

ỨNG DỤNG MÔ HÌNH ARIMA ĐỂ DỰ BÁO CHỈ SỐ GIÁ TIÊU DÙNG Ở VIỆT NAM

Huyền Tấn Nguyên

Trường Đại học Công nghệ Đồng Nai

Nguyễn Văn Lượng

Trường Trung cấp Nghề số 9 Quảng Bình

Tóm tắt. Nghiên cứu này ứng dụng mô hình ARIMA trong dự báo chỉ số giá tiêu dùng CPI ở Việt Nam trong quý 2 năm 2016. Số liệu được thu thập từ Tổng cục Thống kê Việt Nam giai đoạn tháng 1/2010 đến tháng 03/2016 và được xử lý bằng phần mềm Eviews 6. Kết quả cho thấy mô hình ARIMA(2,1,1) là thích hợp cho việc dự báo. Kết quả dự báo CPI quý 2 năm 2016 lần lượt là: 159.5409%, 159.5447%, 159.5476%. Các nhà làm chính sách, các doanh nghiệp có thể sử dụng công cụ này cho công tác dự báo, hoạch định và lập kế hoạch.

Từ khoá: Dự báo, CPI, ARIMA, phương sai

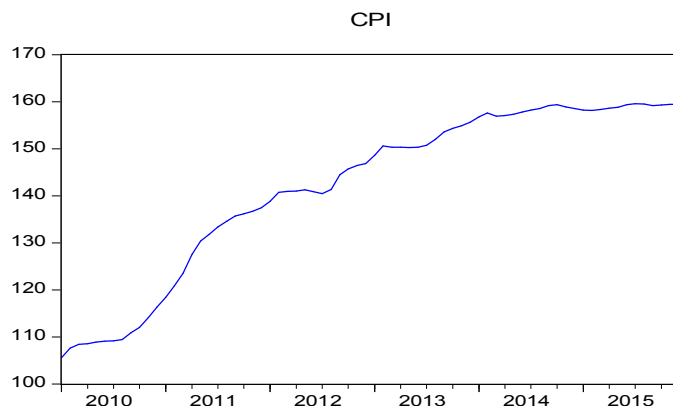
1. ĐẶT VẤN ĐỀ

Dự báo chỉ số giá tiêu dùng (CPI) là hoạt động rất quan trọng đối với các 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. Theo Daehoon Nahm (2015), tình hình CPI ngày càng khó dự báo do sự biến động của tình hình kinh tế quốc gia và thế giới [1]. Vài thập kỷ qua, có nhiều mô hình được sử dụng để dự báo CPI: Marcos Álvarez-Díaz & Rangan Gupta (2016) đã sử dụng mô hình bước ngẫu nhiên (random walk) và mô hình tự hồi quy bậc 1 (AR(1)) để dự báo CPI của Mỹ [2], Al-Tamimi et al., (2011) đã sử dụng mô hình hồi quy bội để dự báo CPI của UEA giai đoạn 1990 - 2005 [3], Hu et al., (2013) sử dụng thuật toán thống kê để dự báo CPI [4], Bernardi et al., (2015) sử dụng phương pháp dự báo san hàm mũ Holt – Winter để dự báo CPI của Italia giai đoạn 2004 – 2014 [5]. Sự phát triển, cải thiện của các công cụ dự báo bằng toán học như trên đã tạo ra những kỹ thuật dự báo chính xác và sai số thấp.

Các phương pháp dự báo trên đều có những ưu và nhược điểm riêng phụ thuộc vào dữ liệu thống kê. Theo Robert et al., (1979) mô hình ARIMA rất phù hợp đối với những quan hệ tuyến tính giữa dữ liệu hiện tại và dữ liệu quá khứ [6]. Hơn nữa, Brockwell et al., (2001) còn cho rằng mô hình ARIMA sẽ dự báo chính xác hơn khi số liệu được thống kê chi tiết theo từng tháng trong năm [7].

Nghiên cứu này ứng dụng mô hình ARIMA để dự báo chỉ số giá tiêu dùng của Việt Nam trong quý 2 năm 2016. Số liệu nghiên cứu được thu thập từ Tổng cục Thống kê Việt Nam giai đoạn tháng 1/2010 đến tháng 03/2016 [8], kỳ được chọn làm gốc để tính CPI là năm 2009. Phương pháp tính CPI được Tổng cục Thống kê Việt Nam ứng dụng là phương pháp Paasche (lượng hàng hóa cố định ở kỳ nghiên cứu),

cụ thể:
$$I_P^{Paasche} = \frac{\sum_{i=1}^n P_{oi} Q_{oi}}{\sum_{i=1}^n P_{oi} Q_{oi}} \times 100\%$$



Hình 1. CPI Việt Nam giai đoạn 2010 – 2015.

(Nguồn: Tổng cục Thống kê Việt Nam, 2016 [8])

Mặt khác, nghiên cứu này muốn tính toán sai số dự báo của phần dữ liệu tham khảo (in-sample) và sai số của phần kiểm tra (out-of-sample), do đó dữ liệu thu thập được chia làm 2 phần gồm: dữ liệu tham khảo gồm các quan sát từ tháng 01/2010 đến tháng 12/2015 và dữ liệu kiểm tra gồm các quan sát của tháng 01/2016, tháng 02/2016, và tháng 3/2016.

Bảng 1. Thống kê chỉ số giá tiêu dùng giai đoạn 2010 – 2016 (ĐVT: %)

Tháng	Năm						
	2010	2011	2012	2013	2014	2015	2016
1	105.57	118.41	138.86	148.67	156.78	158.25	159.5234
2	107.64	120.89	140.76	150.64	157.64	158.17	159.5345
3	108.44	123.51	140.98	150.35	156.95	158.41	159.5411
4	108.59	127.61	141.06	150.38	157.07	158.63	
5	108.89	130.43	141.31	150.29	157.38	158.88	
6	109.13	131.85	140.94	150.37	157.86	159.44	
7	109.19	133.39	140.53	150.77	158.22	159.65	
8	109.44	134.63	141.42	152.02	158.57	159.53	
9	110.88	135.74	144.53	153.63	159.2	159.2	
10	112.04	136.23	145.76	154.39	159.38	159.37	
11	114.13	136.76	146.44	154.91	158.95	159.49	
12	116.39	137.48	146.84	155.7	158.57	159.51	

(Nguồn: Tổng cục Thống kê Việt Nam năm 2016 [8]).

2. PHƯƠNG PHÁP NGHIÊN CỨU

Box & Jenkins (1970) lần đầu tiên giới thiệu mô hình ARIMA (autoregressive integrated moving average) trong phân tích chuỗi thời gian, được hiểu là phương pháp Box-Jenkins [9]. Mô hình ARIMA được kết hợp bởi 3 thành phần chính: AR (thành phần tự hồi quy), I (tính dừng của chuỗi thời gian) và MA (thành phần trung bình trượt) [9]. Theo Gujarati (2006) và R.Carter Hill et al., (2011) để sử dụng mô hình ARIMA trong dự báo chuỗi thời gian, cần trải qua 4 bước như sau [10], [11]:

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. 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),... Để kiểm định tính dừng của chuỗi, chúng tôi sử dụng kiểm định nghiệm đơn vị Dickey–Fuller cải biên (ADF) và kiểm định Phillips-Perron [12]:

$$\Delta Y_t = \beta_0 + \beta_1 t + \pi Y_{t-1} + \sum_{j=1}^p \psi_j \Delta Y_{t-j} + \varepsilon_t \quad (1)$$

Kiểm định giả thuyết $\begin{cases} H_0: \pi = 0 \\ H_1: \pi < 0 \end{cases}$ sử dụng thống kê student (ký hiệu t)

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

Quy trình tự hồi quy bậc p, ký hiệu AR(p), được định nghĩa [9], [10]:

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

trong đó: Y_t là chuỗi theo thời gian, δ là kỳ vọng của chuỗi Y_t , u_t là nhiễu trắng (white noise).

Quy trình trung bình trượt bậc q, ký hiệu MA(q) được định nghĩa [9] [10]:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3)$$

Bản chất của mô hình (2) là kết hợp tuyến tính giữa Y_t và các nhiễu trắng.

Kết hợp (2) và (3) ta có mô hình ARMA(p,q) như sau:

$$Y_t = \Theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (4)$$

Nhận dạng mô hình ARIMA(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 PACF = f(t) và ACF = f(t), chi tiết được thể hiện ở Bảng 2.

Bảng 2. Các dạng đồ thị của mô hình ARIMA

Loại mô hình	Dạng đồ thị ACF	Dạng đồ thị PACF
AR(p)	Giảm nhanh theo hàm mũ hoặc dạng hình sin, hoặc cả hai	Có đỉnh ở trễ thứ p
MA(q)	Có đỉnh sau trễ q	Giảm dần theo hàm mũ
ARMA(p, q)	Giảm nhanh theo hàm mũ	Giảm nhanh theo hàm mũ

(Nguồn: Gujarati et al., [10])

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 Eview. Quá trình lựa chọn mô hình là quá trình thực nghiệm và so sánh các tiêu chí R^2 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 đảm bảo mô hình là phù hợp, sai số của mô hình phải là nhiễu trắng. 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.

Bên cạnh đó để đánh giá độ tin cậy của mô hình dự báo, nghiên cứu sử dụng chỉ số MAPE (Mean Absolute Percent Error). Theo Lewis (1983) thì MAPE lớn hơn hoặc bằng 50% thì dự báo không chính xác, 20% - 50% là hợp lệ, 10%-20% là dự báo tốt, dưới 10% là dự báo hoàn hảo [13]. Chỉ số MAPE được định nghĩa như sau [14]:

$$MAPE = \frac{100}{n} \times \sum_{t=2}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (5)$$

trong đó x_t , \hat{x}_t là giá trị thật và giá trị dự báo ở thời điểm t, n là tổng số dự báo.

Bước 4. Dự báo

Sau khi kiểm định sai số của các mô hình dự báo, nếu phù hợp sẽ được sử dụng vào việc dự báo.

3. KẾT QUẢ NGHIÊN CỨU

3.1. Xây dựng mô hình ARIMA

Để xây dựng mô hình ARIMA, trước tiên ta phải kiểm định tính dừng của chuỗi CPI. Kết quả kiểm định Dickey-Fuller (ADF) và Phillips-Perron (PP) cho thấy giá trị /t-stat/ đều bé hơn giá trị tới hạn. Vì vậy, kết luận chuỗi CPI không dừng.

Bảng 3. Kiểm định tính dừng chuỗi CPI

Kiểm định	Giá trị t-sta	Xác suất
ADF	-0.817240	0.9578
PP	-0.548331	0.9783

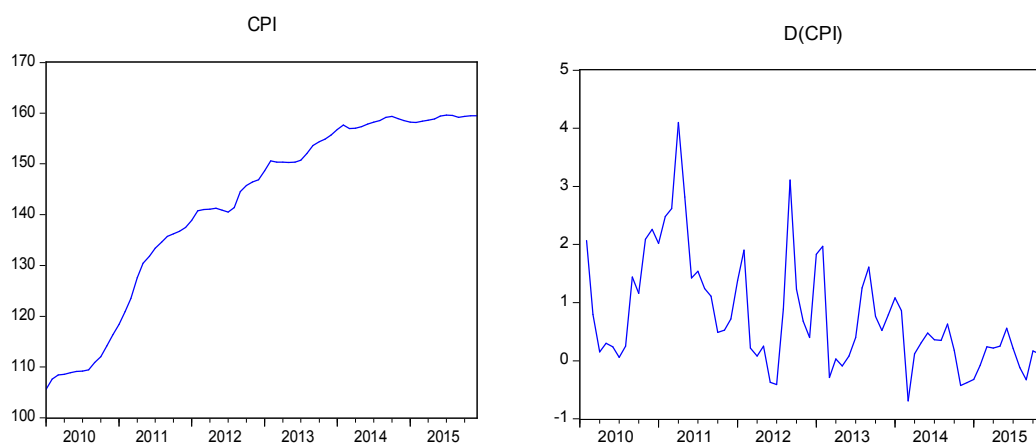
Ghi chú: Các giá trị tới hạn ở mức ý nghĩa 1%, 5% và 10% tương ứng là: -4.121303, -3.487845, -3.172314

Bảng 3 cho thấy chuỗi CPI không dừng tại sai phân bậc 0 nên phải lấy sai phân bậc 1 để kiểm định tính dừng. Sau khi lấy sai phân bậc 1, chuỗi CPI đã dừng. Kết quả kiểm định ADF và PP được thể hiện ở Bảng 4 và Hình 2.

Bảng 4. Kiểm định tính dừng chuỗi CPI sau khi lấy sai phân bậc 1

Kiểm định	Giá trị t sta	Xác suất
ADF	-3.776261	0.0250
PP	-3.789125	0.0242

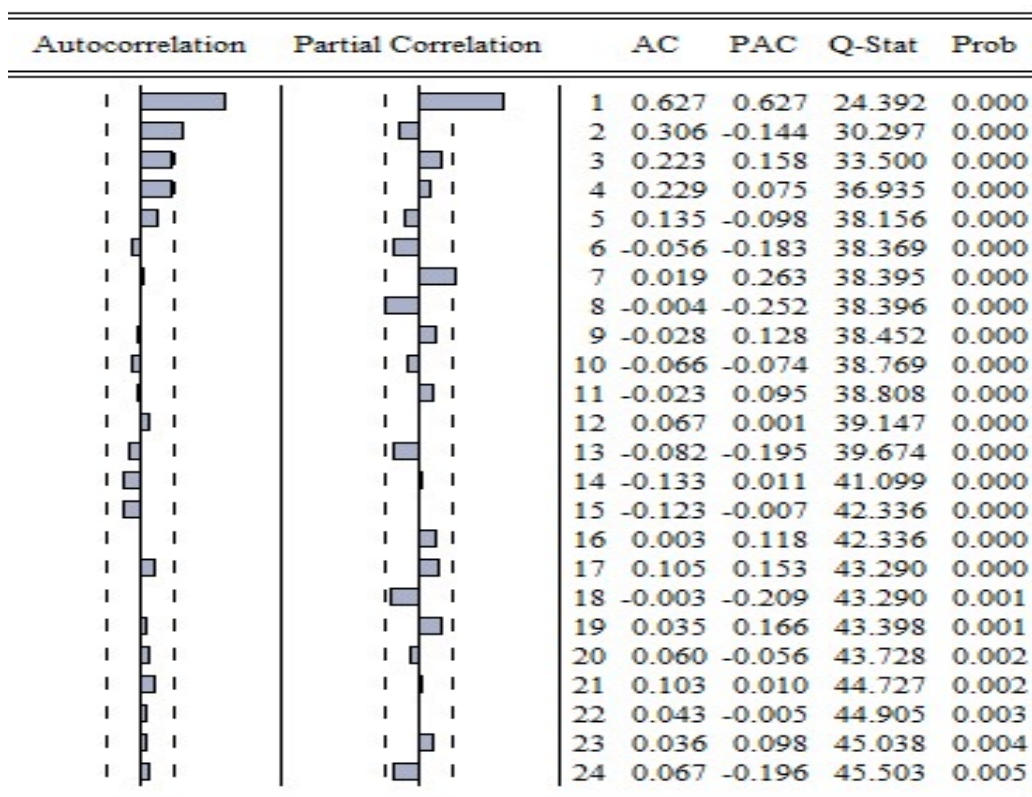
Ghi chú: Các giá trị tới hạn ở mức ý nghĩa 1%, 5% và 10% tương ứng là: -4.121303, -3.487845, -3.172314



Hình 2. Chuỗi CPI không dừng và CPI đã dừng khi lấy sai phân bậc 1.

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

Qua Hình 3 cho thấy: Để xác định giá trị p , q của mô hình ARIMA cần phải dựa vào biểu đồ ACF và PACF các hệ số tương quan khác 0 ở các độ trễ 1, 2. Còn đối với biểu đồ PACF, ta có các hệ số tương quan riêng phần khác 0 ở các độ trễ 1.



Hình 3. Biểu đồ PACF và ACF của chuỗi sai phân bậc 1 CPI.

Để tìm ra mô hình dự báo phù hợp nhất phải dùng phương pháp thực nghiệm bằng cách so sánh các chỉ số R^2 hiệu chỉnh, AIC và Schwarz. Kết quả so sánh cho thấy mô hình ARIMA(2,1,1) 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.

Bảng 5. Kết quả hồi quy mô hình ARIMA(2,1,1)

Dependent Variable: D(CPI)

Method: Least Squares

Sample (adjusted): 2010M04 2015M12

Included observations: 69 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.723846	0.284997	2.539835	0.0135
AR(2)	0.457587	0.136955	3.341152	0.0014
MA(1)	0.914103	0.066091	13.83100	0.0000
R-squared	0.517459	Mean dependent var		0.740145
Adjusted R-squared	0.502837	S.D. dependent var		0.951581
S.E. of regression	0.670958	Akaike info criterion		2.082285
Sum squared resid	29.71220	Schwarz criterion		2.179420

Log likelihood	-68.83883	Hannan-Quinn criter.	2.120822
F-statistic	35.38802	Durbin-Watson stat	2.137175
Prob(F-statistic)	0.000000		
Inverted AR Roots	.68	-.68	
Inverted MA Roots	-.91		

3.3. Kiểm định mô hình

Để biết mô hình ARIMA(2,1,1) có vi phạm các giả định của mô hình hồi quy không, cần 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, kiểm định Jacque-Bera cho thấy sai số ngẫu nhiên có phân phối chuẩn (Phụ lục).

Hơn nữa, việc tính toán và so sánh chỉ số MAPE của 2 mô hình ARIMA(2,1,1) và ARIMA(1,1,1) đã cho thấy ARIMA(2,1,1) có độ chính xác cao hơn xét cả về dữ liệu tham khảo và dữ liệu kiểm tra (Bảng 6).

Bảng 6. Tổng hợp các chỉ số đánh giá độ chính xác của mô hình dự báo.

Mô hình	MAPE (%)			
	In-samples	Đánh giá	Out-of-sample	Đánh giá
ARIMA(2,1,1)	7.21	Hoàn hảo	5.24	Hoàn hảo
ARIMA(1,1,1)	19.33	Tốt	11.07	Tốt

(Nguồn: Tính toán của tác giả)

3.4. Dự báo

Kết quả dự báo CPI từ mô hình ARIMA(2,1,1) như sau:

Bảng 7. Kết quả dự báo CPI bằng mô hình ARIMA(2,1,1)

Tháng	Tháng 4/2016	Tháng 5/2016	Tháng 6/2016
CPI dự báo (%)	159.5409	159.5447	159.5234

(Nguồn: Tính toán của tác giả)

Kết quả dự báo ở Bảng 7 cho thấy CPI tháng 6/2016 đạt 159.5234%, có nghĩa là tăng 59.5234% so với năm làm gốc 2009. Như vậy, ứng dụng kết quả dự báo giúp các nhà hoạch định chính sách có kế hoạch đối phó với tình hình tăng giá hàng tiêu dùng trong tương lai, các nhà quản lý doanh nghiệp có thể sử dụng kết quả dự báo để lập kế hoạch kinh doanh, kế hoạch dự trữ hàng tồn kho để phục vụ tốt nhất nhu cầu của người tiêu dùng, đối với nhà đầu tư khi xác định được CPI có thể tính toán tỷ lệ sinh lời của dự án đầu tư một cách hợp lý,...

4. KẾT LUẬN

Nghiên cứu khả năng ứng dụng của mô hình ARIMA vào việc dự báo CPI và mục đích tìm ra mô hình tốt nhất cho việc dự báo CPI tại Việt Nam. Kết quả nghiên cứu cho thấy, mô hình ARIMA(2,1,1) cho kết quả dự báo CPI tốt và việc ứng dụng mô hình ARIMA đã được thực hiện ở một số nghiên cứu trên thế giới.

Tuy nhiên, nghiên cứu này chưa so sánh kết quả dự báo với một số mô hình khác như: Mô hình dự báo Xám, mô hình chuỗi thời gian mờ.... Những hạn chế này sẽ được quan tâm khắc phục trong các nghiên cứu tiếp theo.

TÀI LIỆU THAM KHẢO

- [1] D. Nahm (2015), “The Effects of New Goods and Substitution on the Korean CPI as a Measure of Cost of Living”, *International Economic Journal*, vol. 29, no. 1, pp. 57-72.
- [2] M. Álvarez-Díaz and R. Gupta (2016), “Forecasting US consumer price index: does nonlinearity matter?”, *Applied Economics*, pp. 1-14.
- [3] H. A. H. Al-Tamimi, A. A. Alwan and A. A. Rahman (2011), “Factors Affecting Stock Prices in the UAE Consumer Price Index”, *Journal of Transnational Management*, vol. 6, no. 1, pp. 3-19.
- [4] H. Z, B. Y and X. T (2013), “CPI forecasting using support vector regression with memetic algorithms”, *Sci World J Article*, p. 10.
- [5] M. Bernardi and L. Petrella (2015), “Multiple seasonal cycles forecasting model: the Italian CPI”, *Journal of Transnational Management*, pp. 671-695.
- [6] R. B. MILLER and J. C. HICKMAN, “Time series analysis and forecasting”, *Transaction of society of actuaries*, vol. 25, no. 1, pp. 267-329, 1973.
- [7] P. J. Brockwell and R. A. Davis (2001), *Introduction to Time Series and Forecasting*, 2nd ed., New York: Springer Link, pp. 180-196.
- [8] Tổng Cục Thống kê Việt Nam (2016), *Trung tâm Tư liệu và Dịch vụ Thống kê*, [Online]. Available: <http://www.gso.gov.vn/default.aspx?tabid=720>. [Accessed 04 10 2016].
- [9] G. Box and Jenkin (1970), *Time Series Analysis, Forecasting and Control*, 4 ed., San Francisco: Holden-Day, 1970, pp. 234-239.
- [10] D. N. Gujarati and D. C. Porter (2009), *Basic Econometrics*, 5 ed., vol. 5, Canada: Mc GrawHill, pp. 777-784.
- [11] R. Hill, W. E. Griffiths and G. C. Lim (2011), *Principles of Econometrics*, 4 ed., New Jersey: John Wiley & Sons, Inc., pp. 512-517.
- [12] D. Dickey and W. Fuller (1979), “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”, *Journal of the American Statistical Association*, vol. 74, pp. 427-431.
- [13] C. Lewis (1983), “Industrial and business forecasting methods”, *Journal of Forecasting*, vol. 2, pp. 194-196.
- [14] Spider Financial (2016), *Spider Financial Corp*, [Online]. Available: <http://www.spiderfinancial.com/support/documentation/numxl/reference-manual/descriptive-stats/mape>. [Accessed 04 10 2016].

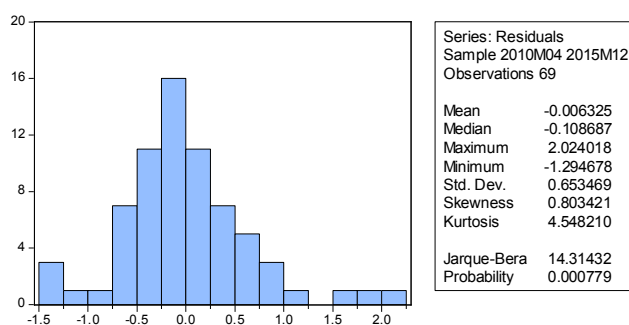
APPLYING ARIMA MODEL TO PREDICT COMSUMER PRICE INDEX IN VIETNAM

Abstract. This paper used ARIMA model to predict Consumer Price Index (CPI) of Vietnam in the second quarter of 2016. Secondary data for research was collected from General Statistics Office of Viet Nam from January 2010 to March 2016 and was estimated via Eviews 6. The result showed that ARIMA(2,1,1) was suitable approach to forecast CPI. The fitted CPI in the second quarter of 2016 was 159.5409%, 159.5447%, 159.5234%. This ARIMA model can be used by policy makers and entrepreneurs for the prediction, planning, and business planning.

Keywords: Forecast, CPI, ARIMA, variance

PHỤ LỤC:

Phụ lục 1: Kiểm định phân phối chuẩn của Sai số ngẫu nhiên



Phụ lục 2: Kiểm định tự tương quan

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.307456	Prob. F(2,63)	0.7364
Obs*R-squared	0.660467	Prob. Chi-Square(2)	0.7188

Phụ lục 3: Kiểm định hiện tượng sai số ngẫu nhiên thay đổi:

Heteroskedasticity Test: White

F-statistic	1.137173	Prob. F(14,54)	0.3491
Obs*R-squared	15.71084	Prob. Chi-Square(14)	0.3313
Scaled explained SS	24.51429	Prob. Chi-Square(14)	0.0397