

# DỰ BÁO CHỈ SỐ GIÁ TIÊU DÙNG TẠI VIỆT NAM BẰNG MÔ HÌNH ARIMA

Nguyễn Thị Thanh Huyền\* - Nguyễn Tuấn Sơn\*\*

*Nghiên cứu này ứng dụng mô hình ARIMA trong dự báo chỉ số giá tiêu dùng CPI ở Việt Nam từ tháng 6 năm 2021 đến tháng 3 năm 2022, với số liệu được dùng để ước lượng từ tháng 1 năm 2009 đến tháng 5 năm 2021. Số liệu được thu thập từ finance.vietstock.vn. Kết quả cho thấy mô hình ARIMA(7,0,5) là thích hợp cho việc dự báo. Kết quả dự báo CPI trong 10 tháng cuối nói trên đã phản ánh được xu hướng biến đổi của chỉ số giá tiêu dùng thực tế.*

• Từ khóa: Chỉ số giá tiêu dùng, ARIMA.

*This study applies the ARIMA model in forecasting the consumer price index CPI in Vietnam from June 2021 to March 2022, with the data used to estimate from January 2009 to May 2021. The data is collected from finance.vietstock.vn. The results show that the ARIMA(7,0,5) model is suitable for forecasting. The CPI forecast results for the last 10 months mentioned above have reflected the trend of the actual consumer price index.*

• Keywords: Consumer price index, ARIMA.

Ngày nhận bài: 10/5/2022

Ngày gửi phản biện: 18/5/2022

Ngày nhận kết quả phản biện: 18/6/2022

Ngày chấp nhận đăng: 20/6/2022

Chỉ số CPI có mối liên hệ đặc biệt mật thiết với tỉ lệ lạm phát, được coi như là một thước đo điển hình của tỉ lệ lạm phát hay còn được gọi là “thuế” lạm phát. CPI được coi như một chỉ số kinh tế, việc tính toán và phân tích chỉ số CPI, đem lại nhiều hiệu quả tích cực cho nền kinh tế, chính phủ, doanh nghiệp, từ đó người dân sẽ có sự chuẩn bị trước khi thay đổi giá cả nền kinh tế. Tính toán CPI là một trong những yếu tố quan trọng, chỉ số CPI còn được ứng dụng làm thước đo lạm phát cho các yếu tố kinh tế khác. Các yếu tố kinh tế này có thể là doanh số bán lẻ, thu nhập hàng giờ, giá trị đồng tiền...

Việc tăng chỉ số giá tiêu dùng đồng nghĩa với việc giá cả của các loại mặt hàng đều tăng, điều này tác động mạnh mẽ đến đời sống của người tiêu dùng đặc biệt là với những người có thu nhập thấp, cuộc sống của họ sẽ trở nên khó khăn và vất vả hơn. Ở thành thị, những người có thu nhập thấp chủ yếu có thu nhập từ tiền lương, tiền công, trợ cấp xã hội, mức tiền lương không thay đổi, nhưng giá thành sản phẩm nhu cầu tiêu dùng lại tăng cao, do đó chi phí chi tiêu cũng tăng, cuộc sống dần trở nên khó khăn và vất vả hơn.

Dự báo chỉ số giá tiêu dùng (CPI) là hoạt động quan trọng đối với không chỉ chính phủ mà cả các doanh nghiệp trong việc lập kế hoạch hoạt động. Kết quả dự báo càng chính xác thì việc lập kế hoạch càng khả thi. Hoạt động dự báo này cũng được tiến hành nhiều trong các nghiên cứu trước. Với các nghiên cứu ngoài nước, Moser và cộng sự (2007) áp dụng các mô hình nhân tố do Stock và Watson đề xuất, cũng như các mô hình VAR và ARIMA để tạo ra các dự báo không lấy mẫu trong 12 tháng về lạm phát HICP của Áo và các chỉ số phụ của nó.

## 1. Giới thiệu

CPI là một chỉ số được sử dụng khi phân tích và đánh giá khả năng tiêu dùng của nền kinh tế. Chỉ số giá tiêu dùng (tiếng Anh: Consumer Price Index - CPI) là chỉ số phản ánh mức thay đổi tương đối của giá hàng tiêu dùng theo thời gian được tính bằng phần trăm (%). CPI đo lường sự thay đổi trung bình về giá theo thời gian mà người tiêu dùng trả cho một giỏ hàng hóa hoặc dịch vụ, thường được gọi là lạm phát.

Cách tính chỉ số CPI gồm 4 bước:

Bước 1: Cố định giỏ hàng hóa: Thông qua báo cáo điều tra sẽ xác định lượng hàng hoá, dịch vụ tiêu biểu mà một người tiêu dùng điển hình mua.

Bước 2: Xác định giá cả: Thống kê giá cả của mỗi mặt hàng trong giỏ hàng hoá tại từng thời điểm.

Bước 3: Tính chi phí vật dụng trong giỏ hàng hoá bằng cách dùng số lượng nhân với giá cả của từng loại sản phẩm rồi cộng lại.

Bước 4: Lựa chọn thời kỳ gốc để làm cơ sở so sánh rồi tính chỉ số giá tiêu dùng bằng công thức sau:

$$CPI_t = \frac{\text{Chi phí để mua giỏ hàng hoá thời kỳ } t}{\text{Chi phí để mua giỏ hàng hoá kỳ cơ sở}} \times 100$$

Thời kỳ gốc sẽ được thay đổi trong vòng 5 đến 7 năm tùy ở từng nước.

\* Công ty Cổ phần Dịch vụ Phát triển Nam Phong - \*\* Cục Thuế Bắc Ninh

Các tác giả áp dụng quy trình lựa chọn mô hình dự báo tuần tự, phù hợp với nhiệm vụ cụ thể này. Kết quả chỉ ra rằng các mô hình nhân tố có độ chính xác dự đoán cao nhất đối với một số chỉ số phụ và độ chính xác dự đoán có thể được cải thiện hơn nữa bằng cách kết hợp thông tin có trong mô hình nhân tố và VAR cho một số chỉ số. Đối với dự báo lạm phát HICP, phân tích của các tác giả đề xuất ưu tiên việc tổng hợp các dự báo của chỉ số phụ hơn là dự báo về lạm phát tiêu đề. Một nghiên cứu của Hassani & Silva (2018) khẳng định tỷ lệ lạm phát là một chỉ số kinh tế quan trọng mà các nhà dự báo không ngừng tìm cách cải thiện độ chính xác của các dự đoán, để cho phép ra quyết định kinh tế vĩ mô tốt hơn. Gần đây, Ergemen (2022) phân tích tỷ lệ lạm phát ở các quốc gia OECD có thu nhập cao bằng cách sử dụng cấu trúc nhân tố đa cấp được ước tính dựa trên phân tích tương quan chính tắc (CCA) và bình phương nhỏ nhất tuần tự (SLS). Các tác giả cho thấy rằng lạm phát có thành phần toàn cầu, chủ yếu do các nước G7, giải thích cho 77% sự thay đổi trung bình và thành phần địa phương chiếm tỷ trọng đáng kể trong một nhóm nhỏ các quốc gia. Các tác giả chỉ ra rằng sự kết hợp giữa các thành phần toàn cầu và địa phương này có khả năng dự báo vượt trội và có thể cải thiện đáng kể hiệu suất dự báo so với thông số kỹ thuật chỉ dành cho thành phần toàn cầu cho các phạm vi chính sách khác nhau, do đó tạo thành một chuẩn mực mới cho dự báo lạm phát. Tuy nhiên, các nghiên cứu trong nước về dự báo tỷ lệ lạm phát sử dụng các phương pháp khác nhau còn chưa được công bố chính thức. Bởi vậy nghiên cứu này sử dụng mô hình ARIMA trong dự báo chỉ số giá tiêu dùng, như đã được sử dụng trong Junttila (2001). Phần còn lại của bài viết được kết cấu như sau: Mục 2 trình bày dữ liệu và phương pháp nghiên cứu, mục 3 trình bày kết quả nghiên cứu, cuối cùng là kết luận trong mục 4.

## 2. Dữ liệu nghiên cứu và phương pháp nghiên cứu

### 2.1. Dữ liệu nghiên cứu

Dữ liệu nghiên cứu là chỉ số giá tiêu dùng tại Việt Nam được cung cấp tại <https://finance.vietstock.vn/>.

### 2.2. Phương pháp nghiên cứu

Mô hình ARMA của Box-Jenkins là sự kết hợp giữa các mô hình AR (Tự hồi quy) và MA (Trung bình trượt) như sau:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} - \alpha_1 u_{t-1} - \alpha_2 u_{t-2} - \dots - \alpha_q u_{t-q} + u_t \quad (1)$$

Phương pháp của Box-Jenkins bao gồm các thủ tục sau:

- Thiết lập sự ổn định của chuỗi thời gian: Hàm tự tương quan (ACF) cũng như các kiểm định Augmented Dickey-Fuller (ADF) (1979) và Phillips-Perron (1988) được sử dụng để kiểm tra tính dừng của các chuỗi thời gian.

- Nhận dạng mô hình của mô hình ARMA(p, q): Để xác định thứ tự của ARMA(p, q), chúng ta sử dụng mẫu của hàm tự tương quan (ACF) và hàm tự tương quan riêng phần (PACF) của chuỗi tĩnh. Hai phần này được dùng để đề xuất mô hình chúng ta nên xây dựng. Tham số p của toán tử tự hồi quy được xác định bởi hệ số tự tương quan riêng phần và tham số q của toán tử trung bình trượt được chỉ định bởi hệ số tự tương quan. Trong thực tế, tác giả sử dụng các giới hạn  $\pm 2/\sqrt{n}$  cho các mức ý nghĩa của hai hàm, vì vậy chúng tôi sẽ có một số mô hình ARIMA(a, b), trong đó  $0 \leq a \leq p$ ,  $0 \leq b \leq q$ . Để lựa chọn mô hình tối ưu, tác giả sử dụng các tiêu chuẩn của Akaike (AIC) và Schwartz (SIC).

- Ước lượng mô hình: Việc kiểm định sự tồn tại của các nhiễu trắng trong mô hình ARIMA đòi hỏi một quá trình lặp phi tuyến trong việc ước lượng các tham số. Phương pháp ước lượng hợp lý tối đa được xem là kỹ thuật phù hợp trong tình huống này.

- Dự báo: Một trong những lý do chính của việc phân tích các mô hình chuỗi thời gian là dự báo. Độ chính xác của các dự báo phụ thuộc vào sai số dự báo. Sau đó, giá trị dự báo sau một giai đoạn có điều kiện trên tất cả các thông tin theo thời gian t, được đưa ra tại thời điểm t + k, như sau:

$$y_{t+k} = (\beta_1 + 2)y_{t+k-1} - (1 + 2\beta_1)y_{t+k-2} + \beta_1 y_{t+k-3} + \alpha_1 \varepsilon_{t+k-1} + \varepsilon_t \quad (2)$$

## 3. Kết quả thực nghiệm

Bảng 1 sau đây trình bày kết quả mô tả thống kê của chỉ số giá tiêu dùng.

**Bảng 1. Thống kê mô tả chỉ số giá tiêu dùng**

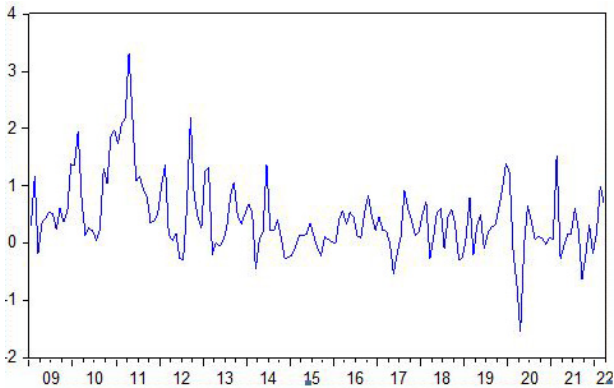
	CPI
Trung bình	0.437358
Trung vị	0.320000
Giá trị lớn nhất	3.320000
Giá trị nhỏ nhất	-1.540000
Độ lệch chuẩn	0.637709
Hệ số bất đối xứng	1.176441
Hệ số nhọn	6.012003
Jarque-Bera	96.77947
Giá trị xác suất	0.000000
Số quan sát	159

(Nguồn: Tác giả tổng hợp)

Bảng thống kê thể hiện các giá trị thống kê cơ bản của từng biến như giá trị trung bình, giá trị lớn nhất, giá

trị nhỏ nhất, độ lệch chuẩn. Đồ thị từng biến sau đây thể hiện trực quan sự vận động của chỉ số giá tiêu dùng trong giai đoạn nghiên cứu.

**Hình 1. Đồ thị mô tả biến động của chỉ số giá tiêu dùng CPI**



(Nguồn: Tác giả vẽ từ số liệu thu thập)

Ngoài việc cung cấp hình ảnh trực quan về sự biến động của từng biến, hình 1 còn giúp chúng ta dự báo tính dừng của các chuỗi dữ liệu. Có thể đưa ra dự đoán rằng, chỉ số giá tiêu dùng là chuỗi dừng. Các kết quả kiểm định tính dừng của chuỗi thời gian được trình bày trong bảng 2 sau đây:

**Bảng 2. Kết quả kiểm định tính dừng của chỉ số giá tiêu dùng**

Null Hypothesis: CPI has a unit root  
Exogenous: Constant, Linear Trend  
Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.745012	0.0000
Test critical values:		
1% level	-4.017185	
5% level	-3.438515	
10% level	-3.143558	

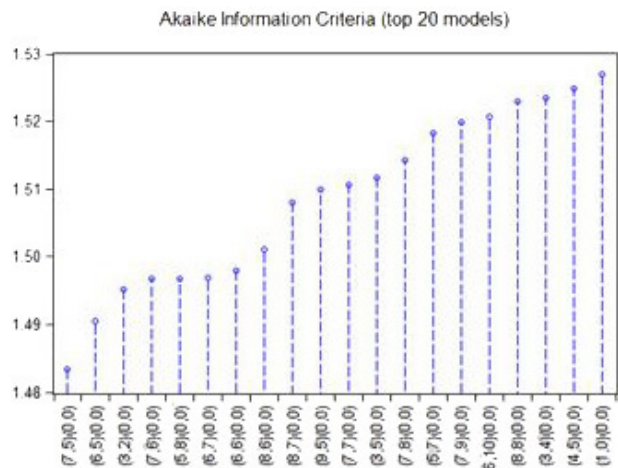
(Nguồn: Tác giả)

Kết quả kiểm định trong Bảng 2 khẳng định chuỗi CPI là chuỗi dừng. Do đó, đối với mô hình ARIMA(p, d, q) trong bài báo này, chúng ta sẽ có giá trị  $d = 0$ .

Để xác định mô hình ARMA (p, q), tức là các giá trị của tham số p và q. Các quan sát từ tháng 1 năm 2009 đến tháng 5 năm 2021, gồm 149 quan sát được sử dụng để ước lượng mô hình, còn các quan sát từ tháng 6 năm 2021 đến tháng 3 năm 2022 được sử dụng để kiểm định mô hình. Như đã được đề cập ở trên, mô hình AR(p) có PACF dừng ở độ trễ p và MA (q) có ACF dừng ở độ trễ q. Trong thực hành,  $\pm 2/\sqrt{n}$  là các giới hạn của hai hàm. Chúng ta sẽ tìm trong các mô hình ARMA (a, b), với  $0 \leq a \leq p, 0 \leq b \leq q$  để tìm một mô hình tối ưu. Để làm điều này, chúng ta sẽ sử dụng các tiêu chuẩn AIC và SIC được tính toán một cách tự động cho mỗi mô hình ước lượng được. Giới hạn cho cả hai hàm (ACF, PACF) là  $\pm 2/\sqrt{149} \approx 0,1638$ .

Để lựa chọn mô hình ARMA tốt nhất, cần ước lượng các mô hình khác nhau và căn cứ vào số lượng các biến có ý nghĩa thống kê trong mô hình, hệ số R<sup>2</sup> hiệu chỉnh, tiêu chuẩn Akaike và tiêu chuẩn Schwarz nhỏ nhất. Ở đây, sau khi đã biết chuỗi chỉ số giá là chuỗi dừng, ta sử dụng trực tiếp chuỗi chỉ số giá trong ước lượng các mô hình. Còn các bậc của các mô hình AR và MA ta thực hiện ước lượng đến tối đa bậc 12. Hình 2 sau đây minh họa 20 mô hình tốt nhất trong quá trình ước lượng.

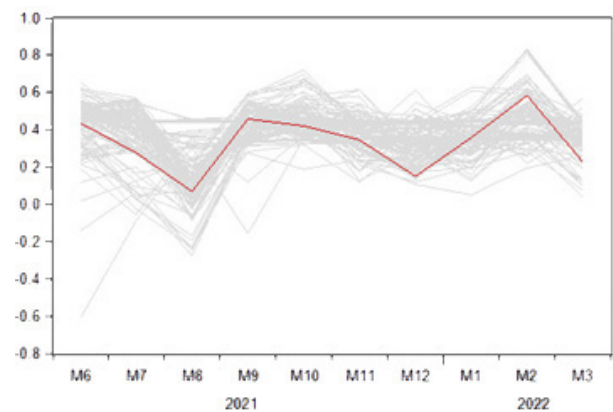
**Hình 2. Minh họa tiêu chuẩn Akaike để lựa chọn dạng hàm tối ưu cho mô hình ARMA**



(Nguồn: Tác giả)

Hoặc hình 3 sau đây, so sánh kết quả ước lượng của các mô hình khác nhau.

**Hình 3. Minh họa tiêu chuẩn Akaike để lựa chọn dạng hàm tối ưu cho mô hình ARMA**



(Nguồn: Tác giả)

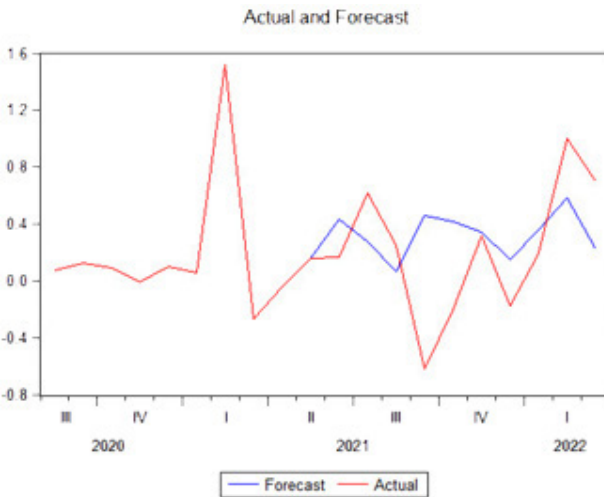
Trong hình 3, ước lượng của mỗi mô hình được vẽ đồ thị để so sánh với các mô hình khác và mô hình được chọn là mô hình đường đậm. Kết quả cuối cùng, mô hình tốt nhất là mô hình ARIMA(7,0,5). Kết quả ước lượng mô hình như trong bảng 3 sau đây.

**Bảng 3. Kết quả ước lượng mô hình ARIMA(7,0,5)**

Các biến	Hệ số hồi quy	Sai số chuẩn	Thống kê t	Giá trị xác suất
C	0.441605	0.187196	2.359047	0.0198
AR(1)	-0.118542	0.161218	-0.735290	0.4634
AR(2)	0.036760	0.212219	0.173217	0.8627
AR(3)	0.256962	0.162001	1.586171	0.1150
AR(4)	0.498826	0.073231	6.811651	0.0000
AR(5)	0.534849	0.076402	7.000455	0.0000
AR(6)	-0.519733	0.149614	-3.473831	0.0007
AR(7)	0.131620	0.111135	1.184322	0.2384
MA(1)	0.845616	17.49655	0.048330	0.9615
MA(2)	0.505917	11.48971	0.044032	0.9649
MA(3)	-0.079017	27.13002	-0.002913	0.9977
MA(4)	-0.536641	20.34210	-0.026381	0.9790
MA(5)	-0.886863	51.37520	-0.017262	0.9863
SIGMASQ	0.198335	5.412158	0.036646	0.9708

(Nguồn: Tác giả)

Cuối cùng, mô hình ARIMA(7,0,5) được sử dụng để dự báo cho 10 tháng tiếp theo từ tháng 6 năm 2021 đến tháng 3 năm 2022. Minh họa ước lượng như trong hình 4 sau đây.

**Hình 4. Minh họa kết quả dự báo nhờ mô hình ARIMA(7,0,5)**

(Nguồn: Tác giả)

Có thể thấy kết quả dự báo có xu thế trùng với xu thế thực tế trong 10 tháng cuối, nhưng có độ trễ nhất định.

#### 4. Kết luận

Trong bài báo này, sử dụng kỹ thuật Box - Jenkins, tác giả đã thực hiện dự báo chỉ số giá tiêu dùng của Việt Nam nhờ mô hình ARIMA. Sau khi kiểm tra tính dừng của chuỗi dữ liệu chỉ số giá tiêu dùng, tác giả tìm xem

mô hình ARIMA(p, d, q) nào là phù hợp. Các tiêu chuẩn về số hệ số ước lượng có ý nghĩa thống kê,  $R^2$  hiệu chỉnh, tiêu chuẩn Akaike, tiêu chuẩn Schwarz được sử dụng để lựa chọn mô hình tốt nhất. Mô hình ARIMA(7,0,5) giúp dự báo 10 tháng từ tháng 6 năm 2021 đến tháng 3 năm 2022, so với chỉ số thực tiễn như trong bảng 4.

Các kết quả dự báo cũng thể hiện được xu thế biến động của chỉ số giá tiêu dùng, nên mô hình phương pháp này có thể sử dụng trong dự báo không chỉ chỉ số giá tiêu dùng mà còn có thể áp dụng cho các chỉ số kinh tế vĩ mô khác nữa.

**Bảng 4. Kết quả dự báo chỉ số giá tiêu dùng nhờ mô hình ARIMA(7,0,5)**

Thời gian	Chỉ số dự báo	Chỉ số thực tế
Tháng 6/2021	0.43	0.17
Tháng 7/2021	0.27	0.62
Tháng 8/2021	0.06	0.25
Tháng 9/2021	0.45	-0.62
Tháng 10/2021	0.41	-0.2
Tháng 11/2021	0.34	0.32
Tháng 12/2021	0.14	-0.18
Tháng 1/2022	0.36	0.19
Tháng 2/2022	0.58	1
Tháng 3/2022	0.22	0.7

(Nguồn: Tác giả tổng hợp)

#### Tài liệu tham khảo:

Dickey, D. & Fuller, W. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Ergemen, Y. E. (2022). "Forecasting inflation rates with multi-level international dependence". *Economics Letters*, Vol. 214, May 2022, 110456.

Hassani, H. & Silva, E. S. (2018). "Forecasting UK consumer price inflation using inflation forecasts". *Research in Economics*, Vol. 72, Issue 3, September 2018, Pages 367-378.

Junttila, J. (2001). "Structural breaks, ARIMA model and Finnish inflation forecasts". *International Journal of Forecasting*, Vol. 17, Issue 2, April-June 2001, Pages 203-230.

Moser, G.; Rumler, F. & Scharler, J. (2007). "Forecasting Austrian inflation". *Economic Modelling*, Vol. 24, Issue 3, May 2007, Pages 470-480.

Phillips, P. C. B. & Perron, P. (1988). "Testing for a unit root in time series regression", *Biometrika*, 75(2), 335-346.