại học Kinh tế

TÓM TẮT

Thị trường chứng khoán trên thế giới nói chung và ở Việt Nam nói riêng luôn là nơi hấp dẫn các tổ chức và cá nhân đầu tư bởi mức sinh lợi cao của nó. Tuy nhiên, đây cũng là một hoạt động tiềm ẩn rất nhiều rủi ro. Vì thế, việc đưa ra dự báo xu hướng biến động của chỉ số giá chứng khoán để có một sách lược phù hợp cho hoạt động đầu tư của cá nhân, tổ chức thu hút rất nhiều sự quan tâm của các nhà kinh tế lượng tài chính trong và ngoài nước. Đề tài này cung cấp cách thức xây dựng mô hình ARIMA trong dự báo chỉ số VNIndex trên thị trường chứng khoán Việt Nam.

ABSTRACT

Stock markets around the world in general and in Vietnam in particular are always attractive to investment institutions and individual investors because of its high level of profitability. However, it is also an operation with a lot of potential risks. Thus, forecasting the trends of the stock index to adapt a consistent investment strategy for individuals and organizations attracts the attention of many financial specialists, both domestic and abroad. This research offers a method to build the ARIMA model in forecasting the VN-Index on this local stock market.

1. Đặt vấn đề

Ra đời vào đầu năm 2000, thị trường chứng khoán Việt Nam đã trở thành một kênh đầu tư hết sức hấp dẫn đối với các nhà đầu tư, từ các tổ chức đầu tư chuyên nghiệp cho đến các nhà đầu tư cá nhân nghiệp dư nhỏ lẻ. Tuy nhiên, bên cạnh mức sinh lợi cao, đây cũng là hoạt động luôn tồn tại nhiều rủi ro tiềm ẩn bởi nhà đầu tư không phải lúc nào cũng dự đoán được chính xác xu hướng của giá cổ phiếu trong tương lai. Do đó, việc dự báo chính xác sự biến động giá của cổ phiếu để có một sách lược nhằm phục vụ cho công việc kinh doanh của các cá nhân, tổ chức hay hoạch định chiến lược của một quốc gia đã thu hút rất nhiều sự quan tâm của các nhà kinh tế lượng tài chính trong và ngoài nước.

Tại thị trường Việt Nam, sự biến động của chỉ số VnIndex phản ánh rủi ro hệ thống, vì vậy, việc dự báo được sự tăng giảm của Vn-Index cũng đồng thời giúp các nhà đầu tư nhận biết chiều hướng biến động giá của các cổ phiếu trên thị trường này.

Trong khuôn khổ đề tài, chúng tôi đề xuất sử dụng mô hình ARIMA và phương pháp Box-jenkins để dự báo chỉ số VnIndex trong ngắn hạn căn cứ vào chuỗi dữ liệu quá khứ. George Box và Gwilym Jenkins (1976) đã nghiên cứu mô hình ARIMA (**Autoregressive Integrated Moving Average** - *Tự hồi qui tích hợp Trung bình trượt*), và tên của họ thường được dùng để gọi tên các quá trình ARIMA tổng quát, áp dụng vào việc phân tích và dự báo các chuỗi thời gian. Phương pháp Box-Jenkins với bốn bước: nhận dạng mô hình thử nghiệm; ước lượng; kiểm định bằng chẩn đoán; và dự báo.

2. Xây dựng mô hình ARIMA cho VnIndex

2.1. Giới thiệu về số liệu

+ Nguồn cập nhật số liệu là trang web cophieu68.com. Đây là trang web chuyên cung cấp số liệu về thị trường chứng khoán Việt Nam.

+ Số liệu VnIndex được lấy từ ngày 2/1/2009 tới ngày 30/3/2010. Sở dĩ nhóm thực hiện quyết định chọn chuỗi thời gian này vì VnIndex trong thời gian này phản ánh tương đối tác động của nền kinh tế vĩ mô lên giá chứng khoán.

2.2. Cơ sở lý luận

Mô hình sử dụng dữ liệu chuỗi thời gian, xem giá trị trong quá khứ của một biến số cụ thể là một chỉ tiêu tốt phản ánh giá trị trong tương lai của nó, cụ thể, cho Y_t là giá trị của biến số tại thời điểm t với $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_0, t)$.

Mục đích của phân tích là để thấy rõ một số mối quan hệ giữa các giá trị Y_t được quan sát đến nay để cho phép chúng ta dự báo giá trị Y_t trong tương lai. Phương pháp này đặc biệt hữu ích cho việc dự báo trong ngắn hạn.

Mô hình tự hồi quy p - AR(p): trong mô hình tự hồi quy quá trình phụ thuộc vào tổng trọng số của các giá trị quá khứ và số hạng nhiễu ngẫu nhiên

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

Mô hình trung bình trượt q - MA(q): trong mô hình trung bình trượt, quá trình được mô tả hoàn toàn bằng tổng trọng số của các nhiễu ngẫu nhiên hiện hành có độ trễ:

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

Mô Hình Hồi Quy Kết Hợp Trung Bình Trượt - ARMA(p, q):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

2.2.1. Xem xét tính dừng của chuỗi quan sát

Điều trước tiên cần phải lưu ý là hầu hết các chuỗi thời gian đều không dừng, và các thành phần AR và MA của mô hình ARIMA chỉ liên quan đến các chuỗi thời gian dừng. Quy trình ngẫu nhiên của Y_t được xem là dừng nếu trung bình và phương sai của quá trình không thay đổi theo thời gian và giá đồng phương sai giữa hai thời đoạn chỉ phụ thuộc vào khoảng cách độ trễ về thời gian giữa các thời đoạn này chứ không phụ thuộc vào thời điểm thực tế mà đồng phương sai được tính. Do đó, để nhận diện mô hình ARIMA, chúng ta phải thực hiện hai bước sau:

Có ba cách để nhận biết tính dừng của một chuỗi thời gian là dựa vào trên đồ thị của chuỗi thời gian, đồ thị của hàm tự tương quan mẫu hay kiểm định Dickey – Fuller.

2.2.2. Nhận dạng mô hình

Nhận dạng mô hình ARMA(p, d, q) là tìm các giá trị thích hợp của p , d , q . Với d là bậc sai phân của chuỗi thời gian được khảo sát, p là bậc tự hồi quy và q là bậc trung bình trượt.

Việc xác định p và q sẽ phụ thuộc vào các đồ thị SPACF = $f(t)$ và SACF = $f(t)$. Với SACF là hàm tự tương quan mẫu và SPACF là hàm tự tương quan mẫu riêng phần (Sample Partial Autocorrelation):

+ Chọn giá trị của p nếu đồ thị SPACF có giá trị cao tại độ trễ 1, 2, ..., p và giảm nhiều sau p và dạng hàm SAC giảm dần.

+ Chọn giá trị của q nếu đồ thị SACF có giá trị cao tại độ trễ 1, 2, ..., q và giảm nhiều sau q và dạng hàm SPAC giảm dần.

2.2.3. Ước lượng các tham số của mô hình

Các hệ số ϕ và ω của mô hình ARIMA được xác định bằng phương pháp ước lượng thích hợp cực đại. Sau đó chúng ta kiểm định ϕ và ω bằng thống kê t . Ước lượng sai số bình phương trung bình của phần dư: S^2

2.2.4. Kiểm định mô hình

Sau khi ước lượng các tham số của một mô hình ARIMA được nhận dạng thử, chúng ta cần phải kiểm định để kiểm nghiệm rằng mô hình là thích hợp. Các cách thức để thực hiện điều này:

Kiểm tra phần dư e_t có phải là nhiễu trắng không. Nếu e_t là nhiễu trắng thì chấp nhận mô hình, trong trường hợp ngược lại chúng ta phải tiến hành lại từ đầu. Các kiểm định có thể sử dụng là kiểm định BP (Box-Priere) hoặc kiểm định Ljung-box với trị thống kê Q , hoặc kiểm định LM.

Nếu tồn tại nhiều hơn một mô hình đúng, mô hình có AIC (Akaike Information Criterion) nhỏ nhất sẽ được lựa chọn.

2.2.5. Dự báo bằng mô hình ARIMA

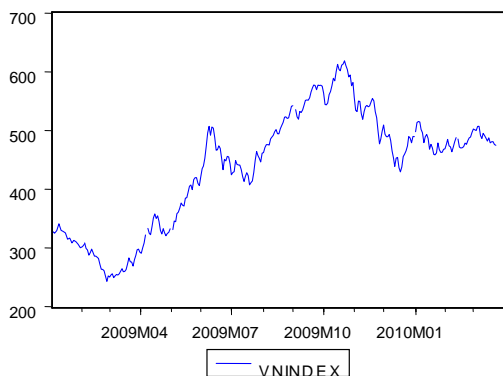
Một trong số các lý do về tính phổ biến của phương pháp lập mô hình ARIMA là thành công của nó trong dự báo. Trong một số trường hợp dự báo thu được từ phương pháp này có tính tin cậy cao hơn so với các dự báo thu được từ các phương pháp lập mô hình kinh tế lượng truyền thống khác, đặc biệt là đối với dự báo ngắn hạn.

Dựa vào mô hình ARIMA ước lượng được, tiến hành xác định giá trị dự báo và khoảng tin cậy cho dự báo với độ tin cậy 95% và $k=1.96$ như sau:

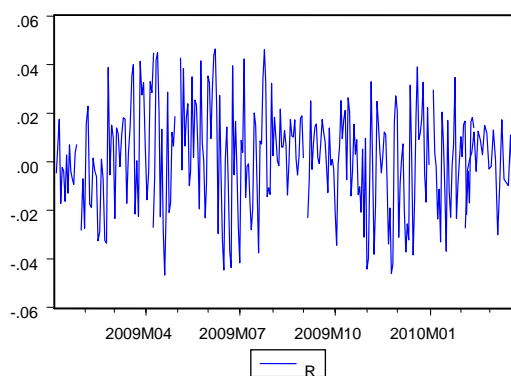
+ Dự báo điểm \hat{Y}_t

+ Khoảng tin cậy $\hat{Y}_t - k\sigma(\varepsilon_t) < \hat{Y}_t < \hat{Y}_t + k\sigma(\varepsilon_t)$

2.2.6. Mô hình ARIMA cho VnIndex



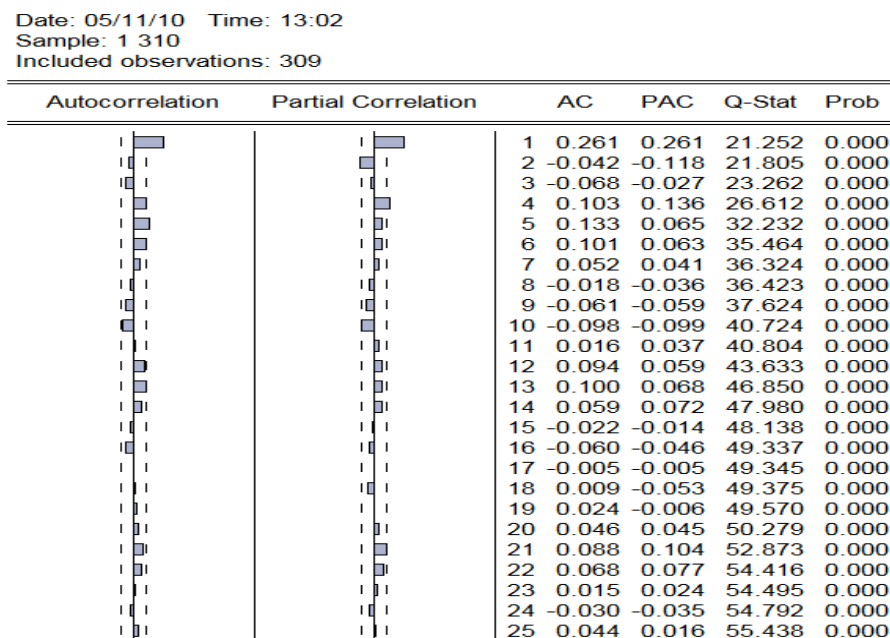
Hình 1 – VnIndex



Hình 2 – Tỷ suất sinh lời R

Để xây dựng mô hình ARIMA nhóm chúng tôi đã sử dụng chuỗi dữ liệu gồm 310 quan sát từ ngày 2/1/2009 tới 30/3/2010. Dữ liệu quá khứ được đặt tên là Vnindex sau đó được lấy logarit tự nhiên trước khi lấy sai phân bậc nhất để có được tỷ suất lợi tức của VnIndex, ký hiệu là r . Đồ thị $Vnindex=f(t)$ và $r=f(t)$ được trình bày ở hình 1 và hình 2.

Từ hình 2 và sử dụng kiểm định Dickey – Fuller cho P-value < 0.05 cho thấy chuỗi R là chuỗi dừng.



Hình 3 – Đồ thị của hàm tự tương quan và tự tương quan riêng phần của chuỗi R

Để định dạng cho mô hình chúng ta sử dụng đồ thị tự tương quan và tự tương quan riêng phần của chuỗi R . Theo đồ thị ở hình 3, tại $k=1$ SAC và PAC đạt cực đại 0.261 và sau đó giảm mạnh xuống. Do đó p và q có thể nhận các giá trị là 1. Các mô hình ARIMA có thể có là ARIMA (0,1,1), ARIMA(1,1,1), ARIMA(1,1,0)

Trước tiên ta dùng kiểm định LB để chọn ra các mô hình có p-value bé hơn 0.05. Dùng đồ thị SAC để kiểm tra chuỗi e_t thấy rằng cả 4 chuỗi e_t của 4 mô hình đề là nhiều trắng. Bây giờ cần lựa chọn mô hình tốt nhất để sử dụng cho công tác dự báo, chúng ta sử dụng tiêu chuẩn kiểm định Chi bình phương $\chi^2(4)$ và tiêu chuẩn AIC. Theo kết quả từ eview cho bởi bảng 1 ta thấy mô hình ARIMA(0,1,1) là mô hình phù hợp với R nhất.

Mô hình	số quan sát	$\chi^2(4)$	AIC
Arima(0,1,1)	476	0.129571	-4.976207
Arima(1,1,1)	476	0.104665	-4.966832
Arima(1,1,0)	476	0.061240	-4.964157

Bảng 1 – Kết quả các thông số kiểm định

Dùng phương pháp bình phương bé nhất để ước lượng các tham số của mô hình. Thực hiện ước lượng bằng Evview ta được mô hình ARIMA(0,1,1) có các hệ số như sau:

$$r_T = 0.001504 + 0.284364 \varepsilon_{T-1}$$

2.3. Dự báo VNindex bằng mô hình xây dựng được

Sử dụng mô hình vừa xây dựng để dự báo điểm và khoảng tin cậy cho r tại thời điểm ngày 31/3/2010 bằng phần mềm Eview với độ tin cậy 95%.

Kết quả thu được $r = -0.1495\%$ và khoảng tin cậy là $[-0.444676\%; 0.145676\%]$. Trong bước đầu xử lý số liệu ta đã chuyển VnIndex thành tỷ suất sinh lời r thông qua việc lấy logarit tự nhiên trước khi lấy sai phân bậc nhất. Do đó từ kết quả này để quy ngược về Vnindex chúng ta sử dụng công thức: $VnIndex_t = e^r \cdot VnIndex_{t-1}$. Từ đó ta có dự báo điểm và khoảng tin cậy cho Vnindex ngày 31/3/2010 với mức tin cậy 95% là: 499.952; [498.46; 501.43]

Giá trị Vnindex thực ngày 31/3/2010 là 499,2. Giá trị này nằm trong khoảng tin cậy 95% và xấp xỉ giá trị dự báo điểm là 499.952. Sai số dự báo là: 0.1506%.

3. Kết luận

Kết quả dự báo cho thấy giá trị dự báo xấp xỉ với giá trị thực tế và khoảng tin cậy 95% cũng chứa giá trị thực tế. Điều này chứng tỏ độ tin cậy của mô hình dự báo là khá cao. Trong một vài phiên giao dịch do tác động của các yếu tố ngoại lai lớn như tâm lý nhà đầu tư, tác động của các thị trường chứng khoán khác, thông tin về sự thay đổi chính sách... sẽ làm cho sai số dự báo tăng cao hơn. Do đó kết quả của mô hình vẫn chỉ mang tính chất tham khảo nhiều hơn. Tuy nhiên có thể nói mô hình ARIMA là một mô hình tốt để dự báo trong ngắn hạn.

TÀI LIỆU THAM KHẢO

- [1] John E.Hanke & Dean W.Wichern, (2005), Business Forecasting, 8th Edition, Chapter 9.
- [2] Cao Hào Thi và các cộng sự (1998), Bản Dịch Kinh Tế Lượng Cơ Sở (Basic Econometrics của Gujarati D.N) Chương trình giảng dạy kinh tế FulBright tại Việt Nam.
- [3] Phùng Thanh Bình, Hướng dẫn sử dụng Eview trong phân tích dữ liệu và hồi quy.
- [4] Nguyễn Thống (2000), *Kinh Tế Lượng Ứng Dụng*, Nhà xuất bản Đại Học Quốc Gia Tp.Hồ Chí Minh, tr.238-278
- [5] Nguyễn Quang Dong (2006), *Kinh Tế Lượng (chương trình nâng cao)*, Nhà xuất bản Khoa học và Kỹ thuật, Hà Nội, chương 3-4-5.