

DỰ BÁO TIÊU THỤ DẦU MỎ CỦA MỸ BẰNG MÔ HÌNH ARIMA

Trần Ngọc Thùy Dung

Tóm tắt. Dầu mỏ là huyết mạch của nền kinh tế nhiều nước. Ngày nay, khi kinh tế ngày càng phát triển thì nhu cầu năng lượng nói chung và dầu mỏ nói riêng ngày càng tăng. Thế giới tiêu thụ khoảng 30 tỷ thùng (4,8 km³) dầu mỏ mỗi năm, trong đó các nước phát triển tiêu thụ nhiều nhất, đặc biệt là Mỹ. Việc dự đoán đúng và sớm mức tiêu thụ dầu mỏ sẽ đem lại lợi ích giúp bình ổn thị trường, bảo đảm cân đối cung - cầu. Bài viết này sử dụng phương pháp Box-Jenkins để xây dựng mô hình ARIMA cho dự báo tiêu dùng dầu mỏ của Mỹ. Kết quả cho thấy, trong số các mô hình ước lượng thử nghiệm thì ARIMA(2,1,1) có biến giả là phù hợp nhất. Các giá trị dự báo của mô hình này ít có sự khác biệt so với giá trị thực tế.

Từ khóa: ARIMA, dự báo, tiêu dùng dầu mỏ của Mỹ

1. ĐẶT VẤN ĐỀ

Dầu mỏ hay còn được gọi là vàng đen, đóng vai trò quan trọng trong nền kinh tế thế giới. Nguồn tài nguyên quý giá này không những mang lại nguồn ngoại tệ lớn khi xuất khẩu cho các quốc gia và các dân tộc trên thế giới đang sở hữu và trực tiếp kinh doanh mà còn là nguồn năng lượng quan trọng hiện nay cho sự phát triển kinh tế.

Hiện nay, dầu mỏ chiếm một tỷ lệ lớn trong tiêu thụ năng lượng trên toàn cầu. Trong các nước phát triển, Mỹ là quốc gia tiêu thụ dầu nhiều nhất. Và nguồn tài nguyên này có vai trò quan trọng trong quá trình thiết lập những sách lược chính trị của các quốc gia. Vì vậy, việc dự đoán đúng và sớm mức tiêu thụ dầu mỏ sẽ đem lại lợi ích to lớn giúp bình ổn thị trường, bảo đảm cân đối cung - cầu.

Với những lý do trên, việc xây dựng một mô hình dự báo về tiêu thụ dầu mỏ ở Mỹ là điều cần thiết. Trong bài viết này, phương pháp Box-Jenkins được sử dụng để xây dựng mô hình ARIMA (Autoregressive Integrated Moving Average) để dự báo tiêu dùng dầu mỏ của Mỹ, trên cơ sở các chuỗi dữ liệu được thống kê trong quá khứ.

2. CƠ SỞ LÝ THUYẾT VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Cơ sở lý thuyết

Hai tác giả George Box và Gwilym Jenkins [2] đã nghiên cứu mô hình tự hồi quy tích hợp trung bình trượt, viết tắt là ARIMA. Tên của Box và Jenkins được dùng để gọi cho các quá trình ARIMA tổng quát áp dụng vào phân tích và dự báo các chuỗi thời gian căn cứ vào chuỗi giá trị quá khứ. Mô hình ARIMA được thực hiện cho dữ liệu “dùng” (hay dữ liệu đã loại bỏ yếu tố xu thế).

Mô hình tự tương quan bậc p, viết tắt là 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:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t, (1)$$

trong đó: Y_t là quan sát dừng hiện tại, Y_{t-1}, Y_{t-2}, \dots là quan sát dừng quá khứ; φ_0 là hằng số; $\varphi_1, \varphi_2, \dots$ là các tham số phân tích hồi quy; ε_t là sai số dự báo ngẫu nhiên của giai đoạn hiện tại, có giá trị trung bình bằng 0 và phương sai không đổi σ^2 (được gọi là yếu tố nhiễu ngẫu nhiên thuần túy - white noise) và p là bậc tự hồi quy. Lưu ý rằng trong mô hình đưa bởi phương trình (1), chỉ có các giá trị hiện tại và quá khứ của Y được đưa vào mô hình, không có biến làm hồi quy nào khác.

Mô hình trung bình trượt bậc q , viết tắt là MA(q): Trong mô hình trung bình trượt, quá trình được mô tả hoàn toàn bằng tổng trọng số của các ngẫu nhiên hiện hành có độ trễ:

$$Y_t = \mu + \Theta_0 \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q} \quad (2)$$

trong đó: Y_t là quan sát dừng hiện tại; ε_t là sai số ngẫu nhiên hiện hành; $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ là các giá trị trễ của sai số ngẫu nhiên hiện hành; μ là hằng số; $\Theta_0, \Theta_1, \dots$ là các tham số phân tích hồi quy và q là bậc trung bình trượt.

Mô Hình Hồi Quy Kết Hợp Trung Bình Trượt - ARMA(p, q) là mô hình hỗn hợp của AR và MA. Hàm tuyến tính sẽ bao gồm những quan sát dừng quá khứ và những sai số dự báo quá khứ và hiện tại:

$$Y_t = \eta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \Theta_0 \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q} \quad (3)$$

trong đó: η là hằng số.

Mô hình ARIMA(p, d, q): do mô hình Box-Jenkins chỉ mô tả chuỗi dừng hoặc những chuỗi đã sai phân hóa, nên mô hình ARIMA(p, d, q) thể hiện những chuỗi dữ liệu không dừng, đã được sai phân (d chỉ mức độ sai phân).

Phương pháp Box - Jenkins bao gồm các bước sau:

Bước 1: Xem xét tính dừng của chuỗi quan sát

Mô hình ARIMA chỉ được áp dụng đối với chuỗi dừng. 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. Để nhận được dữ liệu có tính dừng, sử dụng sai phân của dữ liệu với Z_t là dữ liệu gốc.

Các bậc sai phân:

Sai phân bậc 0 là $I(0)$: chính là dữ liệu gốc Z_t .

Sai phân bậc 1 là $I(1)$: $w_t = Z_t - Z_{t-1}$

Sai phân bậc d ký hiệu là $I(d)$: $w_t = Z_t - Z_{t-d}$

Mô hình ARMA (p, q) áp dụng cho $I(d)$ được gọi là mô hình ARIMA (p, d, q).

Tính dừng của một chuỗi thời gian có thể được nhận biết dựa trên đồ thị của chuỗi thời gian và kiểm định Dickey-Fuller gia tăng (Augmented Dickey-Fuller, ký hiệu ADF)

Bước 2: Nhận dạng mô hình

Nhận dạng mô hình ARIMA (p,d,q) là tìm các giá trị thích hợp của p, d và q với p là bậc tự hồi quy, d là bậc sai phân của chuỗi thời gian được khảo sát, q là bậc trung bình trượt. Việc xác định p và q sẽ phụ thuộc vào các đồ thị SPACF (Sample Autocorrelation Function) và SPACF (Sample Partial Autocorrelation Function). Hàm tự tương quan mẫu (SACF) là một hàm hay đồ thị của độ tương quan của mẫu ở các độ trễ $k = 1, 2, \dots$. Hàm tự tương quan riêng phần của mẫu (SPACF) là một danh sách hay đồ thị của các trị số tương quan riêng phần của mẫu ở các độ trễ $k = 1, 2, \dots$. Các đồ thị này chỉ đơn giản là các điểm của SACF và SPACF vẽ theo độ trễ.

Chọn giá trị của p nếu đồ thị SPACF có giá trị cao tại các độ trễ 1, 2, ..., p và giảm đột ngột sau đó, đồng thời dạng hàm SAC tắt lịm dần. Tương tự, chọn giá trị của q nếu đồ thị SACF có giá trị cao tại độ trễ 1, 2, ..., q và giảm mạnh sau q, đồng thời dạng hàm SPAC tắt lịm dần.

Bước 3: Ước lượng các tham 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 (Ordinary Least Squares, ký hiệu OLS).

Bước 4: Kiểm tra chẩn đoán mô hình

Sau khi đã lựa chọn mô hình ARIMA cụ thể và ước lượng các tham số, 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 số hạng sai số ε_t của mô hình này có tính ngẫu nhiên thuần túy (white noise) hay không; nếu có, mô hình được chấp nhận; nếu không, phải tiến hành lại từ đầu. Các kiểm định có thể sử dụng là kiểm định BP (Box-Priore) hoặc kiểm định Ljung-Box với trị thống kê Q.

Bước 5: Dự báo bằng mô hình ARIMA

Dựa vào mô hình ARIMA phù hợp nhất, thực hiện dự báo điểm và dự báo khoảng tin cậy.

2.2. Phương pháp nghiên cứu

2.2.1. Phương pháp định lượng

Dự báo chuỗi thời gian dựa trên phân tích chuỗi quan sát của một biến duy nhất theo biến số độc lập là thời gian.

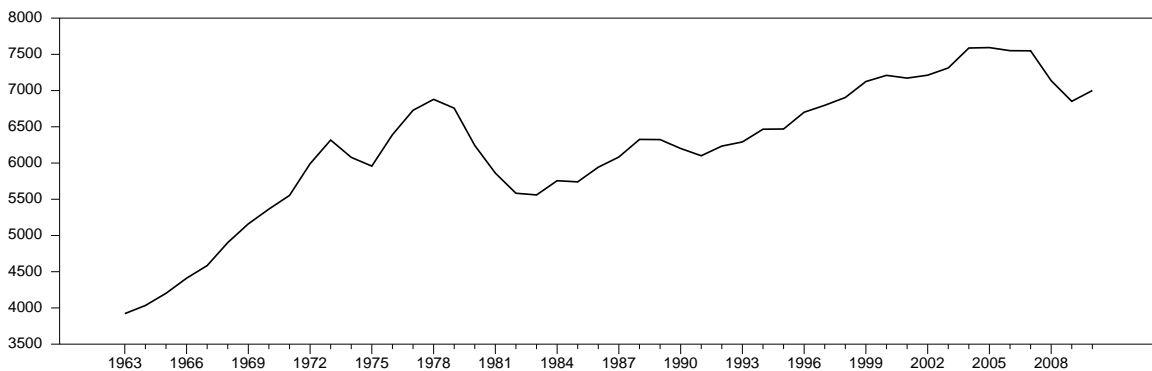
2.2.2. Thu thập và xử lý số liệu

Chuỗi dữ liệu được sử dụng gồm 48 quan sát từ năm 1963 đến năm 2010 của mức tiêu thụ dầu mỏ ở Mỹ.

Số liệu lấy tại website: www.inflationdata.com, được tổng hợp theo năm và xử lý trên phần mềm Excel và Winrats.

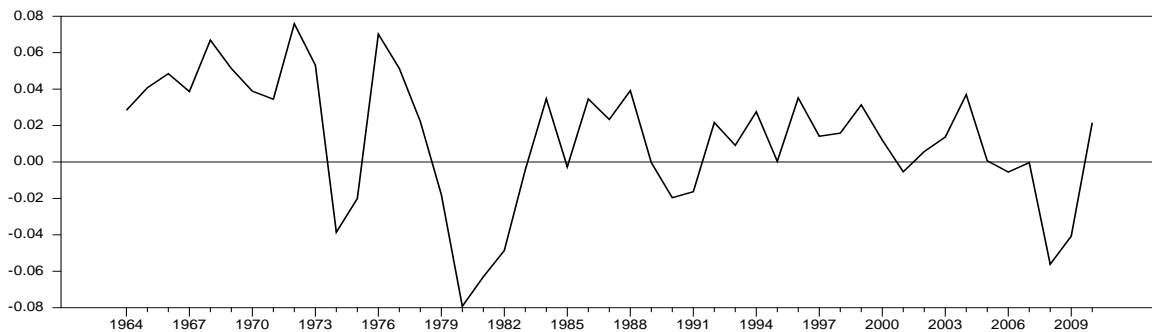
3. KẾT QUẢ VÀ THẢO LUẬN

Để xây dựng mô hình dự báo, sử dụng chuỗi dữ liệu gồm 48 quan sát trong giai đoạn mẫu (1963 - 2010). Các kết quả được đưa ra dựa trên việc sử dụng phần mềm Winrats. Chuỗi số liệu sử dụng trong mô hình ARIMA được giả định là chuỗi dừng. Để dự báo tiêu dùng dầu mỏ của Mỹ bằng mô hình ARIMA, đầu tiên cần phải xem xét liệu các chuỗi đó có phải là chuỗi dừng hay không. Muốn vậy, trước hết dựa vào quan sát trực tiếp đồ thị của chuỗi số liệu, sau đó tiến hành kiểm định.



Hình 1. Mức tiêu thụ dầu mỏ của Mỹ tính bằng triệu thùng/năm từ năm 1963 đến năm 2010

Kết quả trong Hình 1 cho thấy diễn biến mức tiêu thụ dầu mỏ của Mỹ (viết tắt là CONSP) theo năm (từ năm 1963 đến năm 2010) không ổn định và có xu hướng tăng. Có thể suy đoán rằng chuỗi số liệu gốc là chuỗi không dừng. Chuỗi CONSP được lấy logarit tự nhiên trước khi lấy sai phân bậc nhất để có được chuỗi sai phân bậc 1 (viết tắt là DLOGCONSP), chính là tốc độ tăng trưởng tiêu thụ dầu mỏ qua các năm. Chuỗi này không rõ xu hướng và xoay quanh một giá trị trung bình của chuỗi, thể hiện ở Hình 2. Đây được xem như là biểu hiện của một chuỗi dừng.



Hình 2. Tốc độ tăng trưởng tiêu thụ dầu mỏ của Mỹ từ năm 1963 đến năm 2010

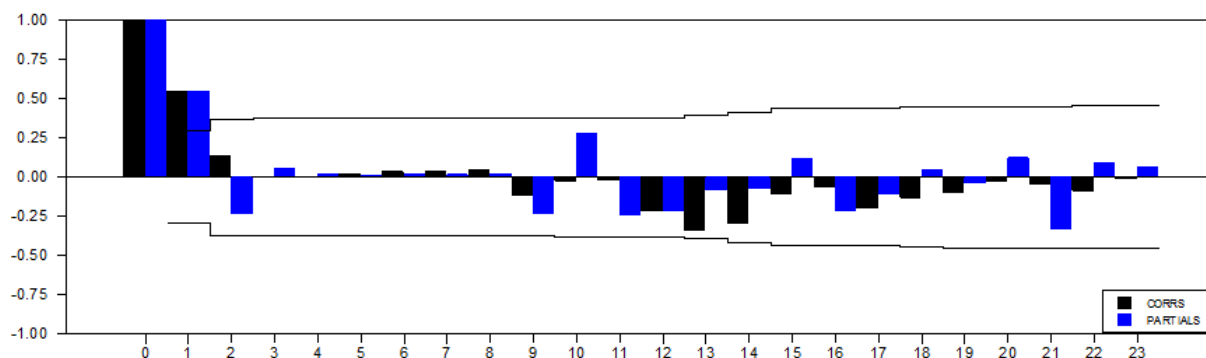
Thực hiện kiểm định ADF để xem liệu kết quả có nhất quán với những suy đoán trên hay không. Kết quả kiểm định tính dừng cho chuỗi số liệu gốc và chuỗi sai phân bậc 1 được trình bày ở Bảng 1.

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

<u>Chuỗi gốc</u>		<u>Chuỗi sai phân bậc 1</u>	
Sig Level	Crit Value	Sig Level	Crit Value
1%**	-	1%**	-
3.57117		3.57438	
5%*	-	5%*	-2.92414
2.92277		10%	-2.59972
10%	-		
2.59900			
		T-Statistic	-
T-Statistic	-	3.58956**	
2.63898			

Dựa vào Bảng 1, chuỗi gốc cho thấy giá trị kiểm định (T-Statistic) ở phương pháp ADF đều vượt quá giá trị tới hạn (Crit Value) ở mức ý nghĩa thống kê (Sig Level) 1%, 5% và 10%. Kiểm định ADF đã chấp nhận giả thiết rằng chuỗi gốc (CONSP) là chuỗi không dừng. Chuỗi sai phân bậc 1 cho thấy giá trị kiểm định ở phương pháp ADF nhỏ hơn giá trị tới hạn ở mức ý nghĩa thống kê 1%. Giả thiết chuỗi sai phân bậc 1 là chuỗi dừng được chấp nhận. Đồ thị trong Hình 2 và kết quả kiểm định Dickey – Fuller gia tăng cho thấy chuỗi DLOGCONSP là chuỗi dừng với mức ý nghĩa 1%.

Tiếp theo, xác định các giá trị p, d, q để nhận dạng mô hình ARIMA (p,d,q) cho dữ liệu dừng (chuỗi sai phân bậc 1). Chuỗi dữ liệu CONSP kiểm định ở trên cho thấy chuỗi này dừng ở sai phân bậc 1, nên d=1. Để xác định các giá trị p và q, đồ thị hàm tự tương quan mẫu và tự tương quan riêng phần mẫu của chuỗi DLOGCONSP được sử dụng.



Hình 3. Đồ thị của hàm tự tương quan mẫu và tự tương quan riêng phần mẫu của chuỗi DLOGCONSP

Ghi chú: -1.00,.....,1.00 : các hệ số tự tương quan và tự tương quan riêng phần
1,.....,23: các độ trễ

CORRS và PARTIALS là hàm tự tương quan mẫu và tự tương quan riêng phần mẫu

Đồ thị này cho thấy hàm SAC và SPAC có giá trị khá lớn ở các độ trễ 1 và 2, có giá trị nhỏ và rất nhỏ ở các độ trễ còn lại. Do vậy, p và q có thể nhận các giá trị là 1 và 2.

Kết quả trong hình 1 cũng cho thấy mức tiêu thụ dầu mỏ đã có sự biến động mạnh trong giai đoạn 1974-1984. Do đó, biến giả, ký hiệu DUM_{7481} , có thể được đưa vào mô hình ước lượng. Và có thể có 8 mô hình ARIMA. Dùng đồ thị hàm tự tương quan mẫu và tự tương quan riêng phần mẫu của chuỗi phần dư ϵ_t và kiểm định Ljung - Box để kiểm tra, thấy rằng cả 8 chuỗi ϵ_t của 8 mô hình đều là nhiễu trắng.

Tiếp theo, cần lựa chọn mô hình phù hợp nhất để sử dụng cho công tác dự báo theo Akaike Information Criterion (AIC) và Bayesian Information Criterion (BIC) [2]. Đây là hai tiêu chuẩn thông dụng nhất để so sánh các mô hình. Và mô hình có tiêu chuẩn AIC và BIC nhỏ nhất là mô hình được lựa chọn.

Ta có bảng xác định các tiêu chuẩn đánh giá sau khi đã thử với một vài mô hình khác nhau ở Bảng 2.

Bảng 2. Kết quả thống kê tiêu chuẩn AIC và BIC của các mô hình ARIMA thử nghiệm

Mô hình ARIMA(p,d,q)	AIC	BIC
ARIMA(1,1,1) không có biến giả	- 4.151	- 4.032
ARIMA(1,1,1) có biến giả	- 4.169	- 4.010
ARIMA(1,1,2) không có biến giả	- 4.108	- 3.949

ARIMA(1,1,2) có biến giả	- 4.126	- 3.927
ARIMA(2,1,1) không có biến giả	- 4.096	- 3.935
ARIMA(2,1,1) có biến giả	- 4.473	- 4.272
ARIMA(2,1,2) không có biến giả	- 4.199	- 3.998
ARIMA(2,1,2) có biến giả	- 4.349	- 4.108

Các kết quả ở Bảng 2 cho thấy mô hình ARIMA (2,1,1) có biến giả là mô hình có tiêu chuẩn AIC và BIC nhỏ nhất.

Sau khi thực hiện kiểm tra, so sánh nhiều mô hình, nhận thấy rằng mô hình ARIMA (2,1,1) có biến giả là phù hợp nhất. Kết quả ước lượng mô hình được trình bày ở Bảng 3.

Bảng 3: Kết quả ước lượng mô hình ARIMA(2,1,1) có biến giả cho chuỗi DLOGCONSP theo phương pháp OLS

Tham số	Hệ số ước lượng	Sai số tiêu chuẩn	T - kiểm định	Xác suất
Hằng số	0.003648824	0.052660698	0.06929	0.94510456
AR{1}	1.600511994	0.161157979	9.93132	0.00000000
AR{2}	-0.630235300	0.177956878	-3.54151	0.00102762
MA{1}	-1.291791537	0.247397946	-5.22151	0.00000583
Dum ₇₄₈₁	-0.024390178	0.024938267	-0.97802	0.33394169

Ghi chú : OLS (Ordinary Least Square) là phương pháp bình phương bé nhất

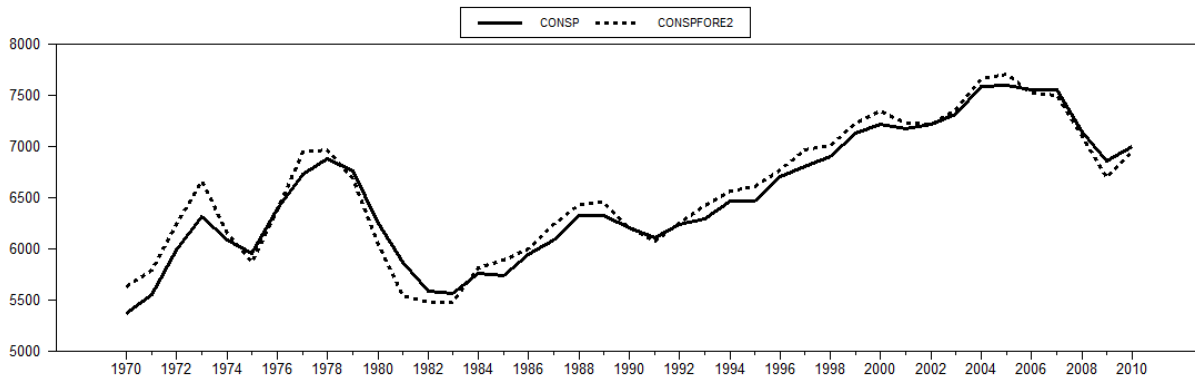
Đặt $Z_t = DLCONSP$ và dựa vào Bảng 3, phương trình ARIMA (2,1,1) có biến giả là:

$$Z_t = 0,004 + 1,601Z_{t-1} - 0,63Z_{t-2} - 1,29\varepsilon_{t-1} - 0,024Dum_{7481} + \varepsilon_t \quad (4)$$

(0,053) (0,161) (0,178) (0,247) (0,025)

Mức độ thích hợp của mô hình là $R^2 = 0.9580955$. Mô hình này giải thích được 95.8% sự biến động của tiêu dùng dầu mỏ ở Mỹ. Kết quả ở Bảng 3 chỉ ra rằng các hệ số hồi quy có ý nghĩa thống kê, nói cách khác là chúng khác 0 với mức ý nghĩa 5%, ngoại trừ hệ số hồi quy riêng và hệ số hồi quy của biến giả. Như vậy, biến giả không thật sự có ý nghĩa trong mô hình này. Tuy nhiên, trong số các mô hình ước lượng thử nghiệm thì ARIMA(2,1,1) có biến giả vẫn là phù hợp nhất. Lựa chọn giai đoạn dự báo bất kỳ trong khoảng thời gian từ năm 1963 đến năm 2010, cụ thể là 1970 - 2010. Tiến hành dự báo trong mẫu giai đoạn 1970 - 2010 dựa trên mô hình

ARIMA (2,1,1) có biến giả. Đồ thị về giá trị thực tế và giá trị dự báo với xác suất tin cậy 95% dựa trên mô hình ARIMA (2,1,1) có biến giả được trình bày trong Hình 4.



Hình 4. Giá trị thực tế và giá trị dự báo cho tiêu dùng dầu mỏ của Mỹ trong giai đoạn từ năm 1970 đến năm 2010

Ghi chú: Tiêu dùng dầu mỏ của Mỹ tính bằng triệu thùng/năm

Lưu ý rằng, đường nét liền (ký hiệu CONSSP) là các giá trị thực tế. Đường nét đứt (ký hiệu CONSPFORE2) là các giá trị của mô hình dự báo ARIMA (2,1,1) có biến giả từ năm 1970 đến năm 2010. Kết quả hình 4 cho thấy các giá trị dự báo tiêu thụ dầu mỏ của Mỹ là sát với các giá trị thực tế.

Cuối cùng, thực hiện việc kiểm tra tính chính xác của mô hình ARIMA (2,1,1) có biến giả bằng việc tiến hành kiểm định tính hợp lý và kiểm tra số hạng sai số ε_t .

Trước hết, tiến hành kiểm định tính hợp lý để kiểm tra sự khác biệt giữa giá trị thực tế đã có và giá trị dự báo trong giai đoạn 1970 - 2010. Việc kiểm tra được thực hiện thông qua kiểm định Holden - Peel [1]. Ước lượng các tham số (α, β) của mô hình hồi quy sau, theo phương pháp OLS:

$$y_t = \alpha + \beta y_{t-1} + u_t \quad (5)$$

Trong đó:

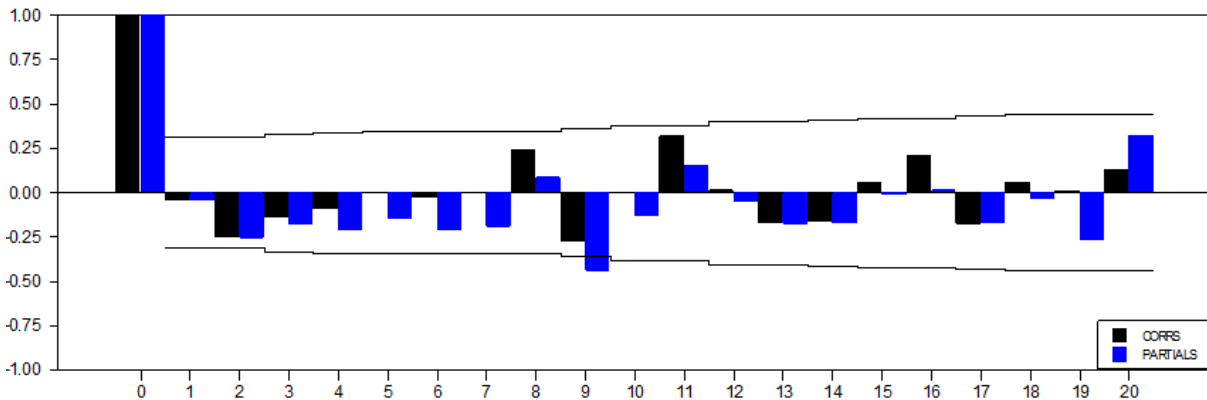
y_t, y_{t-1} lần lượt là giá trị thực tế và giá trị dự báo tại mốc thời gian trong giai đoạn từ năm 1970 đến năm 2010.

Giả thiết không của kiểm định là : $(\alpha, \beta) = (0, 1)$

Tiến hành kiểm định Holden - Peel, xác suất (signif = 0,59) lớn hơn 0,05 (5%). Giả thiết không của kiểm định được chấp nhận. Nghĩa là, giá trị dự báo sát với thực tế.

Sau đó, kiểm tra xem số hạng sai số ε_t của mô hình (phần dư ε_t ở phương trình (4)) có phải là ngẫu nhiên thuần túy (white noise) hay không. Việc kiểm tra này đã được tiến hành ở phần trước cho 8 chuỗi phần dư của 8 mô hình ARIMA. Đối với mô hình ARIMA (2,1,1) có biến giả được lựa chọn, các kết quả được trình bày cụ thể như sau:

Quan sát trực tiếp đồ thị của hàm tự tương quan mẫu (SACF) và tự tương quan riêng phần mẫu (SPACF) của chuỗi phần dư, sau đó là kết quả kiểm định.



Hình 5. Đồ thị hàm tự tương quan mẫu và tự tương quan riêng phần mẫu của chuỗi phần dư mô hình ARIMA (2,1,1) có biến giả

Ghi chú: -1,00.....1,00 là các hệ số tự tương quan và tự tương quan riêng phần

1....23 là các độ trễ

CORRS và PARTIALS là hàm tự tương quan mẫu và tự tương quan riêng phần mẫu

Quan sát đồ thị trong Hình 5, SPACF và SACF của chuỗi phần dư ε_t giảm đều theo dạng dao động lên xuống. Số hạng sai số ε_t là ngẫu nhiên thuần túy.

Kiểm định Ljung-Box về tương quan chuỗi với trị thống kê Q được sử dụng để xem kết quả có nhất quán với quan sát trên hay không. Kết quả kiểm định được trình bày ở Bảng 4.

Bảng 4. Kết quả thống kê của kiểm định Ljung-box

<u>Ljung-Box Q-Statistics</u>		
Lags	Statistic	Signif Lvl
2	2.983	0.225017
4	4.335	0.362605
6	4.358	0.628332
8	5.952	0.652602
10	8.610	0.569436
12	9.711	0.641304
14	12.041	0.602984

16	13.531
	0.633641
18	14.350
	0.705993
20	14.986
	0.777182

Ghi chú : Lags là độ trễ

Statistic là trị thống kê Q

Signif Lvl là mức ý nghĩa

Dựa vào Bảng 4, thấy rằng các trị thống kê Q không có ý nghĩa thống kê (Signif Lvl đều lớn hơn 5%). Kiểm định Ljung-Box về tương quan chuỗi đã chấp nhận giả thiết rằng chuỗi phần dư không tồn tại tự tương quan. Những kết quả này cho thấy số hạng sai số của mô hình ARIMA (2,1,1) có biến giả là ngẫu nhiên thuần túy (white noise).

Như vậy, mô hình ARIMA (2, 1, 1) có biến giả là một mô hình dự báo phù hợp.

4. KẾT LUẬN

Kết quả dự báo cho thấy giá trị dự báo xấp xỉ với giá trị thực tế. Điều này cho thấy độ tin cậy của mô hình dự báo là khá cao. Theo tiêu chuẩn AIC và BIC, trong số các mô hình ước lượng thử nghiệm thì ARIMA (2,1,1) có biến giả là phù hợp nhất để dự báo tiêu thụ dầu mỏ của Mỹ. Mô hình ARIMA là một mô hình tốt để dự báo trong ngắn hạn.

Bên cạnh những kết quả đạt được, đề tài nghiên cứu còn có thể bổ sung thêm một số điểm để xây dựng một mô hình dự báo hoàn thiện hơn. Đó là, tăng số lượng quan sát, bởi hiện tại số quan sát của bài nghiên cứu chưa nhiều (mới chỉ có 48 mẫu quan sát). Cần xây dựng các mô hình dự báo tổng hợp có cả mô hình ARIMA và mô hình hồi quy. Thực hiện dự báo dựa trên mô hình hồi quy để nối kết biến tiêu thụ dầu (biến phụ thuộc) và biến tăng trưởng tổng sản phẩm quốc nội GDP (biến độc lập). Mô hình tổng hợp này sẽ tạo ra được tiềm năng cải thiện các dự báo cả ngắn hạn và dài hạn.

TÀI LIỆU THAM KHẢO

- [1] K. Holden, D. A. Peel, and J. L. Thompson (1990) , *Economic forecasting: an introduction*, Cambridge University Press, Cambridge.
- [2] Ramanathan, Ramu (2002), *Introductory Econometrics with Applications*, 5th edition, Harcourt College Publishers.