

SỬ DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO GIÁ TRỊ XUẤT KHẨU CỦA VIỆT NAM

Ths. Bùi Thị Minh Nguyệt - Ths. Nguyễn Thị Quỳnh Nga - Ths. Nguyễn Thị Quỳnh Châm*

Nghiên cứu sử dụng mô hình ARIMA dự báo giá trị xuất khẩu của Việt Nam trong 6 tháng cuối năm 2018 bằng phần mềm Eviews 8. Kết quả nghiên cứu thực nghiệm cho thấy mô hình phù hợp nhất để dự báo giá trị xuất khẩu của Việt Nam với dữ liệu từ tháng 1 năm 2010 đến tháng 6 năm 2018 đó là ARIMA (1,1,16). Kết quả mô hình có giá trị dự báo với mức sai số khá nhỏ so với thực tiễn.

• Từ khóa: dự báo, xuất khẩu, ARIMA.

Research using ARIMA model predicts export value of Vietnam in the last 6 months of 2018 by Eviews software 8. The empirical results show that the most suitable model for forecasting export value of Vietnam With data from January 2010 to June 2018 that is ARIMA (1,1,16). The model results have a predictive value with a relatively small error rate compared to reality.

• Keywords: forecast, export, ARIMA.

Ngày nhận bài: 3/11/2018

Ngày chuyển phản biện: 7/11/2018

Ngày nhận phản biện: 15/11/2018

Ngày chấp nhận đăng: 20/11/2018

1. Đặt vấn đề

Từ khi Việt Nam ra nhập các tổ chức thương mại, ký kết các hiệp định thương mại song phương và đa phương đã mở ra cho Việt Nam phát huy nhiều thế mạnh. Trong đó xuất khẩu là một trong những chỉ tiêu kinh tế quan trọng của nền kinh tế quốc dân. Nhiều năm qua nước ta đã không ngừng chú trọng tăng xuất khẩu để phát triển nền kinh tế. Hoạt động xuất khẩu đã tạo nhiều công ăn việc làm và là một kênh giúp tăng nguồn thu ngoại tệ của Việt Nam. Dự báo giá trị xuất khẩu là hoạt động rất quan trọng đối với Chính phủ và các doanh nghiệp trong việc lập kế hoạch cho đơn vị của mình. Kết quả dự báo càng chính xác thì việc hoạch định chính sách càng khả thi. Đã có nhiều nghiên cứu mà trong đó có sử dụng mô

hình ARIMA dự báo tình hình xuất nhập khẩu các nước trên thế giới chẳng hạn như:

Ahmad Farooqi (2014) đã xây dựng được mô hình chuỗi thời gian ARIMA (Tự động chuyển đổi tích hợp tự động) với cách tiếp cận Box và Jenkins để dự báo tổng số hàng nhập khẩu và xuất khẩu của Pakistan với dữ liệu cụ thể từ năm 1947 đến năm 2013 dựa trên phần mềm thống kê hữu ích R. Kết quả nghiên cứu cho thấy rằng mô hình ARIMA (2, 2, 2) và ARIMA (1, 2, 2) phù hợp để dự báo nhập khẩu và xuất khẩu hàng năm của Pakistan tương ứng.

Gulshan Kumar, Sanjeev Gupta (2010) dự báo xuất khẩu hàng công nghiệp từ PUNJAB (Ấn Độ) đến năm 2020 bằng mô hình ARIMA.

Sanjeev Panwar, Susheel Kumar Sarkar, Anil Kumar, K.N. Singh, Samir Farooqi and Vipin Kumar Choudhary (2013) đã áp dụng mô hình ARIMA để dự báo xuất khẩu thịt và sản phẩm thịt hàng tháng từ Ấn Độ.

Jinzhao Lu (2015) Dự báo tổng xuất khẩu hàng dệt may của Mỹ sang thế giới trong 10 năm tiếp theo (2015-2025) với hai mô hình dự báo: mô hình hồi quy và mô hình ARIMA. Bằng cách sử dụng hai phương pháp dự báo khác nhau đã cho kết quả dự báo gần như nhau.

Do tính ứng dụng cao của phương pháp Box - Jenkins và hiệu quả của mô hình ARIMA trong lĩnh vực dự báo ngắn hạn. Trong bài nghiên cứu này chúng tôi vẫn sử dụng cách tiếp cận đó để lập mô hình và dự báo giá trị xuất khẩu ở Việt Nam trong 6 tháng tiếp theo của năm 2018. Số liệu nghiên cứu được thu thập từ Tổng cục

* Học viện Tài chính

Bảng 1. Thống kê số liệu xuất khẩu Việt Nam giai đoạn 2010 - 2018

DVT: Triệu USD

Tháng	Năm								
	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	5.013,00	7.091,00	7.095,00	11.472,00	11.460,00	13.405,00	13.800,00	14.600,00	19.000,00
2	3.740,00	4.848,00	8.300,00	7.146,00	9.540,00	9.600,00	10.300,00	13.000,00	13.400,00
3	5.592,00	7.447,00	9.479,00	11.029,00	12.000,00	13.345,00	14.200,00	16.300,00	19.800,00
4	5.332,00	7.437,00	8.962,00	10.033,00	13.071,00	13.348,00	14.100,00	16.700,00	18.200,00
5	6.312,00	7.233,00	9.100,00	11.675,00	12.000,00	13.500,00	14.600,00	17.200,00	19.200,00
6	6.317,00	8.460,00	9.750,00	11.009,00	12.378,00	14.300,00	14.800,00	17.800,00	19.600,00
7	6.029,00	9.323,00	9.600,00	11.599,00	12.916,00	14.500,00	14.700,00	17.500,00	
8	6.857,00	9.247,00	10.308,00	11.923,00	13.272,00	14.500,00	15.200,00	18.200,00	
9	6.098,00	7.944,00	9.483,00	11.184,00	12.400,00	14.200,00	15.000,00	19.000,00	
10	6.227,00	8.394,00	10.322,00	12.613,00	14.068,00	14.314,00	15.500,00	19.400,00	
11	6.641,00	8.854,00	10.347,00	11.992,00	13.200,00	14.300,00	15.600,00	19.200,00	
12	7.498,00	9.087,00	10.355,00	11.637,00	12.876,00	14.200,00	16.000,00	19.300,00	

Nguồn: Tổng cục Thống kê Việt Nam

Thống kê Việt Nam giai đoạn tháng 1/2010 đến tháng 06/2018.

2. Phương pháp nghiên cứu

Mô hình ARIMA (AutoRegressive Integrate Moving Average) do Box-Jenkins giới thiệu lần đầu tiên vào năm 1974. Mô hình ARIMA được kết hợp bởi 3 thành phần chính: AR (thành phần tự hồi quy); I (tính dừng của chuỗi thời gian); MA (thành phần trung bình trượt). ARIMA là mô hình dự báo định lượng theo thời gian, giá trị tương lai của biến số dự báo sẽ phụ thuộc vào xu thế vận động của chính đối tượng đó trong quá khứ. Mô hình ARIMA phân tích tính tương quan giữa các dữ liệu quan sát để đưa ra mô hình dự báo thông qua các giai đoạn nhận dạng mô hình, ước lượng các tham số từ dữ liệu quan sát và kiểm tra các tham số ước lượng để tìm ra mô hình thích hợp. Mô hình kết quả của quá trình trên gồm các tham số thể hiện mức độ tương quan trên dữ liệu, và được chọn để dự báo giá trị tương lai. Giới hạn độ tin cậy của dự báo được tính dựa trên phương sai của sai số dự báo.

Cơ sở toán học của mô hình chuỗi thời gian đơn biến là lý thuyết về quá trình ngẫu nhiên, việc tính toán được đảm bảo về mặt toán học. Các mô hình này cung cấp một công cụ dự báo vững chắc về mặt lý thuyết, có sức thuyết phục đối với các nhà nghiên cứu và ứng dụng. Mô hình ARIMA khá linh hoạt, thông qua nhiều cách kết hợp các tham số của mô hình, có thể lựa chọn để đưa ra một mô hình dự báo thích hợp nhất, phản ánh khách quan nhất kết quả dự báo. Mô hình ARIMA đặc biệt phù hợp cho các dự báo ngắn hạn và cho dữ liệu có yếu tố mùa vụ cao. Tuy nhiên, để dự

báo chính xác mô hình đòi hỏi nhiều quan sát (khoảng từ 50 quan sát trở lên).

Theo Gujarati (2004), phương pháp Box - Jenkins bao gồm 4 bước chung:

Bước 1. Nhận dạng mô hình

Để sử dụng mô hình ARIMA(p,d,q) trong dự báo cần nhận dạng ba thành phần p, d và q của mô hình.

- Thành phần d của mô hình được nhận dạng thông qua kiểm định tính dừng của chuỗi thời gian. Để kiểm tra

tính dừng của chuỗi thời gian ta có thể sử dụng đồ thị của Y_t theo thời gian, đồ thị tương quan mẫu hoặc kiểm định bước ngẫu nhiên (hay gọi là kiểm định nghiệm kiểm đơn vị Dickey-Fuller (1979) hoặc có thể kết hợp với kiểm định Phillips-Perron. Nếu chuỗi thời gian dừng ở bậc 0 ta ký hiệu I(d=0), nếu sai phân bậc 1 của chuỗi dừng ta ký hiệu I(d=1), nếu sai phân bậc 2 của chuỗi dừng ta ký hiệu I(d=2),...

- Sau khi kiểm định tính dừng, ta sẽ xác định bậc p của quá trình tự hồi quy (AR) và quy trình trung bình trượt bậc q của (MA) thông qua biểu đồ tự tương quan (ACF) và biểu đồ tự tương quan riêng phần (PACF).

Giả sử có chuỗi thời gian là $Y_t (t=1, 2, \dots, n)$. Giá trị dự báo của Y_t trong các thời kỳ $n+1, n+2, \dots, n+i$ được ký hiệu là Y_{t+i} .

Quy trình tự hồi quy bậc p, ký hiệu AR(p), được định nghĩa:

$$Y_t - \delta = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t$$

Trong đó: Y_t là chuỗi theo thời gian, δ là kỳ vọng của chuỗi Y_t , u_t là nhiễu trắng.

Giá trị p được xác định thông qua biểu đồ ACF và PACF. Nếu chuỗi có dạng AR(p) thì biểu đồ PACF sẽ có các hệ số tương quan riêng phần có ý nghĩa thống kê từ 1 tới p và các giá trị sau đó sẽ giảm nhanh về không, đồng thời ACF có các hệ số tương quan sẽ giảm dần về không.

Quy trình trung bình trượt bậc q, ký hiệu MA(q) được định nghĩa:

$$Y_t = \mu + \gamma_0 u_t + \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \dots + \gamma_q u_{t-q}$$

Nếu chuỗi có dạng MA(q) thì biểu đồ ACF sẽ có các hệ số tương quan có ý nghĩa thống kê từ 1 tới q và các giá trị sau đó sẽ giảm nhanh về không. Còn đối với PACF các hệ số tương quan riêng phần sẽ giảm dần về không.

Kết hợp hai dạng mô hình trên ta có mô hình ARMA(p,q) như sau:

$$Y_t = \theta + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \alpha_0 u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_q u_{t-q}$$

Bước 2. Ước lượng các tham số và lựa chọn mô hình

Các tham số của mô hình sẽ được ước lượng bằng phần mềm Eviews. Quá trình lựa chọn mô hình là quá trình thực nghiệm và so sánh các tiêu chuẩn R² hiệu chỉnh, AIC và Schwarz cho đến khi ta chọn được mô hình tốt nhất cho việc dự báo.

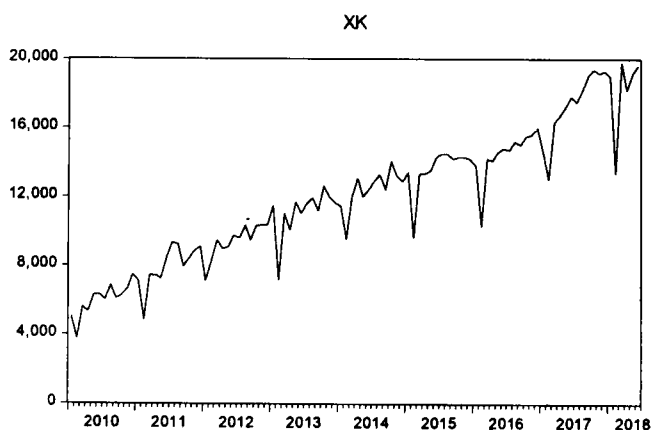
Bước 3. Kiểm định mô hình

Để kiểm định sự phù hợp của mô hình SARIMA được lựa chọn, bao gồm kiểm định các tham số và kiểm định sai số ngẫu nhiên. Sai số của mô hình phải là nhiễu trắng (white noise). Ta có thể sử dụng biểu đồ tự tương quan ACF hoặc kiểm định Breusch-Godfrey kiểm tra tính tự tương quan của sai số. Đối với phương sai sai số thay đổi, ta có thể sử dụng kiểm định White hoặc ARCH. Nếu kiểm định mô hình được lựa chọn không thỏa mãn thì quay lại từ giai đoạn nhận dạng để lựa chọn mô hình khác hợp lý hơn.

Bước 4. Dự báo

Dựa trên mô hình được lựa chọn thực hiện dự báo giá trị tương lai của dữ liệu chuỗi mùa vụ, cũng như đưa ra khoảng tin cậy của dự báo. Giá trị tương lai có thể được dự báo cho thời điểm kế tiếp hoặc mùa vụ kế tiếp.

Hình 1. Xuất khẩu Việt Nam giai đoạn 2010 - 2018



Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

3. Kết quả nghiên cứu

Dữ liệu được sử dụng là Tổng giá trị xuất khẩu của Việt Nam giai đoạn từ tháng 1 năm 2010 đến tháng 6 năm 2018 (ký hiệu XK, đơn vị tính: triệu USD), dữ liệu này được thu thập từ Tổng cục Thống kê 2018, tổng cộng 102 quan sát (Hình 1).

Dựa vào đồ thị phân bố cho thấy chuỗi XK là chuỗi có cả yếu tố xu thế và yếu tố mùa vụ. Ta thực hiện kiểm định tính dừng của chuỗi XK bằng kiểm định Dickey-Fuller (ADF) kết hợp kiểm định Phillips-Perron (PP). Kết quả cho thấy giá trị |t - Statistic| đều nhỏ hơn các giá trị tới hạn, đồng thời giá trị P_{value} = 0.9963 > 0.05; 0.01; 0.1, kết luận chuỗi không dừng ở cả 3 mức ý nghĩa. Kết quả được cho trong Bảng 2.

Như vậy, chuỗi dữ liệu ban đầu vừa có yếu tố xu thế và yếu tố mùa vụ đồng thời là chuỗi không dừng. Ta thực hiện hiệu chỉnh yếu tố mùa vụ theo phương pháp trung bình trượt dạng tổng thu được chuỗi dữ liệu ký hiệu XKSA (Bảng 3).

Kiểm định tính dừng của chuỗi XKSA bằng kiểm định ADF và PP cho kết quả trong Bảng 4.

Tiếp tục kiểm định tính dừng ở sai phân bậc 1 của chuỗi XKSA bằng hai kiểm định Dickey-Fuller và kiểm định Phillips-Perron đều cho kết quả: giá trị /t-stat/ lớn hơn giá trị tới hạn, các giá trị Pvalue đều nhỏ hơn các mức ý nghĩa, cho kết

Bảng 2. Kiểm định tính dừng của chuỗi XK

Kiểm định	T - Statistic	P - Value
ADF	0.995633	0.9963
PP	-1.2294	0.6594

Nguồn: Tính toán bởi tác giả

Bảng 3. Bảng dữ liệu XKSA

Year	Value
2010M01	5273.710
2010M02	6449.222
2010M03	5436.108
2010M04	5352.603
2010M05	6249.056
2010M06	5943.490
2010M07	5624.192
2010M08	6157.853
2010M09	6047.816
2010M10	5626.812
2010M11	6262.957
2010M12	7232.078
2011M01	7351.710
2011M02	7557.222
2011M03	7291.109
2011M04	7457.603
2011M05	7170.056
2011M06	8086.490
2011M07	8918.192
2011M08	

Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

luận chuỗi đã dừng ở bậc 1. Kết quả kiểm định Dickey-Fuller kết hợp kiểm định Phillips-Perron được thể hiện ở Bảng 5.

Do chuỗi XKSA dừng ở sai phân bậc 1, ta sẽ tạo chuỗi sai phân của xuất khẩu ký hiệu DXK

Bảng 4. Chuỗi XKSA không dừng ở sai phân bậc 0

Kiểm định	T - Statistic	P - Value
ADF	0.995633	0.9963
PP	-1.2294	0.6594

Nguồn: Tính toán bởi tác giả

Bảng 5. Kiểm định tính dừng chuỗi XKSA ở sai phân bậc 1

Kiểm định	T - Statistic	P - Value
ADF	-12.17322	0.0001
PP	-8.58584	0.0000

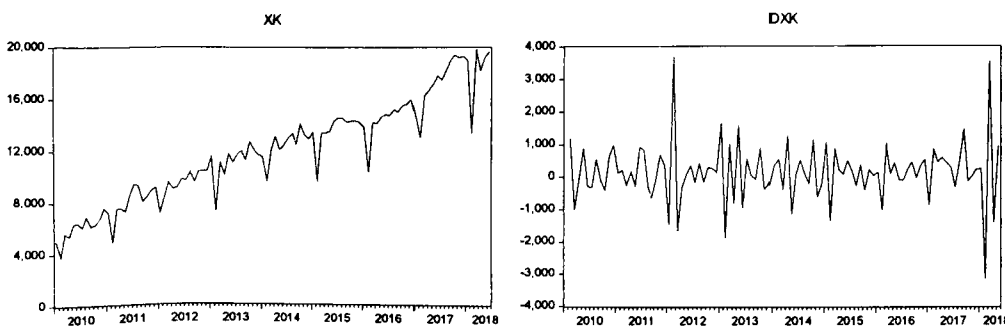
Nguồn: Tính toán bởi tác giả

Bảng 6. Bảng dữ liệu chuỗi DXK

Year	Value
2010M01	NA
2010M02	1175.512
2010M03	-1013.113
2010M04	-83.50595
2010M05	896.4524
2010M06	-305.5855
2010M07	-319.2984
2010M08	533.6615
2010M09	-109.9375
2010M10	-421.1042
2010M11	636.1458
2010M12	969.1220
2011M01	119.6310
2011M02	205.5119
2011M03	-266.1131
2011M04	166.4940
2011M05	-287.5476
2011M06	916.4345
2011M07	

Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

Hình 2. Chuỗi XK không dừng và dừng sau khi loại bỏ xu thế



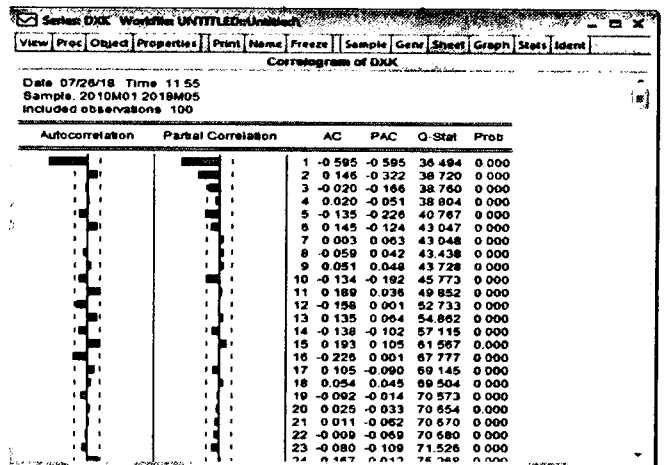
Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

thu được từ chuỗi XKSA. Khi đó sẽ sử dụng chuỗi này để thực hiện dự báo (Bảng 6).

b. Ước lượng các tham số và lựa chọn mô hình

Do chuỗi dừng để xác định giá trị p, q của mô hình ARIMA ta phải dựa vào biểu đồ ACF và PACF. Dựa vào biểu đồ PACF ở Hình 1 ta thấy, các hệ số tương quan khác không ở các độ trễ 1,2,5. Còn đối với biểu đồ ACF, ta có các hệ số tương quan riêng phần khác nhau ở các độ trễ 1,16 (Bảng 7).

Bảng 7. Biểu đồ PACF và ACF của xuất khẩu



Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

Để tìm ra mô hình dự báo phù hợp nhất ta phải dùng phương pháp thực nghiệm bằng cách so sánh các chỉ số R² hiệu chỉnh, AIC và Schwarz. Kết quả so sánh cho thấy mô hình ARIMA(1,1,16) là mô hình phù hợp nhất đối với bộ dữ liệu đã cho. Bảng sau đây thể hiện kết quả hồi quy của mô hình đã được lựa chọn.

c. Kiểm định mô hình

Để kiểm định mô hình ARIMA(1,1,16) có vi phạm các giả định của mô hình hồi quy không, ta thực hiện thêm một số kiểm định. Kiểm định White cho thấy mô hình không có phương sai sai số thay đổi. Kiểm định Breusch-Godfrey cho thấy sai số không có tự tương quan. Ta có thể kết luận mô hình trên thích hợp cho việc dự báo (Bảng 9).

Bảng 8. Kết quả hồi quy mô hình ARIMA(1,1,16)

Dependent Variable: DXK
 Method: Least Squares
 Date: 08/02/18 Time: 20:30
 Sample (adjusted): 2010M03 2018M06
 Included observations: 100 after adjustments
 Convergence achieved after 22 iterations
 MA Backcast: 2008M11 2010M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	127.6033	10.16695	12.55080	0.0000
AR(1)	-0.235446	0.113623	-2.072161	0.0409
MA(1)	-0.696734	0.069652	-10.00303	0.0000
MA(16)	-0.254451	0.061102	-4.164350	0.0001
R-squared	0.509246	Mean dependent var	127.7727	
Adjusted R-squared	0.493910	S.D. dependent var	897.7136	
S.E. of regression	638.6336	Akaike info criterion	15.79572	
Sum squared resid	39153881	Schwarz criterion	15.89992	
Log likelihood	-785.7858	Hannan-Quinn criter.	15.83789	
F-statistic	33.20576	Durbin-Watson stat	1.966515	
Prob(F-statistic)	0.000000			

Nguồn: tính toán bởi tác giả từ phần mềm Eviews

d. Dự báo

Sử dụng mô hình ARIMA(1,1,16) tiến hành dự báo cho giá trị xuất khẩu. Theo tính toán với kết quả dự báo 6 tháng đầu năm tương đối sát với thực tế với sai số thấp nhất 0,7%, sai số trung bình là 2%. Với các mô hình dự báo ứng dụng trong kinh tế - xã hội nói chung và cho giá trị xuất khẩu nói riêng thì với mức sai số như vậy là chấp nhận được và do đó, mô hình có thể vận dụng trong thực tiễn. Kết quả dự báo trong mẫu được thể hiện trong Bảng 10.

Bảng 9. Kết quả kiểm định Breusch-Godfrey và White

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.013611	Prob. F(12,84)	0.4441
Obs*R-squared	12.01994	Prob. Chi-Square(12)	0.4441

Heteroskedasticity Test: ARCH

F-statistic	0.099904	Prob. F(1,97)	0.7526
Obs*R-squared	0.101859	Prob. Chi-Square(1)	0.7496

Nguồn: Tính toán bởi tác giả từ phần mềm Eviews

Do đó, kết quả dự báo ngoài mẫu cho 6 tháng tiếp theo của năm 2018 có thể nói là đáng tin cậy (Bảng 11).

Từ kết quả dự báo của mô hình ta thấy tình hình xuất khẩu 6 tháng cuối năm 2018 vẫn có xu hướng tăng cao. Theo kết quả dự báo tổng Kim ngạch xuất khẩu 6 tháng cuối năm 2018 ước tăng 115 tỷ USD, tăng 1,3% so với 6 tháng đầu năm (Hình 3).

4. Kết luận và kiến nghị

Bài viết đã nghiên cứu khả năng ứng dụng của mô hình ARIMA vào việc dự báo nhằm tìm ra mô hình tốt nhất cho việc dự báo giá trị xuất khẩu tại Việt Nam. Kết quả nghiên cứu cho thấy ứng với chuỗi dữ liệu đã thu thập thì mô hình ARIMA(1,1,16) cho kết quả dự báo giá trị xuất khẩu của Việt Nam tương đối sát với thực tế. Tuy nhiên, kết quả dự báo trong bài viết này chỉ là kỹ thuật dựa trên dữ

Bảng 10. Kết quả dự báo XK trong mẫu

Tháng	Tháng 1/2018	Tháng 2/2018	Tháng 3/2018	Tháng 4/2018	Tháng 5/2018	Tháng 6/2018
Giá trị thực	19.267	16.115	19.650	18.227	19.143	19.223
Giá trị dự báo	18.994	19.167	19.088	18.465	18.998	18.974
Sai số dự báo	- 0,01	- 0,18	- 0,02	0,01	- 0,007	- 0,01

Nguồn: Tính toán bởi tác giả

Bảng 11. Kết quả dự báo XK ngoài mẫu bằng mô hình ARIMA (1,1,16)

Tháng	Tháng 7/2018	Tháng 8/2018	Tháng 9/2018	Tháng 10/2018	Tháng 11/2018	Tháng 12/2018
XK dự báo	19.090	19.114	19.080	19.144	19.344	19.467

Nguồn: Tính toán bởi tác giả

liệu. Ta có thể kết hợp nhiều phương pháp dự báo khác nhau hoặc thử nghiệm trên các mô hình dự báo khác nhau để có thể so sánh kết quả dự báo nhằm xác định một phương pháp dự báo tối ưu nhất.

Kết quả dự báo của mô hình cho thấy giá trị xuất khẩu Việt Nam 6 tháng cuối năm 2018 vẫn có xu hướng tăng cao hơn. Cho thấy tình hình xuất khẩu của Việt Nam vẫn sẽ tiếp tục khởi sắc. Tuy

(Xem tiếp trang 65)