

SỬ DỤNG MÔ HÌNH ARIMA ĐỂ DỰ BÁO GIÁ BÁN THAN PHỤC VỤ CÔNG TÁC LẬP KẾ HOẠCH KHAI THÁC CHO CÁC MỎ THAN LỘ THIÊN VIỆT NAM

NGUYỄN HOÀNG, BÙI XUÂN NAM,
NGUYỄN NGỌC KHÁNH - *Trường Đại học Mở-Địa chất*
Email: nguyenhoang.mdc@gmail.com

1. Giới thiệu

ARIMA là một mô hình dự báo do Box và Jenkins phát minh năm 1976 [2] được sử dụng để phân tích và dự báo các chuỗi thời gian. Ở nước ta, mô hình ARIMA đã được ứng dụng trong một số lĩnh vực như dự báo chỉ số VN-Index [3], ứng dụng mô hình ARIMA-ARCH/GARCH dự báo giá vàng ngắn hạn" với mục tiêu phân tích tình hình biến động của giá vàng Việt Nam thông qua phân tích giá vàng SJC trong thời gian 10 năm [5], dự báo chỉ số giá tiêu dùng [7], dự báo khách du lịch vào Việt Nam [8], nghiên cứu cảnh báo hạn hán [11],...

Trên thế giới, mô hình ARIMA cũng được ứng dụng rộng rãi để dự báo sự biến động của các đối tượng có liên quan tới yếu tố thời gian. Nhu cầu sử dụng điện năng cho một thành phố cũng được dự báo dựa trên mô hình ARIMA [13], ở một yêu cầu cao hơn về độ chính xác, nhóm nghiên cứu này đã nghiên cứu và cải thiện mô hình ARIMA với độ chính xác cao hơn mô hình ARIMA theo mùa duy nhất. Ngoài ra, mô hình ARIMA còn được sử dụng để dự báo sản lượng dầu thô, than đá và khí ga phục vụ cho nhu cầu nhiên liệu toàn cầu [9]. Bằng cách áp dụng mô hình ARIMA, dự báo về tổng lượng tiêu thụ năng lượng, than đá và lượng tiêu thụ phi hoá thạch ở Trung Quốc trước năm 2020 cũng đã được nghiên cứu [14],... và còn nhiều các nghiên cứu khác sử dụng mô hình ARIMA hoặc mô hình lai giữa ARIMA và một số mô hình khác cho các mục đích tương tự.

Tuy nhiên, việc dự báo sự biến động của nhiên liệu nói chung và giá than trong nước nói riêng lại là một vấn đề khá mới mẻ và chưa được quan tâm nhiều. Có rất nhiều phương pháp dự báo có thể sử dụng để dự báo sự biến động của giá bán than, tuy

nhiên mô hình ARIMA là một trong những phương pháp dễ tiếp cận và cho kết quả khá chính xác trong dự báo ngắn hạn về sự biến đổi giá than trong tương lai gần. Bài báo sử dụng mô hình ARIMA để dự báo sự biến động của giá bán than nhiệt điện từ 05/2017 đến 12/2017 dựa trên cơ sở dữ liệu giá bán than nhiệt điện từ 01/2007 đến 04/2017.

2. Tổng quan về mô hình ARIMA

Mô hình ARIMA (Autoregressive Integrated Moving Average) là tập hợp một số mô hình phân tích và tiên lượng theo dữ liệu thời gian, là phương pháp dựa vào những giả định về mối tương quan giữa các giá trị liên tiếp nhau trong dãy số liệu (hay còn gọi là tương quan liên hoàn - autocorrelation). ARIMA là mô hình thống kê dùng để mô hình hóa ngẫu nhiên của một dãy số liệu theo thời gian.

Mô hình ARIMA (p,d,q) là sự kết hợp của các mô hình AR(p), MA(q) và sai phân I(d) [10]. Ý tưởng chính của mô hình AR(p) là hồi quy trên chính số liệu quá khứ ở những chu kỳ trước [1], [6], [12]:

$$Y(t) = a_0 + a_1 y(t-1) + a_2 y(t-2) + \dots + a_p y(t-p) + e(t) \quad (1)$$

Trong đó: $y(t)$ - Quan sát dừng hiện tại; $y(t-1)$, $y(t-2)$,... - Quan sát dừng quá khứ (thường sử dụng không quá 2 biến này); a_0 , a_1 , a_2 ,... - Các tham số phân tích hồi quy; $e(t)$ - Sai số dự báo ngẫu nhiên của giai đoạn hiện tại. Giá trị trung bình được mong đợi bằng 0.

$Y(t)$ là một hàm tuyến tính của những quan sát dừng quá khứ $y(t-1)$, $y(t-2)$,...

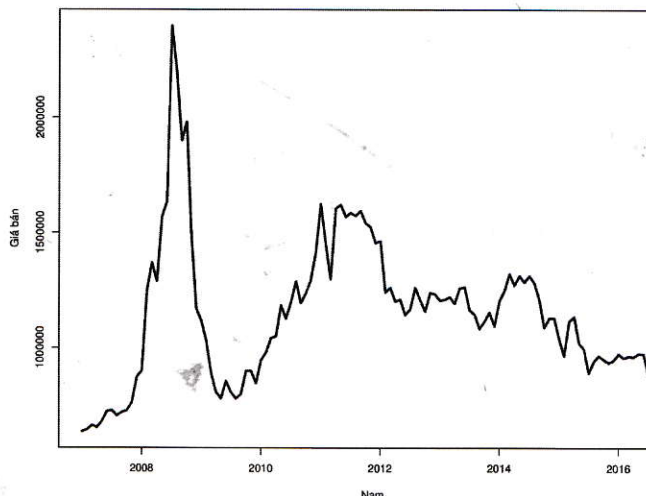
Mô hình MA(q): Quan sát dừng hiện tại $y(t)$ là một hàm tuyến tính phụ thuộc các biến sai số dự báo quá khứ và hiện tại. Mô hình bình quân di động là một trung bình trọng số của những sai số mới nhất.

$$Y(t) = b_0 + e(t) + b_1e(t-1) + b_2e(t-2) + \dots + b_qe(t-q) \quad (2)$$

Trong đó: $y(t)$ - Quan sát dừng hiện tại; $e(t)$ - Sai số dự báo ngẫu nhiên, giá trị của nó không được biết và giá trị trung bình của nó là 0; $e(t-1), e(t-2), \dots$ - Sai số dự báo quá khứ (thông thường mô hình sẽ sử dụng không quá 2 biến này); b_0, b_1, b_2, \dots - Giá trị trung bình của $y(t)$ và các hệ số bình quân di động; q - Sai số quá khứ được dùng trong mô hình bình quân di động, nếu ta sử dụng hai sai số quá khứ thì sẽ có mô hình bình quân di động bậc 2 là MA(2).

Sai phân I(d) thể hiện chuỗi dừng của số liệu. Chuỗi thời gian được coi là dừng nếu như trung bình và phương sai của nó không đổi theo thời gian và giá trị của đồng phương sai giữa hai thời đoạn chỉ phụ thuộc vào khoảng cách và độ trễ về thời gian giữa hai thời đoạn này chứ không phụ thuộc vào thời điểm thực tế mà đồng phương sai được tính.

Sai phân chỉ sự khác nhau giữa giá trị hiện tại và giá trị trước đó. Phân tích sai phân nhằm làm cho ổn định giá trị trung bình của chuỗi dữ liệu, giúp cho việc chuyển đổi chuỗi thành một chuỗi dừng.



H.2. Giá bán than nhiệt điện từ 1/2007 đến 4/2017

Phương pháp nghiên cứu của bài báo dựa trên mô hình ARIMA với các bước thực hiện như sau [4]:

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

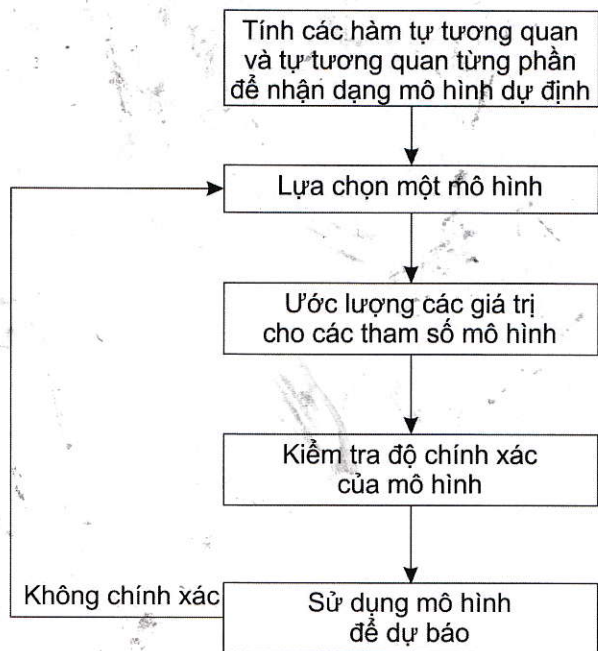
> Bước 2: ước lượng các thông số của mô hình ARIMA (p, d, q). Các tham số của mô hình ARIMA sẽ được ước lượng theo phương pháp bình phương nhỏ nhất;

> 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;

> 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.

4. Kết quả nghiên cứu và thảo luận

Phương pháp Box-Jenkins trong lý thuyết chuỗi thời gian được thông qua để thiết lập một mô hình ARIMA trong nghiên cứu này. Các bước tính toán

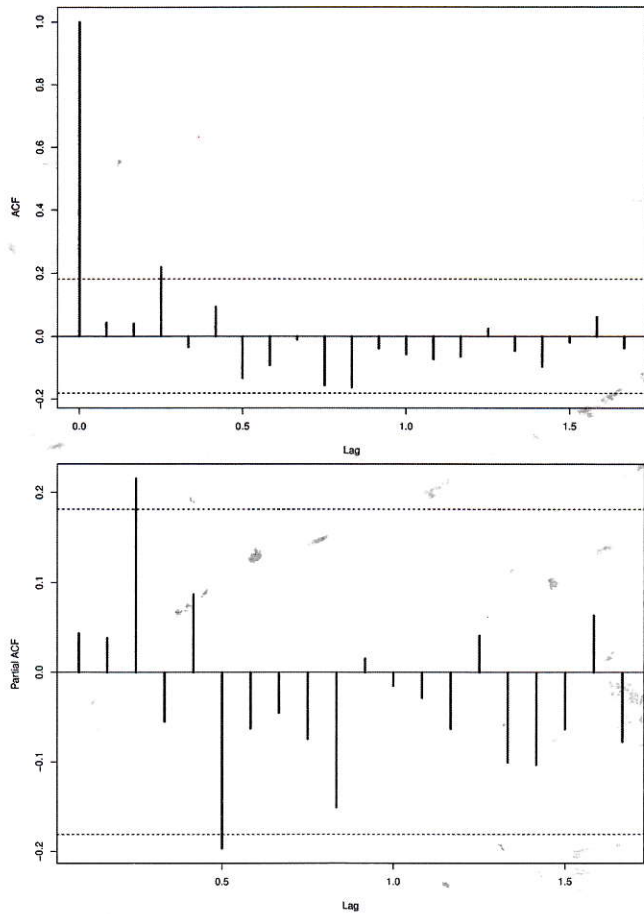


H1. Mô hình Box-Jenkins

3. Số liệu sử dụng và phương pháp nghiên cứu

Số liệu sử dụng cho bài báo là giá bán than nhiệt điện được tổng hợp trong thời gian từ 01/2007 đến 04/2017 (H.2). Số liệu được tổng hợp theo từng tháng, được xử lý bằng ngôn ngữ lập trình R với 124 số liệu được chia thành 2 bộ. Bộ số liệu chạy mô hình bao gồm 118 số liệu, bộ số liệu kiểm tra bao gồm 6 số liệu.

của phương pháp này được thực hiện theo 4 bước như trên [4]. Trong phương pháp Box-Jenkins, một cách tiếp cận khác biệt sẽ được sử dụng để ổn định dữ liệu ban đầu, chức năng tự tương quan (ACF) và chức năng tự tương quan từng phần (PACF) sẽ được sử dụng để quyết định thành phần trung gian tự động hoặc chuyển động (nếu có) trong mô hình ARIMA. Các giá trị ACF và PACF cho lần thứ 1÷118 được thể hiện trong hình H.3.

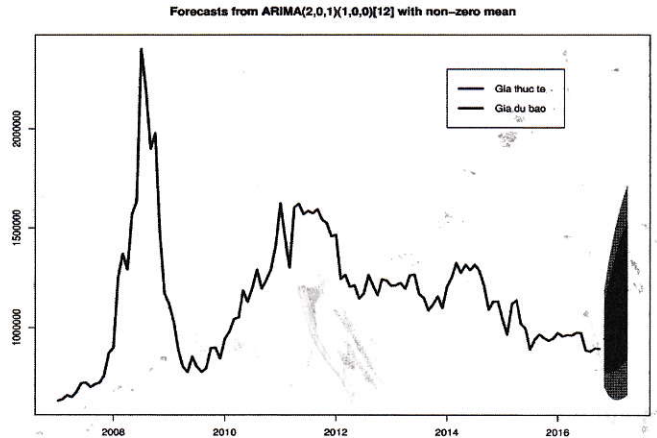


H.3. Biểu đồ tự tương quan ACF và tự tương quan từng phần PACF

Từ hình H.3 có thể thấy rằng từ giá trị 1 đến 118 của chuỗi gốc sau khi tính toán sai phân bậc 1 đều đáp ứng mô hình AR bậc cao hoặc mô hình ARMA vì cả ACF lẫn PACF đều có hiện tượng giảm dần và dao động. Để kiểm tra độ chính xác của các mô hình, sai số chuẩn (standard error - s.e.) được sử dụng để đánh giá các mô hình. Quá trình tính toán và chạy các mô hình đã đưa ra mô hình ARIMA tối ưu nhất cho giá bán than nhiệt điện là mô hình ARIMA(2,0,1)(1,0,0) [12] với các kết quả như sau:

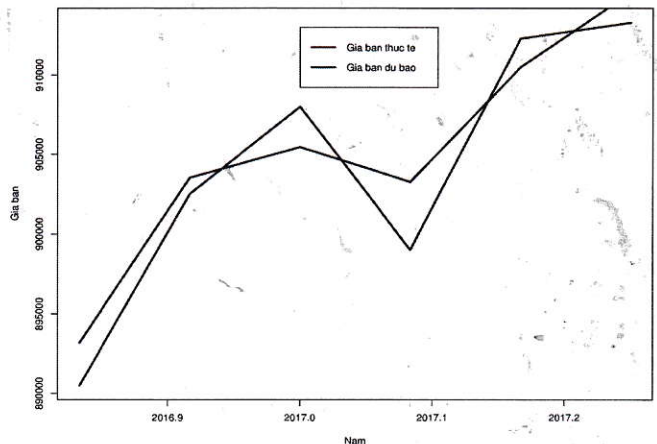
ar1	ar2	ma1	sar1
1.8640	-0.8968	-0.8691	0.0425
s.e.	0.0601	0.0558	0.0794
	0.0943		

Kết quả chạy mô hình cho ra mô hình ARIMA (2,0,1)[12] hoặc mô hình ARIMA(1,0,0) [12] có giá trị tương tự nhau và có tính mùa vụ. Qua khảo sát cho thấy giá than nhiệt điện thường tăng lên vào mùa hè và giảm vào mùa đông do nhu cầu tiêu thụ điện mùa hè tăng cao hơn so với mùa đông.



H.4. Đồ thị dự báo giá bán than nhiệt điện theo mô hình ARIMA đã chọn

H.4 minh họa kết quả dự báo giá bán than nhiệt điện khi sử dụng mô hình ARIMA(2,0,1)(1,0,0) [12] và được thể hiện bằng đường màu xanh nước biển trong khoảng tin cậy 95 % (vùng màu tím) và khoảng tin cậy 80 % (vùng màu xám). Để đánh giá mức độ chính xác của mô hình ARIMA đã chọn, nhóm tác giả sử dụng bộ số liệu kiểm tra bao gồm 6 số liệu (từ 11/2016 đến 04/2017) để so sánh với kết quả dự báo từ mô hình ARIMA đã chọn. H.5 minh họa sự sai khác giữa giá than thực tế so với giá than dự báo, các kết quả dự báo được minh họa trong hình H.5 cho thấy mô hình ARIMA(2,0,1)(1,0,0) [12] cho kết quả dự báo khá chính xác, sai số so với giá bán than thực tế là chấp nhận được.



H.5. Đồ thị so sánh sự khác nhau giữa giá bán dự báo với giá bán thực tế

5. Kết luận

Phương pháp sử dụng mô hình ARIMA là một phương pháp có khả năng ứng dụng rộng rãi trong các lĩnh vực bao gồm cả lĩnh vực khai khoáng, đặc biệt là dự báo sự biến động của giá bán khoáng sản.

Nghiên cứu này cho phép người sử dụng có thể dự báo ngắn hạn, trung hạn hoặc dài hạn sự biến động về giá bán than trong tương lai để từ đó có kế hoạch khai thác phù hợp cho các mỏ than Việt Nam nói chung và các mỏ than lộ thiên Việt Nam nói riêng. Nghiên cứu này đồng thời cũng là tài liệu tham khảo cho các mỏ than lộ thiên Việt Nam để sử dụng các mô hình dự báo sự biến động về giá than phục vụ cho công tác lập kế hoạch khai thác và chiến lược phát triển sản xuất, kinh doanh có hiệu quả của mỏ trong tương lai. □

TÀI LIỆU THAM KHẢO

1. Mohammed S Ahmed và Allen R Cook (1979), Analysis of freeway traffic time-series data by using Box-Jenkins techniques.
2. George Box (2012), "Box and Jenkins: Time series analysis forecasting and control", A Very British Affair: Six Britons and the Development of Time Series Analysis during the 20th Century, tr.161.
3. Đặng Nữ Hà My (2015), Ứng dụng mô hình ARIMA trong dự báo chỉ số VN-Index", Trường Đại học Kinh tế Huế.
4. Ergin Erdem và Jing Shi (2011), "ARMA based approaches for forecasting the tuple of wind speed and direction", Applied Energy. 88(4), tr. 1405-1414.
5. Hoàng Thị Bích Vân (2014), "Phân tích diễn biến giá vàng Việt Nam. Ứng dụng mô hình ARIMA-ARCH/GARCH dự báo giá vàng ngắn hạn".
6. Ross Ihaka (2005), "Time Series Analysis (Lecture Notes for 475.726)", Statistics Department, University of Auckland, New Zealand.
7. Huỳnh Tấn Nguyên và Nguyễn Văn Lượng (2017), "ỨNG DỤNG MÔ HÌNH ARIMA ĐỂ DỰ BÁO CHỈ SỐ GIÁ TIÊU DÙNG Ở VIỆT NAM", Tạp chí Khoa học và Công nghệ. 5(3).
8. Đỗ Quang (2012), "Xây dựng mô hình ARIMA cho dự báo khách du lịch đến Việt Nam".
9. Shahriar Shafiee và Erkan Topal (2008), "An econometrics view of worldwide fossil fuel consumption and the role of US", Energy Policy. 36(2), tr. 775-786.
10. Nguyễn Ngọc Thiệp (2010), Một số phương pháp khai phá dữ liệu quan hệ trong tài chính và chứng khoán, Luận văn Thạc sĩ, Đại học Quốc gia Hà Nội - Đại học Công nghệ.
11. Ngo Van Quan (2016), "Nghiên cứu cảnh báo hạn hán sử dụng mô hình ARIMA cho lưu vực sông Nakdong của Hàn Quốc", Tạp chí Khoa học kỹ thuật Thủy lợi và Môi trường(46), tr. 158.
12. Boris Kovalerchuk; Evgenii Vityaev (2001), Advances in Relational and Hybrid Methods, Data Mining in Finance, Kluwer Academic Publishers, Boston, Dordrecht-London.
13. Yuanyuan Wang and others. (2012), "Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China", Energy Policy. 48, tr. 284-294.
14. Li-ming Xue and others. (2011), "Chinese energy consumption structure prediction by application of ARIMA", China Mining Magazine. 20(4), tr. 24-27.

Ngày nhận bài: 24/04/2017

Ngày gửi phản biện: 5/05/2017

Ngày nhận phản biện: 15/07/2017

Ngày chấp nhận đăng bài: 15/10/2017

Từ khóa: ARIMA, mô hình dự báo, giá bán than, kế hoạch khai thác, mỏ lộ thiên

SUMMARY

Meeting the coal demand for thermal power to ensure energy security is one of the urgent issues nowadays as coal prices fluctuate constantly leading to coal mines without plans to produce the coal, the excessive (or shortage) of coal supplied to thermal power affects the production and business efficiency of the coal industry in general and open-pit coal mines in particular.

Faced with this situation, the study forecasting the fluctuation of the thermal coal price for the exploitation planning for Vietnam coal mines in general and the Vietnam open-pit coal mines in particular is very necessary. The authors have studied the use of the ARIMA model to predict the variation in the price of coal with the optimal model chosen as ARIMA (2,0, 1) (1,0,0) [12]. This model helps the coal mines in general and open-pit coal mines in particular can predict the coal price at specific times to set the optimal mining schedule.