

XÂY DỰNG MÔ HÌNH ARIMA CHO DỰ BÁO KHÁCH DU LỊCH QUỐC TẾ ĐẾN VIỆT NAM

Building ARIMA Model for Forecasting Vietnam's Outbound Tourists

¹Đỗ Quang Giám, ²Vũ Thị Hân, ²Lý Thị Lan Phương, ²Nguyễn Thu Thủy

¹Khoa Kế toán và QTKD, Trường Đại học Nông nghiệp Hà Nội

²Lớp QTKD Tiên tiến, Trường ĐH Nông nghiệp Hà Nội

Địa chỉ email tác giả liên hệ: giamdq@hua.edu.vn

Ngày gửi bài: 17.12.2011

Ngày chấp nhận: 14.03.2012

TÓM TẮT

Du lịch Việt Nam đang trở thành ngành kinh tế mũi nhọn của đất nước, chiếm tới 4% GDP và cũng là ngành đóng góp khoản thu lớn vào ngân sách nhà nước. Việt Nam được đánh giá có tiềm năng du lịch dồi dào, tuy nhiên chúng ta chưa khai thác hết thế mạnh đó. Vì vậy, dự báo lượng khách du lịch quốc tế đến nước ta có ý nghĩa đối với các nhà quản lý, các nhà đầu tư để có kế hoạch phát triển ngành công nghiệp không khói này. Bài viết này sử dụng phương pháp Box-Jenkins để xây dựng mô hình ARIMA cho dự báo lượng khách quốc tế đến Việt Nam dựa trên số liệu công bố hàng tháng của Tổng cục Du lịch Việt Nam. Kết quả cho thấy trong số các mô hình ước lượng thử nghiệm thì ARIMA(12,1,12) là phù hợp nhất. Bài cũng đưa ra dự báo ngắn hạn về lượng khách quốc tế đến Việt Nam những tháng đầu năm 2011 với mức độ sai số so với thực tế từ 14 đến 27%. Tuy vậy, mô hình ARIMA có thể dùng để dự báo, song chưa phải là tối ưu, bởi vì sự phụ thuộc trong mô hình được giả định là tuyến tính.

Từ khóa: ARIMA, dự báo, khách du lịch quốc tế.

SUMMARY

Vietnam Tourism has become a focal industry of the country, accounted for 4% of the GDP. The sector has also contributed a considerable amount to the state budget. Vietnam has been rated as a fully potential destination for international tourists, however we have not taken advantage of such strengths. Therefore, forecasting tourist arrivals in coming years is very useful for managers and investors, who can base on this to perform their businesses better. The paper applied the Box-Jenkins method to build an appropriate ARIMA model for forecasting outbound tourist arrivals to Vietnam, using monthly data from Vietnam National Administration of Tourism. The findings show that among the trial estimated models, ARIMA(12,1,12) is the most appropriate model. Based on this model, short-term forecasts on outbound tourist arrivals to Vietnam in early months, 2011 has been done with errors varying from 14 to 27%. However, ARIMA model could be used for forecast, but it could not be optimal selection since autocorrelation and moving average in the model are assumed to be linear.

Keywords: ARIMA, Forecast, Vietnam's outbound tourists.

1. ĐẶT VẤN ĐỀ

Việt Nam là quốc gia có đường biên giới tiếp giáp cả đất liền và biển, du khách quốc tế có thể tới nước ta theo nhiều hướng, với nhiều phương tiện khác nhau. Lịch sử kiến tạo địa chất qua nhiều niên đại đã tạo cho bề mặt lãnh thổ nước ta nhiều đường nét hình khối độc đáo và đẹp. Cùng với sự phát triển

kinh tế-xã hội, du lịch ngày càng trở thành nhu cầu không thể thiếu của người dân nhiều nước. Với tiềm năng du lịch đa dạng và phong phú, nước ta đã thu hút nhiều du khách quốc tế đến thăm quan. Nhận thức được điều này, Đảng và Chính phủ đã rất quan tâm đến phát triển ngành du lịch nước nhà. Nghị quyết Trung ương IX xem hoạt động du lịch là một trong những ngành kinh

tế mũi nhọn. Quyết định Số 97/2002/QĐ-TTg của Thủ tướng Chính phủ đã phê duyệt Chiến lược phát triển du lịch Việt Nam 2001-2010, với mục tiêu phát triển du lịch thành một ngành kinh tế mũi nhọn, từng bước đưa nước ta trở thành một trung tâm du lịch tâm cổ của khu vực (Chính phủ, 2002).

Thực tế cho thấy, từ năm 1991 đến năm 2010 lượng khách du lịch quốc tế đến nước ta tăng từ 300.000 lên 5.049.855 lượt khách. Việt Nam là một trong 5 nước có mức tăng trưởng du lịch cao nhất thế giới, với mức tăng trưởng lượng khách hàng năm khoảng 36%, thu nhập từ du lịch năm 2008 đạt gần 4 tỉ USD/năm, đóng góp khoảng 4% vào GDP. Tuy nhiên, ngành du lịch nước ta vẫn được xem là phát triển chưa tương xứng tiềm năng, điều này lý giải vì sao xếp hạng cạnh tranh của du lịch của Việt Nam khá thấp (Lê Minh, 2010). Năm 2008, Việt Nam xếp thứ 97 trên tổng số 113 nước, trong khi Singapore xếp thứ 7, Malaysia xếp 32, Thái Lan thứ 42 (Tổng cục Du lịch, 2008). Thực trạng đầu tư cho du lịch nước ta chưa đồng bộ, chạy theo số lượng mà ít tính đến sự chuyên nghiệp. Việc kinh doanh du lịch chấp nhận có nhiều rủi ro không chỉ đến từ yếu tố khách quan như thời tiết, khí hậu, dịch bệnh, v.v... mà còn từ yếu tố quản lý, do không có dự báo tốt để có những quy hoạch, kế hoạch phát triển ngành hợp lý. Để khai thác có hiệu quả các tiềm năng du lịch, tạo dấu ấn tốt trong lòng du khách, khắc phục những rủi ro trong kinh doanh dịch vụ du lịch, mục đích của bài viết này là xây dựng một mô hình ARIMA phù hợp để làm căn cứ dự báo lượng khách du lịch quốc tế đến Việt Nam thời gian tới.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thu thập và xử lý dữ liệu

Số liệu phục vụ cho đề tài được tổng hợp từ các báo cáo thường niên, báo cáo hàng

tháng của Tổng cục Du lịch từ tháng 1/1995 đến tháng 12/2010 tại website của Tổng cục Du lịch

<http://www.vietnamtourism.gov.vn/index.php?cat=2020> và từ số liệu thống kê của Tổng cục Thống kê. Số liệu được tổng hợp theo tháng và được xử lý trên phần mềm Excel và Eviews 5.1.

2.2. Phương pháp nghiên cứu

Hai tác giả George Box & Gwilym Jenkins (1976) đã nghiên cứu mô hình tự hồi qui tích hợp trung bình trượt (Autoregressive Integrated Moving Average), viết tắt là ARIMA. Tên của họ (Box-Jenkins) được dùng để gọi cho các quá trình ARIMA tổng quát áp dụng vào phân tích và dự báo các chuỗi thời gian. Mô hình tự tương quan bậc p (viết tắt là AR(p)) là quá trình phụ thuộc tuyến tính của các giá trị trễ và sai số ngẫu nhiên được diễn giải như sau:

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

Mô hình trung bình trượt bậc q , viết tắt là MA(q), là quá trình được mô tả hoàn toàn bằng phương trình tuyến tính có trọng số của các sai số ngẫu nhiên hiện hành và các giá trị trễ của nó. Mô hình được viết như sau:

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

Mô hình tự tương quan tích hợp với trung bình trượt có dạng ARIMA (p, d, q), được xây dựng dựa trên 2 quá trình (1) và (2) được tích hợp. Phương trình tổng quát là:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Phương pháp Box-Jenkins gồm bốn bước lặp là (i) Nhận dạng mô hình thử nghiệm, (ii) Ước lượng, (iii) Kiểm định bằng chẩn đoán và (iv) Dự báo, được trình bày dưới đây:

Bước 1: Nhận dạng mô hình

Nhận dạng mô hình ARIMA (p, d, q) là việc tìm các giá trị thích hợp của p , d và q . Với d là bậc sai phân của chuỗi thời gian được khảo sát, p là bậc tự hồi qui và q là bậc trung bình trượt. Việc xác định p và q sẽ phụ thuộc vào các đồ thị SPAC = $f(t)$ và SAC = $f(t)$, với SAC là hàm tự tương quan của mẫu (Sample Autocorrelation) và SPAC là hàm tự tương quan từng phần của mẫu (Sample Partial Autocorrelation).

Việc lựa chọn mô hình AR(p) phụ thuộc vào đồ thị SPAC nếu nó có giá trị cao tại các độ trễ $1, 2, \dots, p$ và giảm đột ngột sau đó, đồng thời dạng hàm SAC tắt lịm dần. Tương tự, việc chọn mô hình MA(q) dựa vào đồ thị SAC nếu nó có giá trị cao tại độ trễ $1, 2, \dots, q$

và giảm mạnh sau q , đồng thời dạng hàm SPAC tắt lịm dần.

Bước 2: Ước lượng các thông số của mô hình ARIMA (p, d, q)

Các tham số của mô hình ARIMA sẽ được ước lượng theo phương pháp bình phương nhỏ nhất.

Bước 3: Kiểm tra chẩn đoán mô hình

Sau khi xác định các tham số của quá trình ARIMA, điều cần phải làm là tiến hành kiểm định xem số hạng sai số e_t của mô hình có phải là một nhiễu trắng (white noise) hay không. Đây là yêu cầu của một mô hình tốt (Wang & Lim, 2005).

Bước 4: Dự báo: Dựa trên phương trình của mô hình ARIMA, tiến hành xác định giá trị dự báo điểm và khoảng tin cậy của dự báo.

Bảng 1. Lượng khách quốc tế đến Việt Nam thời kỳ 1995 - 2010

Năm	Số lượt khách	Tốc độ phát triển liên hoàn (%)
1995	1.351.296	-
1996	1.607.155	118,93
1997	1.715.637	106,75
1998	1.520.128	88,60
1999	1.748.754	115,04
2000	2.168.335	123,99
2001	2.330.791	107,49
2002	2.628.237	112,76
2003	3.328.884	126,66
2004	2.927.873	87,95
2005	3.534.650	120,72
2006	3.632.001	102,75
2007	4.224.401	116,31
2008	4.197.317	99,36
2009	3.764.756	89,69
2010	5.049.855	134,80
Trung bình	2.858.129	-

Nguồn: Tổng hợp từ website của Tổng cục Du lịch www.vietnamtourism.gov.vn/index.php?cat=202035

3. KẾT QUẢ VÀ THẢO LUẬN

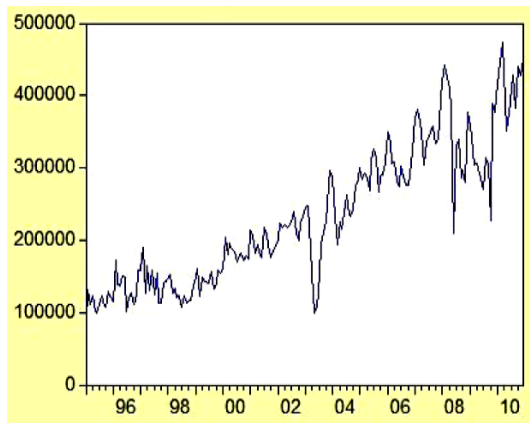
3.1. Kiểm định tính dừng

Trong toán học, *tính dừng* (Stationarity) được sử dụng như một công cụ trong phân tích chuỗi số liệu theo thời gian. Để hình thành một mô hình đầy đủ ý nghĩa thống kê thì chuỗi số liệu theo thời gian trước tiên cần kiểm tra tính *dừng* của chuỗi. Một quá trình có tính dừng là một quá trình ngẫu nhiên, được thể hiện bởi trung bình mẫu và phương sai của sai số không đổi theo thời gian. Trên thực tế, hầu hết các chuỗi số liệu kinh tế (chuỗi gốc) đều không dừng. Điều này được hiểu là các chuỗi thời gian đó có trung bình mẫu và phương sai thay đổi theo thời gian. Nhưng khi ta lấy sai phân thì các chuỗi thời gian thường trở thành chuỗi dừng (Do Q. Giam & cs., 2009).

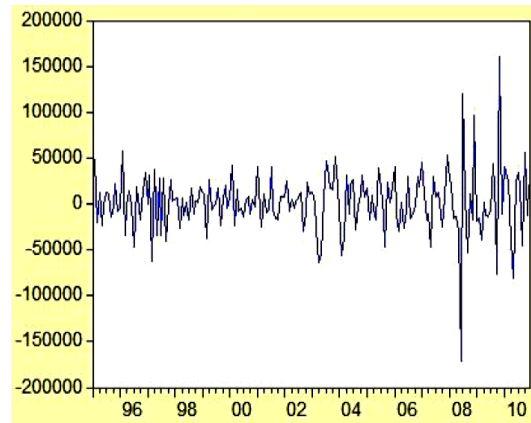
Chuỗi số liệu sử dụng trong mô hình ARIMA được giả định là chuỗi dừng. Vì vậy, để dự đoán số lượng khách quốc tế đến Việt Nam bằng mô hình ARIMA, chúng ta cần phải xem xét liệu các chuỗi đó có phải là chuỗi dừng hay không. Để khẳng định điều này, trước tiên có thể dựa vào quan sát trực

tiếp đồ thị của chuỗi số liệu, sau đó phải kiểm định. Hai phương pháp kiểm định rất phổ biến là Augmented Dickey-Fuller (ADF) và Perron-Phillips (PP) mà kinh tế lượng gọi là kiểm định gốc đơn vị (unit root test) cho các chuỗi số liệu gốc và các chuỗi sai phân.

Hình 1 cho thấy diễn biến số lượng du khách quốc tế đến Việt Nam (viết tắt là TOURISTARRIVAL) theo tháng (từ tháng 1/1995 đến 12/2010) không ổn định và có xu hướng tăng. Cụ thể, trung bình của nó có xu hướng tăng hoặc giảm theo từng thời kỳ. Như vậy, có thể suy đoán rằng chuỗi số lượng khách du lịch là không dừng. Tuy vậy, khi lấy sai phân bậc nhất của chuỗi này thì ta được chuỗi mới, biến động lượng du khách quốc tế qua các tháng (viết tắt là DTOURISTARRIVAL), chuỗi này không rõ xu hướng và xoay quanh một giá trị trung bình nào đó (Hình 2). Đây được xem như là biểu hiện của một chuỗi dừng. Để khẳng định những suy đoán trên, 2 phương pháp kiểm định là ADF và PP được sử dụng để xem liệu kết quả có nhất quán hay không. Kết quả kiểm định cho chuỗi số liệu gốc và chuỗi sai phân bậc 1 được trình bày ở Bảng 2.



Hình 1. Diễn biến lượng khách quốc tế đến Việt Nam, 1995-2010 (lượt)

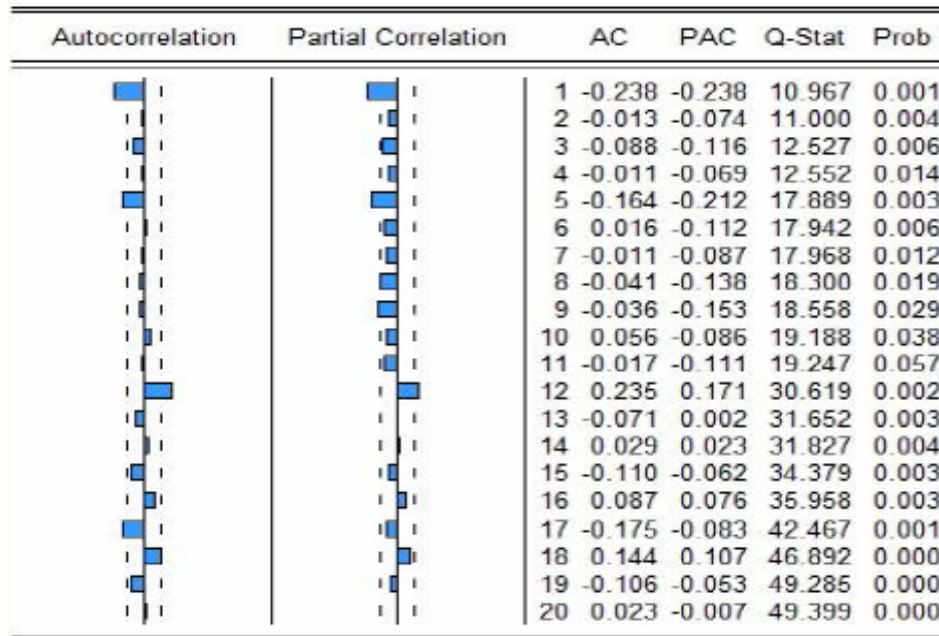


Hình 2. Biến động lượng khách quốc tế đến Việt Nam, 1995-2010 (lượt)

Bảng 2. Kết quả kiểm định ADF và PP đối với chuỗi gốc và chuỗi sai phân bậc 1

	Kí hiệu chuỗi	Giá trị kiểm định	
		ADF	PP
Chuỗi gốc	TOURISTARRIVAL	-0,992	-1,601
Chuỗi sai phân bậc 1	DTOURISTARRIVAL	-8,274	-17,284

Ghi chú: Các giá trị tới hạn ở mức ý nghĩa thống kê 1%, 5% và 10% tương ứng là -3,34, -2,86 và -2,57.



Hình 3. Tương quan chuỗi biến động lượng khách du lịch quốc tế đến Việt Nam

Đối với chuỗi gốc cho thấy giá trị kiểm định ở cả 2 phương pháp đều vượt quá giá trị tới hạn ở mức ý nghĩa thống kê 1%, 5% và 10%. Vì vậy, giả thiết H_0 (chuỗi gốc là chuỗi không dừng) là không thể bác bỏ, tức chuỗi TOURISTARRIVAL là chuỗi không dừng, điều này cũng dễ hiểu bởi những chuỗi số liệu gốc thường không ổn định. Tuy nhiên, giả thiết H_0 ở chuỗi sai phân bậc 1 là bị bác bỏ bởi vì cả 2 phương pháp kiểm định ADF và PP đều cho thấy các giá trị kiểm định đều nhỏ hơn giá trị các giá trị tới hạn ở mức ý nghĩa thống kê 1%, 5% và 10%. Như vậy, cả 2 phương pháp kiểm định đều nhất quán và có thể kết luận chuỗi sai phân bậc 1 hay DTOURISTARRIVAL là chuỗi dừng.

3.2. Xây dựng mô hình ARIMA cho chuỗi biến động lượng khách quốc tế đến Việt Nam

Để xây dựng mô hình ARIMA chúng tôi sử dụng chuỗi dữ liệu gồm 192 quan sát từ tháng 1/1995 đến tháng 12/2010. Dữ liệu quá khứ được đặt tên là TOURISTARRIVAL sau đó lấy logarit tự nhiên trước khi lấy sai phân bậc nhất của TOURISTARRIVAL, ký hiệu là DLOGTOURISTARRIVAL.

Bước 1: Nhận dạng (xác định các giá trị p, d, q)

Chuỗi dữ liệu TOURISTARRIVAL kiểm định ở trên cho thấy chuỗi này dừng ở sai phân bậc 1, ta có $d=1$.

Để xác định p , Box & Jenkins (1976) đưa ra phương pháp nhận dạng như sau: một chuỗi dừng tự tương quan bậc p nếu (i) Các hệ số tự tương quan giảm từ từ theo dạng mũ hoặc hình sin, (ii) Các hệ số tương quan riêng phần giảm đột ngột xuống bằng 0 có ý nghĩa ngay sau độ trễ p .

Hình 3 cho thấy đồ thị tự tương quan và tương quan riêng phần của chuỗi DLOGTOURISTARRIVAL cho thấy tồn tại năm hệ số khác 0, có nghĩa tại các độ trễ 1, 5, 8, 9 và 12, trong đó sau độ trễ 1, 5, 9, 12 các hệ số tương quan riêng phần giảm đột ngột về giá trị bằng 0 có ý nghĩa. Như vậy, p có thể mang 1 trong các giá trị: 1, 5, 9, hoặc 12. Tương tự như cách xác định p , quan sát đồ thị tự tương quan và tương quan riêng phần của chuỗi DLOGTOURISTARRIVAL

ta nhận thấy q có thể mang một trong các giá trị: 1, 5, 12. Như vậy ta có mô hình ARIMA $(p,1,q)$ với các tổ hợp của p và q đã tìm thấy:

$$p \in \{1,5,9,12\} \text{ và } q \in \{1,5,12\}$$

Bước 2: Ước lượng mô hình

Để ước lượng các hệ số của các mô hình ARIMA $(p,1,q)$ như đã nhận dạng ở trên, phần mềm Eviews đã được sử dụng.

Bước 3: Kiểm tra

Để kiểm tra tính phù hợp của các mô hình chúng ta dựa trên tiêu chuẩn Schwarz (BIC) và sai số chuẩn SEE càng nhỏ càng tốt. Sau khi ước lượng thử các mô hình ARIMA có được bảng tổng hợp kết quả thống kê ở Bảng 3.

Bảng 3. Kết quả thống kê một số tiêu chuẩn của các mô hình ARIMA thử nghiệm

Mô hình ARIMA(p,d,q)	BIC	Hệ số tương quan điều chỉnh (R^2)	SEE
(1,1,1)	-1,14	0,17	0,13
(1,1,5)	-1,04	0,09	0,14
(1,1,12)	-1,04	0,08	0,13
(5,1,1)	-1,05	0,09	0,14
(5,1,5)	-0,98	0,02	0,14
(5,1,12)	-1,00	0,04	0,14
(9,1,1)	-0,99	0,05	0,14
(9,1,5)	-0,96	0,02	0,14
(9,1,12)	-0,97	0,03	0,14
(12,1,1)	-1,05	0,11	0,14
(12,1,5)	-1,00	0,07	0,14
(12,1,12)	-1,21	0,25	0,12

Bảng 4. Kết quả ước lượng mô hình ARIMA(12,1,12) cho chuỗi DLOGTOURISTARRIVAL

Tham số	Hệ số ước lượng	Sai số tiêu chuẩn	t-kiểm định	Xác suất
Hằng số	0,00456	0,01745	0,26120	0,79420
AR(12)	0,83744	0,04137	20,24339	0,00000
MA(12)	-0,79736	0,05702	-13,98352	0,00000

Bảng 5. Kết quả dự báo lượng khách du lịch quốc tế đến Việt Nam

Thời gian	Dự báo	Thực tế	Sai số (%)
Tháng 1/2011	443.853	506.424	14,10
Tháng 2/2011	426.238	542.671	27,32
Tháng 3/2011	611.361	475.733	22,18
Tháng 4/2011	558.587	460.000	17,65
Tháng 5/2011	609.786	480.866	21,14

Thực hiện kiểm tra, so sánh nhiều mô hình và nhận thấy mô hình ARIMA (12,1,12) là phù hợp nhất. Kết quả ước lượng mô hình được trình bày ở Bảng 4.

Đặt $Z_t = \text{DLOGTOURISTARRIVAL}$, ta có $Z_t = 0,0045 + 0,8374Z_{t-12} - 0,7973U_{t-12} + U_t$

Bước 4: Dự báo

Những dự báo ngắn hạn về lượng khách du lịch quốc tế đến Việt Nam dựa trên mô hình ARIMA (12, 1,12) được trình bày trong Bảng 5.

Bảng 5 cho thấy số liệu dự báo lượng khách quốc tế đến Việt Nam tháng 1/2011 khá bám sát với thực tế. Điều này cho thấy mô hình ARIMA (12,1,12) này đã giải thích được sự biến động của lượng khách du lịch quốc tế đến Việt Nam. Tuy nhiên dự báo các điểm tiếp theo có sai số lớn hơn, chính vì vậy cần cập nhật dữ liệu thường xuyên để đưa ra dự báo sát với thực tế hơn.

4. KẾT LUẬN

Biến động lượng khách quốc tế đến Việt Nam theo tháng là một chuỗi thời gian tuân theo quá trình tự hồi qui tích hợp với trung bình trượt với độ trễ thời gian là 12 hay ARIMA (12,1,12). Dựa vào mô hình này chúng ta có thể đưa ra dự báo ngắn hạn về lượng khách quốc tế đến Việt Nam. Kết quả dự báo những tháng đầu năm 2011 cho thấy mức độ sai số so với thực tế từ 14 đến 27%, điều này không phải là ngạc nhiên vì ngành du lịch là ngành chịu nhiều rủi ro. Tuy vậy, mô hình ARIMA có thể dùng để dự báo, song chưa phải là tối ưu, bởi vì sự phụ thuộc trong mô hình được giả định là tuyến tính.

Trong thời gian tới, các cơ quan quản lý du lịch cần có chiến lược cụ thể để qui hoạch phát triển, thu hút đầu tư, cân đối nguồn lực, hạn chế những rủi ro trong kinh doanh

du lịch. Đây chính là cơ sở để chuẩn bị tốt nguồn lực phục vụ du lịch, hoàn thiện công tác marketing và tạo thuận lợi cho các tổ chức, cá nhân hoạt động trong lĩnh vực này. Để sẵn sàng cho công tác tiếp đón du khách quốc tế, ngành du lịch cần chú trọng phát triển cơ sở hạ tầng du lịch, chuẩn bị tốt nguồn nhân lực du lịch đáp ứng yêu cầu về chất lượng, cơ cấu ngành nghề và tính chuyên nghiệp, tăng cường khai thác các công nghệ thông tin hiện đại, khai thác hiệu quả internet phục vụ công tác tuyên truyền quảng bá du lịch Việt Nam tại các thị trường trọng điểm, tăng cường hội nhập hợp tác quốc tế về du lịch.

TÀI LIỆU THAM KHẢO

- Box, G.E.P., and G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden Day, San Francisco.
- Chính phủ (2002). Quyết định Số 97/2002/QĐ-TTg của Thủ tướng Chính phủ về việc “*Phê duyệt Chiến lược phát triển du lịch Việt Nam 2001 - 2010*”.
- Do Q. Giam, Michael Mcaleer and Songsak Sriboonchitta (2009). "Effects of international gold market on stock exchange volatility: Evidence from ASEAN emerging stock markets ", *Economics Bulletin*, Vol. 29 no.2 pp. 599-610.
- Lê Minh (2010). Điều gì hạn chế du lịch Việt Nam cất cánh? Nguồn: <http://vccinews.vn/?page=detail&folder=112&Id=1271> (15/01/2012).
- Tổng cục du lịch (2008), Du lịch Việt Nam: Năng lực cạnh tranh còn thấp. Nguồn: <http://www.tourism.edu.vn/index.php?catid=05&itemid=157> (15/01/2012).
- Wang, Y., & Lim, C. (2005). Using time series models to forecast tourist flows. Proceedings of the 2005 International Conference of Simulation and Modelling, V. Kachitvichyanukul, U. Purintrapiban, P., Utayopas, eds.