

# ÁP DỤNG MÔ HÌNH ARIMA TRONG DỰ BÁO BIẾN ĐỘNG CHỈ SỐ VN30 VÀ LỢI SUẤT CHỈ SỐ VN30

Phan Trần Trung Dũng

Khoa Tài chính Ngân hàng, Trường Đại học Ngoại Thương

Email: fandzung@fuu.edu.vn

Lương Ngọc Tuấn Dũng

Trường Đại học Ngoại Thương

Email: lntd2101@gmail.com

Ngày nhận: 14/02/2020

Ngày nhận bản sửa: 13/3/2020

Ngày duyệt đăng: 05/6/2020

## Tóm tắt

Nghiên cứu này xác định khả năng ứng dụng mô hình tự hồi quy tích hợp trung bình trượt (ARIMA) để dự đoán biến động của chỉ số VN30 và lợi suất của chỉ số VN30 tại Việt Nam. Nghiên cứu sử dụng giá đóng cửa theo ngày của chỉ số VN30 tại sàn HOSE với tần suất theo tuần trong giai đoạn từ ngày 01 tháng 01 năm 2013 đến ngày 20 tháng 12 năm 2019. Kết quả cuối cùng cho thấy mô hình ARIMA (1,1,1) là phù hợp và tốt nhất nhằm dự báo biến động chỉ số VN30 khi được lược bỏ đi giai đoạn có biến động lớn của chỉ số. Ngoài ra, nghiên cứu cũng chỉ ra rằng mô hình ARIMA không phù hợp để dự báo biến động lợi suất của chỉ số VN30.

**Từ khóa:** ARIMA, Chỉ số chứng khoán, Dự báo, VN30 Index

**Mã JEL:** G10, G12, G17

## Application of ARIMA model to forecast the volatility of VN30 Index and VN30 Index's returns

### Abstract:

This study aims at showing the applicability of using Autoregressive integrated moving average (ARIMA) to forecast the volatility of VN30 Index and VN30 Index's returns. Data used in the study consists of weekly series of close index of VN30 at Ho Chi Minh Stock Exchange (HOSE) during 01/01/2013 and 20/12/2019. The results show that ARIMA (1,1,1) is the most suitable model to forecast the volatility of VN30 Index when excluding the large volatility. Furthermore, the results also show that ARIMA Models are not applicable in forecasting the volatility of VN30 Index's returns.

**Keywords:** ARIMA Model, Forecasting, Index, VN30 Index

**JEL Code:** G10, G12, G17

## 1. Mở đầu

Thị trường chứng khoán tại Việt Nam ra đời vào đầu năm 2000 và đã trở thành một trong các kênh đầu tư rất hấp dẫn đối với các nhà đầu tư, từ các tổ chức đầu tư lớn và chuyên nghiệp tới các nhà đầu tư cá nhân nhỏ lẻ. Theo quy tắc, song song với mức sinh lợi cao thì kênh đầu tư này luôn tồn tại nhiều

rủi ro tiềm ẩn vì biến động về giá của cổ phiếu trên thị trường thường rất khó để dự đoán. Chính vì vậy, việc đưa ra được các dự báo đối với biến động giá của các cổ phiếu cũng như biến động của thị trường nhằm đưa ra được các quyết định đầu tư cho việc kinh doanh cá nhân và tổ chức hay đưa ra các hoạch định cho tổ chức quản lý thị trường đã nhận được

nhieu sự quan tâm của các chuyên gia trong và ngoài nước. Tại Việt Nam, sự biến động của chỉ số VN30 về cơ bản cũng phản ánh được rủi ro của hệ thống bởi trong rô VN30 bao gồm 30 cổ phiếu lớn được chọn lọc theo các tiêu chí và thay đổi với tần suất 6 tháng một lần.

Trên thế giới, đã có nhiều các nghiên cứu áp dụng mô hình ARIMA với mục đích nhằm dự báo dữ liệu chuỗi thời gian như: CPI, GDP, tỷ giá hối đoái. Tại Việt Nam cũng đã có các nghiên cứu về mô hình ARIMA áp dụng nhằm dự báo chỉ số VN-Index nhưng số lượng chưa nhiều, ngoài ra thì các nhóm chỉ số của các cổ phiếu dẫn đầu như VN30 hay HNX30 chưa nhận được sự quan tâm nhiều. Trong bài nghiên cứu này, chúng tôi sử dụng dữ liệu với tần suất theo tuần với lý do các thông tin giao dịch trên thị trường chứng khoán Việt Nam diễn ra vào các ngày làm việc từ thứ hai tới thứ sáu, vì thế nếu sử dụng dữ liệu ngày sẽ gây khó khăn khi xử lý dữ liệu, nghiên cứu này không tập trung vào các biến động ở mức ngày, nên mô hình sử dụng dữ liệu theo tuần tạo ra sự ổn định trong ước lượng và vẫn giúp đạt được mục tiêu nghiên cứu.

## 2. Tổng quan nghiên cứu

Với các nghiên cứu nước ngoài, đã có rất nhiều các nghiên cứu sử dụng mô hình ARIMA đối với dữ liệu chuỗi thời gian là tài chính - kinh tế cũng như dữ liệu là giá của các mặt hàng thiết yếu trên thị trường. Đối với dữ liệu chuỗi thời gian về giá cổ phiếu và các chỉ số cổ phiếu của các sản chứng khoán, đã có nhiều các nghiên cứu lựa chọn mô hình ARIMA, công trình nghiên cứu của Onwukwe (2010) đã hệ thống hoá cơ sở lý thuyết của mô hình ARIMA và dùng chỉ số NSE cùng với dữ liệu theo tháng nhằm xây dựng mô hình ARIMA (1,1,1) nhằm dự báo thị trường cổ phiếu của Nigeria trong giai đoạn 1985 đến 2009. Kết quả dự báo của tác giả khá tương đồng với thực tế trong cùng giai đoạn, song lại vượt trội so với thực tế khi thị trường trong giai đoạn phải chịu tác động của khủng hoảng kinh tế - tài chính toàn cầu trong năm 2009. Kết quả nghiên cứu cho thấy hạn chế của mô hình ARIMA khi xuất hiện yếu tố khủng hoảng kinh tế, phá huỷ mối tương quan giữa hiện tại và quá khứ. Vì vậy, việc sử dụng mô hình ARIMA trong dự báo là rất tốt tại nhiều sai lệch khi gặp các biến chuyển bất thường lớn đối với thị trường chứng khoán. Ngoài ra, nghiên cứu này cũng sử dụng dự báo trên giả định rằng phương sai của tỷ suất sinh lợi không thay đổi theo thời gian, song thực tế thì chỉ số NSE là chuỗi thời gian có

tính biến động cao nên giả định đã không thực sự phù hợp.

Jung-Hua & Jia-Yuan (1996) nghiên cứu hệ thống các bước nhằm tiến hành xây dựng mô hình ARIMA để dự báo chỉ số TSEWSI. Trong bài nghiên cứu, tác giả đưa ra kết quả dự báo xu hướng chỉ số giá TSEWSI của thị trường chứng khoán Đài Loan trong trung hạn với mô hình ARIMA (1,2,1). Với kết quả trong bài thì việc dự báo cho thị trường chứng khoán Đài Loan cho 6 tuần có độ chính xác có thể chấp nhận được và tốt hơn các mô hình dự báo khác. Nghiên cứu cũng thực hiện việc dự báo dựa trên giả định rằng phương sai của tỷ suất sinh lợi không đổi theo thời gian, song thực tế thì giả định trên lại không phù hợp.

Al-Shiab (2006) sử dụng giá đóng cửa của chỉ số tại sản chứng khoán ASE theo ngày từ 4/1/2004 đến 10/8/2004, với phần mẫu ngoài được dùng để kiểm định là 7 ngày tiếp theo. Sau nhiều lần kiểm định các mô hình khác nhau, mô hình được lựa chọn là ARIMA (4,1,5) dự đoán rằng chỉ số ASE sẽ tiếp tục tăng trưởng ở mức 0.195% trong 7 ngày tiếp theo bắt đầu từ 11/8/2004. Tuy vậy, dự đoán này không thực sự khớp với các diễn biến thực tế trong khoảng thời gian dự đoán (11/8/2004 - 19/8/2004) bởi thực tế chỉ số ASE đã giảm 0.003%. Kết luận của tác giả cho rằng trong khoảng thời gian dự đoán, chỉ số ASE có diễn biến gần với lý thuyết thị trường hiệu quả ở dạng yếu.

Ayodel & cộng sự (2014) phát triển mô hình nhằm dự báo giá cổ phiếu của Nokia và Zenith Bank với lần lượt hai mô hình phù hợp nhất là ARIMA (2,1,0) và ARIMA (1,0,1). Theo kết quả nghiên cứu, nhóm tác giả cho rằng sử dụng mô hình ARIMA có đem lại kết quả dự báo tốt trong ngắn hạn và các nhà đầu tư có thể sử dụng mô hình này nhằm đưa ra các quyết định đầu tư có lợi. Prapanna & cộng sự (2014) nghiên cứu về tính hiệu quả trong việc sử dụng mô hình ARIMA với mục đích dự báo giá của 56 cổ phiếu trên nhiều ngành tại Ấn Độ. Theo nghiên cứu, mô hình ARIMA (1,0,2) được chọn là mô hình phù hợp thông qua giá trị của chỉ số AIC. Nhóm tác giả kết luận rằng độ chính xác của mô hình ARIMA trong dự báo giá cổ phiếu là 85% và nhóm ngành FMCG cho kết quả dự báo với độ chính xác cao nhất.

Fatai & cộng sự (2014) sử dụng mô hình ARIMA nhằm dự báo biến động chỉ số chứng khoán tại Botswana và Nigeria. Nhóm tác giả sử dụng các chỉ số tiêu chuẩn như AIC, BIC, HQC, RMSE và

MAE nhằm chọn ra mô hình ARIMA phù hợp nhất cho hai chỉ số chứng khoán được nghiên cứu. Theo kết quả nghiên cứu, nhóm tác giả kết luận mô hình ARIMA (3,1,1) là phù hợp dự báo biến động chỉ số chứng khoán tại Botswana, mặt khác mô hình ARIMA (1,1,4) phù hợp dự báo chỉ số chứng khoán tại Nigeria. Nhìn chung, các nghiên cứu sử dụng chuỗi thời gian đều thể hiện tính tương thích trong dự báo giá chứng khoán và tính sinh lợi của chứng khoán, chỉ số chứng khoán.

Ngoài ra, các nghiên cứu về dữ liệu chuỗi thời gian là các cặp tỷ giá hối đoái được giao dịch trên thị trường ngoại hối cũng áp dụng mô hình ARIMA với mục đích dự báo. Ince & Trafalis (2006) đã sử dụng các phương pháp khác nhau để dự báo tỷ giá hối đoái, bao gồm ARIMA, VAR, SVR và ANN. Nghiên cứu này chỉ ra rằng muốn áp dụng các phương thức khác nhau một cách phù hợp, kỹ thuật lựa chọn đầu vào đóng vai trò vô cùng quan trọng. Babu & Reddy (2015) nghiên cứu và so sánh tính hiệu quả của việc sử dụng mô hình ARIMA, Neural Network và Fuzzy Neuron với mục đích dự báo tỷ giá hối đoái của đồng Indian Rupee (INR) so với đô la Mỹ (USD), bảng Anh (GBP), Euro (EUR) và đồng Yên (JPY). Theo nghiên cứu, ARIMA (1,1,1) là mô hình tốt nhất đối với USD, GBP và JPY, đối với EUR mô hình ARIMA(1,1,0) được chọn. Kết luận của hai tác giả cho rằng nhằm dự báo tỷ giá hối đoái, sử dụng mô hình ARIMA cho kết quả tốt hơn sử dụng hai mô hình phi tuyến tính là Neural Network và Fuzzy Neuron.

Hơn nữa, mô hình ARIMA cũng được sử dụng trong một vài nghiên cứu dự báo giá của mặt hàng thiết yếu (gạo) hoặc dự báo biến động của chỉ số kinh tế vĩ mô. Ohyver & Herena (2018) phát triển mô hình chuỗi thời gian nhằm dự đoán giá gạo bằng mô hình ARIMA nhằm giúp cho chính phủ kiểm soát và điều chỉnh. Trong nghiên cứu, hai tác giả cho rằng việc kiểm soát giá gạo là cần thiết để giữ vững mức giá nhằm không làm ảnh hưởng tới nhà sản xuất và người tiêu dùng. Theo kết quả nghiên cứu, mô hình ARIMA (1,1,2) là phù hợp cho việc dự báo giá gạo trong khoảng thời gian từ tháng 1 năm 2015. Stanley & Mubita (2016) phát triển mô hình nhằm dự báo biến động lạm phát tại Zambia thông qua chỉ số CPI theo tháng từ tháng 5/2010 tới tháng 5/2014. Kết quả nghiên cứu chỉ ra mô hình ARIMA(12,1,0) là mô hình phù hợp để dự báo dữ liệu chuỗi thời gian CPI, qua đó phù hợp để dự báo biến động của lạm phát tại Zambia.

Tại Việt Nam, cũng đã có khá nhiều các nghiên cứu sử dụng mô hình ARIMA trong mục đích dự đoán biến động của dữ liệu chuỗi thời gian. Lê Tuấn Bách (2010) nêu ra các vấn đề nổi bật cần lưu ý và hướng dẫn mở rộng ứng dụng của mô hình. Tuy nhiên, nghiên cứu mới chỉ lựa chọn mô hình thích hợp để dự báo chỉ số mà chưa đưa ra kết quả dự báo từ mô hình dự báo. Hơn nữa, các ưu cũng như nhược điểm đến từ các mô hình thực hiện dự báo vẫn không được tác giả nêu rõ.

Bùi Quang Trung (2010) sử dụng dữ liệu chuỗi thời gian chứng khoán, xem giá trị trong quá khứ của chỉ số VNIndex là một chỉ tiêu nhằm phản ánh giá trị trong tương lai mô hình ARIMA với nguyên lý Box – Jenkins. Dù kết quả dự báo cho thấy rằng giá trị dự báo từ mô hình xấp xỉ giá trị thực tế với khoảng tin cậy 95% chứa giá trị thực tế song do nhiều yếu tố khác như tâm lý nhà đầu tư, các thay đổi lớn trong chính sách của chính quyền làm cho sai số của dự báo tăng. Ngoài ra, hạn chế của mô hình đó là việc giả định phương sai sai số của chuỗi là không thay đổi theo thời gian tuy nhiên trên thực tế thì chỉ số VNIndex có tính biến động lớn và phương sai có điều kiện.

Phạm Thị Thuý Liên (2014) cũng đã xây dựng khái quát về cơ sở lý thuyết của mô hình kinh tế lượng nhằm mục đích dự báo. Tác giả sử dụng chuỗi dữ liệu đóng cửa của chỉ số VNIndex trong khoảng thời gian từ 02/01/2007 đến 29/04/2014 để dự báo cho chỉ số VNIndex từ 05/05/2014 đến 31/05/2014. Theo kết quả trong luận văn, tác giả cho rằng mô hình ARMA mô phỏng tốt diễn biến hành vi của thị trường nên đã thực hiện khá tốt chức năng dự báo của nó trong tương lai. Ngoài ra, tác giả cũng kết luận rằng có sự tồn tại cân xứng giữa thông tin tốt và thông tin xấu trên thị trường chứng khoán tại Việt Nam, nói cách khác là tin tốt và tin xấu có tác động như nhau trên sản chứng khoán. Tuy nhiên, theo như tác giả kết luận thì mô hình ARMA – GARCH chỉ phù hợp cho việc dự báo ngắn hạn bởi lẽ trong tương lai luôn luôn tiềm ẩn các cú sốc lớn khó lường trước được trong khi chính các yếu tố đó lại không được nhận diện hay được đưa vào mô hình như một biến độc lập (mô hình chỉ sử dụng dữ liệu quá khứ làm biến hồi quy).

Trịnh Thị Phan Lan (2018) cũng đã sử dụng mô hình ARIMA nhằm dự báo biến động của tỷ giá hối đoái. Tác giả dự báo tỷ giá VND/USD và sử dụng chuỗi sai phân bậc 1 của chuỗi tỷ giá sau khi kiểm định thấy chuỗi gốc không có tính dừng. Thông qua

đồ thị ACF và PACF thì tác giả đã lựa chọn mô hình ARIMA (3,1,3) để dự báo tỷ giá, sau đó tác giả tiếp tục sử dụng mô hình SARIMA (3,1,0). Qua kết quả nghiên cứu, tác giả kết luận rằng mô hình SARIMA cho kết quả dự báo tốt nhất trong các mô hình được nghiên cứu. Sai số của mô hình SARIMA là không quá lớn và điều này cho thấy hoàn toàn có thể sử dụng mô hình để phục vụ mục đích dự báo tỷ giá trong tương lai. Tuy nhiên, tác giả cho rằng khi dự báo cũng cần sử dụng kết hợp nhiều mô hình để cho kết quả tốt hơn. Nhìn chung, bước đầu các nghiên cứu sử dụng dữ liệu chuỗi thời gian cũng cho thấy tính tương thích trong điều kiện thị trường chứng khoán Việt Nam.

Từ việc nghiên cứu tổng quan các công trình nghiên cứu đã công bố, các tác giả nhận thấy hoàn toàn có cơ sở để sử dụng dữ liệu dạng chuỗi thời gian và phương pháp phù hợp để ứng dụng trong việc dự báo chỉ số chứng khoán. Hơn nữa, qua tổng quan các nghiên cứu tại Việt Nam, nhóm tác giả nhận thấy các công trình nghiên cứu chỉ dừng lại ở dự báo biến động giá của các mã cổ phiếu hoặc biến động của chỉ số VNIndex nói chung; chưa có nghiên cứu nào được áp dụng vào dự báo biến động của chỉ số chứng khoán đại diện cho toàn sản chứng khoán, cụ thể ở đây là chỉ số VN30. Qua đó, để tạo ra tính mới trong nghiên cứu cũng như tính áp dụng thực tiễn, nhóm tác giả đã lựa chọn sử dụng mô hình ARIMA nhằm dự báo biến động chỉ số VN30.

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

#### 3.1. Dữ liệu

Nghiên cứu xem xét chỉ số VN30 đóng cửa theo ngày trong giai đoạn từ năm 2013 đến thời điểm ngày 20/12/2019, dữ liệu của nghiên cứu là dữ liệu thứ cấp và được cung cấp bởi phần mềm Finpro. Ngoài ra, dữ liệu về thành phần của rổ cổ phiếu VN30 cũng được cung cấp nhằm làm cho mục đích bài viết được sáng tỏ hơn. Dữ liệu được thu thập khoảng 7 năm (tương đương với 360 quan sát) do trong khoảng thời gian này sẽ bao gồm cả giai đoạn khó khăn của nền kinh tế cũng như giai đoạn hồi phục và tăng trưởng. Sau khi thu thập dữ liệu đóng cửa chỉ số VN30 theo ngày, để dễ dàng hơn trong việc xử lý số liệu, tác giả sử dụng dữ liệu giá đóng cửa chỉ số theo tuần. Lý do là các thông tin giao dịch trên thị trường chứng khoán chỉ diễn ra vào các ngày làm việc thứ 2 đến thứ 6, như vậy nếu lấy dữ liệu ngày thì hai ngày cuối trong tuần sẽ không có dữ liệu và gây khó khăn trong việc xử lý dữ liệu. Bên cạnh đó, các tác giả cũng cho rằng dữ liệu theo tuần giúp

làm trơn mô hình dự báo và không gây ảnh hưởng tới kết quả nghiên cứu.

#### 3.2. Phương pháp nghiên cứu

Mô hình sử dụng dữ liệu chuỗi thời gian, xem giá trị trong quá khứ của một biến số cụ thể là một chỉ tiêu tốt phản ánh giá trị trong tương lai của nó, cụ thể, cho  $Y_t$  là giá trị của biến số tại thời điểm  $t$  với  $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_0, t)$ . Mục đích của phân tích là để thấy rõ một số mối quan hệ giữa các giá trị  $Y_t$  được quan sát đến nay để cho phép chúng ta dự báo giá trị  $Y_t$  trong tương lai. Phương pháp này đặc biệt hữu ích cho việc dự báo trong ngắn hạn. Mô hình tự hồi quy  $p$  - 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:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \epsilon_t$$

Mô hình trung bình trượt  $q$  - MA( $q$ ): trong mô hình trung bình trượt, quá trình được mô tả hoàn toàn bằng tổng trọng số của các nhiễu nhiên hiện hành có độ trễ:

$$Y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

Mô Hình Hồi Quy Kết Hợp Trung Bình Trượt - ARMA( $p, q$ ):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

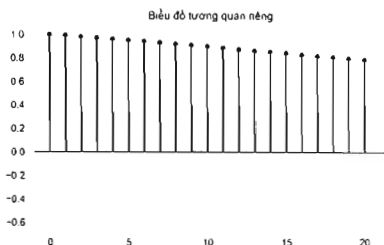
##### 3.2.1. Kiểm định tính dừng

Điều trước tiên cần phải lưu ý là hầu hết các chuỗi thời gian đều không dừng, và các thành phần AR và MA của mô hình ARIMA chỉ liên quan đến các chuỗi thời gian dừng. Quy trình ngẫu nhiên của  $Y_t$  được xem là dừng nếu trung bình và phương sai của quá trình không thay đổi theo thời gian và giá đồng phương sai giữa hai thời đoạn chi phụ thuộc vào khoảng cách độ trễ về thời gian giữa các thời đoạn này chứ không phụ thuộc vào thời điểm thực tế mà đồng phương sai được tính. Do đó, để nhận diện mô hình ARIMA, chúng ta phải thực hiện hai bước sau: Có ba cách để nhận biết tính dừng của một chuỗi thời gian là dựa vào trên đồ thị của chuỗi thời gian, đồ thị của hàm tự tương quan mẫu hay kiểm định Dickey - Fuller.

##### 3.2.2. Nhận dạng mô hình

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 SPACF =  $f(t)$  và SACF =  $f(t)$ . Với SACF là hàm tự tương quan mẫu và SPACF là

**Hình 1: Hệ số tương quan ACF chuỗi đóng cửa chỉ số VN30**



*Nguồn. Tổng hợp từ Python.*

hàm tự tương quan mẫu riêng phần (Sample Partial Autocorrelation).

Chọn giá trị của  $p$  nếu đồ thị SPACF có giá trị cao tại độ trễ 1, 2, ...,  $p$  và giảm nhiều sau  $p$  và dạng hàm SAC giảm dần. Chọn giá trị của  $q$  nếu đồ thị SACF có giá trị cao tại độ trễ 1, 2, ...,  $q$  và giảm nhiều sau  $q$  và dạng hàm SPAC giảm dần.

### 3.2.3. Ước lượng mô hình

Sau khi ước lượng các tham số của một mô hình ARIMA được nhận dạng thử, chúng ta cần phải kiểm định để kiểm nghiệm rằng mô hình là thích hợp. Các cách thức để thực hiện điều này: Kiểm tra phần dư et có phải là nhiễu trắng không. Nếu et là nhiễu trắng thì chấp nhận mô hình, trong trường hợp ngược lại chúng ta phải tiến hành lại từ đầu. Các kiểm định có thể sử dụng là kiểm định BP (Box-Priere) hoặc kiểm định Ljung-box với trị thống kê Q, hoặc kiểm định LM. Nếu tồn tại nhiều hơn một mô hình đúng, mô hình có AIC (Akaike Information Criterion) nhỏ nhất sẽ được lựa chọn.

## 4. Kết quả nghiên cứu

### 4.1. Kết quả nghiên cứu đối với dữ liệu chuỗi giá đóng cửa VN30

Từ Hình 1, có thể thấy dữ liệu chỉ số đóng cửa theo tuần của VN30 là một chuỗi thời gian có xu thế tăng theo thời gian và có yếu tố bất thường và không có tính dừng. Không thể ứng dụng mô hình ARIMA trực tiếp cho chuỗi thời gian này. Để chắc chắn đây là chuỗi không có tính dừng, tác giả sử dụng đồ thị hệ số tương quan ACF như hình 1, có thể thấy các hệ số tương quan của các giá trị trễ trong mô hình không có dạng theo hình sin hoặc không có dạng

của đồ thị hàm mũ, mà giảm về 0 một cách từ từ. Điều này khẳng định dữ liệu chuỗi thời gian không có tính dừng.

Chính vì vậy, thay vì sử dụng trực tiếp để dự báo thì chuỗi thời gian cần được biến đổi về chuỗi có tính dừng rồi mới tiếp tục sử dụng mô hình ARIMA để dự báo.

Chuỗi sai phân đã được loại bỏ xu thế tăng của chuỗi gốc, tuy nhiên vẫn còn một số kì có giá trị bất thường, rơi vào đầu năm 2018. Tiếp tục kiểm định tính dừng của chuỗi sai phân cấp 1 thu được kết quả như sau:

Thống kê ADF: -13.772776

Tham số p-value 0.000000

Giá trị tới hạn:

1%: -3.4486972813047574.3f

5%: -2.8696246923288418.3f

10%: -2.571077032068342.3f

Bác bỏ  $H_0$  – Chuỗi thời gian có tính dừng

Kiểm định này có cấp giả thuyết ban đầu:

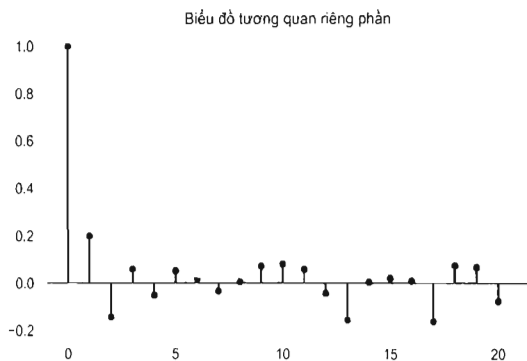
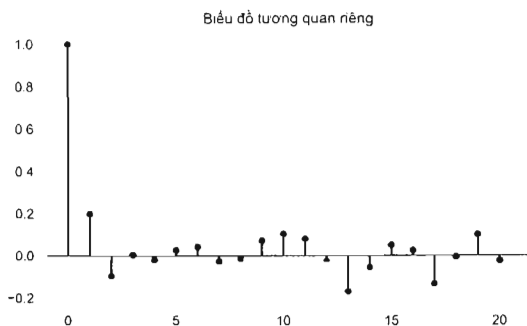
$$\begin{cases} H_0: \text{Chuỗi thời gian không có tính dừng} \\ H_1: \text{Chuỗi thời gian có tính dừng} \end{cases}$$

Dựa vào kết quả trên, giá trị thu về là p-value = 0 <  $\alpha = 0.05 \Rightarrow$  Bác bỏ giả thuyết  $H_0 \Rightarrow$  Chuỗi thời gian có tính dừng, đến đây có thể tiếp tục thực hiện mô hình ARIMA.

Do đã xác định chuỗi lợi suất ban đầu có tính dừng nên d có thể nhận giá trị 1 hoặc lớn hơn 1.

Tiếp theo, sử dụng đồ thị ACF và PACF nhằm xác định hai chỉ số p và q trong mô hình

Hình 2: Biểu đồ ACF và PACF của chuỗi đóng cửa chỉ số VN30



Nguồn: Tổng hợp từ Python.

#### ARIMA (p,d,q).

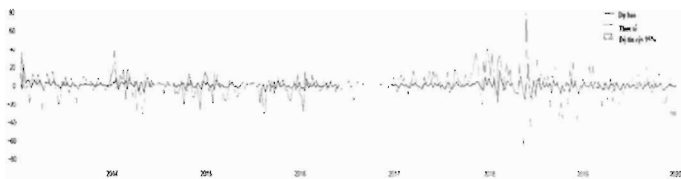
Chọn p dựa trên đồ thị ACF (phía trái), những giá trị nằm ngoài phần in đậm màu xanh là các giá trị có ý nghĩa thống kê ở mức 5%, tương tự nguyên tắc chọn q cũng là những giá trị nằm ngoài đường màu xanh là các đường có ý nghĩa thống kê. Căn cứ vào đồ thị cũng có thể chọn các giá trị  $p=1$  hoặc  $p=13$ . q có thể nhận các giá trị 1,2,13.

Do tại đồ trễ 1 cả giá trị ACF và PACF đều có dạng như nhau, tác giả quyết định chọn  $p=2$  và  $q=1$ .

Sau khi tiến hành chạy mô hình ARIMA (2.1.1) với chuỗi dữ liệu đã có kết quả như tại Hình 3.

Từ kết quả dự báo có thể thấy, tại các thời điểm chỉ số ít biến động thì kết quả dự báo có thể chấp nhận được, tuy nhiên tại những điểm có biến động lớn thì dự báo vẫn chưa thực sự hiệu quả. Có thể khắc phục

Hình 3: Kết quả dự báo chuỗi đóng cửa chỉ số VN30



Nguồn: Tổng hợp từ Python

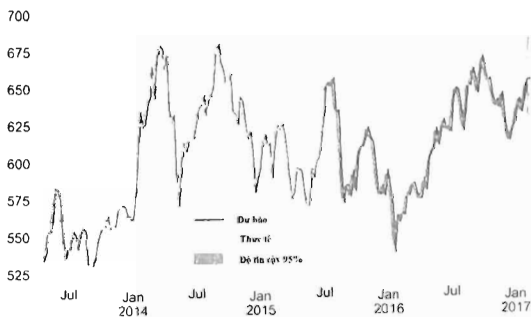
bằng cách thêm biến giả đại diện cho giai đoạn bất thường vào cuối năm 2017 và đầu năm 2018. Điều này có nghĩa là sẽ mở rộng mô hình ARJMA. Độ chính xác của ước lượng:

```
{'mape': 0.32676746454787753,  
'me': -339.70415834523925,  
'mae': 339.8243155145582,  
'mpe': -0.3266629190208065,  
'rmse': 352.642196103886,  
'corr': 0.832900183088708,  
'minmax': 0.326766712985994}
```

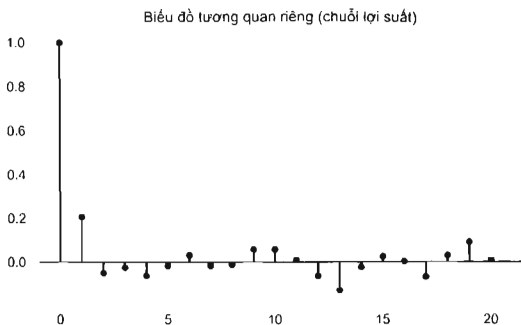
Từ kết quả đo lường độ tin cậy của dự báo, thông qua chỉ số phần trăm sai số tuyệt đối trung bình (MAPE) có thể nhận thấy rằng độ tin cậy của dự báo đang ở mức 32,67%. Điều này đồng nghĩa với việc dự báo này chỉ dừng ở mức độ hợp lệ chứ chưa có độ tin cậy cao (theo Lewis 1982).

Để khẳng định lại kết luận trên, bài báo tiến hành dự báo lại chuỗi VN30 trong giai đoạn có ít biến động bất thường hơn. Cụ thể, nghiên cứu sẽ điều chỉnh lại khoảng thời gian dự báo từ năm 2013 đến hết tháng 1 năm 2017 (bỏ bớt khoảng thời gian từ tháng 2 năm 2017 cho đến hết năm 2019 vì khoảng

Hình 4: Kết quả dự báo 2013-2017



Nguồn: Tổng hợp từ Python



Nguồn: Tổng hợp từ Python.

thời gian này có chứa biến động tăng cao bất thường của chuỗi giá đóng cửa VN30).

Tiến hành thực hiện chạy mô hình ARIMA (1,1,1), chỉ số AIC của mô hình là 1638.5, trong khi chỉ số AIC của mô hình trên là 3025.61. Như vậy mô hình của chuỗi này có giá trị nhỏ hơn, chứng tỏ độ phù hợp tốt hơn so với mô hình trước. Đồ thị dự báo của chuỗi VN30 khi sử dụng mô hình ARIMA (1,1,1) như sau:

Có thể nhận thấy các giá trị dự báo và giá trị thực thể ở trong mẫu gần sát nhau. Giá trị tổng bình phương các phần dư của mô hình này là 14544.54. Trong khi phần dư của mô hình trên là 124356.52. Như vậy, khi không có biến động bất thường lớn sai số của mô hình dự báo nhỏ hơn nhiều so với mô hình có yếu tố biến động lớn. Độ chính xác trong mẫu của dự báo mô hình này.

`['mape': 0.15937402014840196,`

`'me': -114.71459306935606,`

`'mae': 114.71459306935606,`

`'mpe': -0.15937402014840196,`

`'rmse': 120.60077142771605,`

`'corr': 0.5052258308914742,`

`'mnmmax': 0.15937402014840174]`

Có thể thấy chỉ số MAPE = 15.93% nhỏ hơn con số 32.67% của mô hình bên trên. Điều này đồng

nghĩa với việc mô hình phía dưới sẽ có độ chính xác cao hơn mô hình trước đó.

#### 4.2. Kết quả nghiên cứu đối với dữ liệu chuỗi lợi suất chỉ số VN30

Từ hình 5, có thể thấy các giá trị tương quan đã tuân theo quy luật hình sin và giảm nhanh về 0 hoặc về âm nên có thể kết luận dữ liệu lợi tức có tính dừng. Tuy nhiên để chắc chắn về kết luận trên, các tác giả sử dụng kiểm định nghiệm đơn vị thu được kết quả như sau:

Thông kê ADF: -13.471716

Tham số p-value 0.000000

Giá trị tới hạn:

1%: -3.44880082033912.3f

5%: -2.869670179576637.3f

10%: -2.5711012838861036.3f

Không chấp nhận  $H_0$  – Chuỗi thời gian có tính dừng

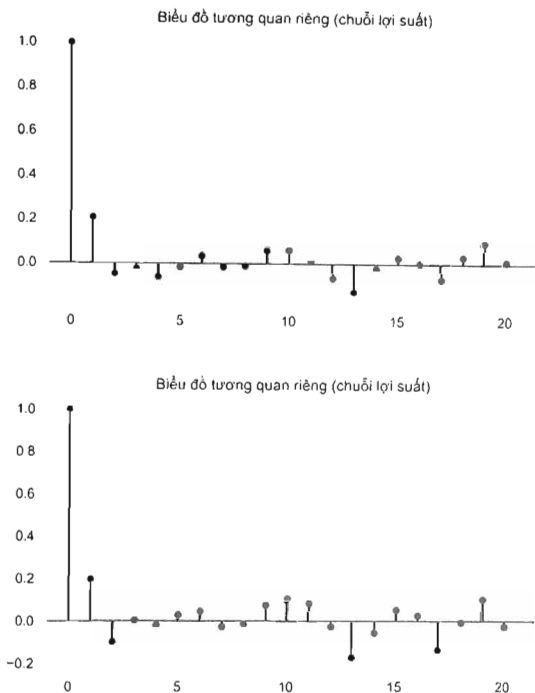
Kiểm định này có cấp giả thuyết ban đầu:

$\begin{cases} H_0: \text{Chuỗi thời gian không có tính dừng} \\ H_1: \text{Chuỗi thời gian có tính dừng} \end{cases}$

Dựa vào kết quả phía trên thu được p-value = 0 <  $\alpha$  = 0.05 => Bác bỏ giả thuyết  $H_0$  => Chuỗi thời gian có tính dừng, đến đây có thể tiếp tục thực hiện



Hình 6: Giản đồ ACF và PACF chuỗi lợi suất chỉ số VN30



Nguồn: Tổng hợp từ Python.

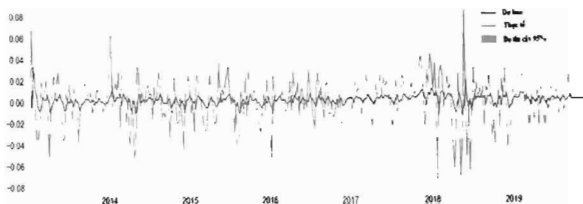
mô hình ARIMA.

Bước tiếp theo trong quá trình sử dụng mô hình ARIMA để dự báo là nhận dạng mô hình ARIMA(p,d,q) thích hợp, đồng nghĩa với việc tìm các giá trị thích hợp của p, d và q (với d là bậc sai phân của chuỗi dữ liệu thời gian được khảo sát, p là bậc tự hồi quy và q là bậc trung bình trượt). Các giá trị này được xác định dựa vào biểu đồ tự tương quan (ACF) và biểu đồ tự tương quan riêng phần (PACF). Trong đó, việc lựa chọn mô hình AR (p) phụ thuộc vào biểu đồ PACF nếu nó có giá trị cao tại các độ trễ

1, 2, ..., p và giảm đột ngột sau đó, đồng thời dạng hàm ACF tắt lịm dần. Tương tự, việc lựa chọn mô hình MA (q) dựa vào biểu đồ ACF nếu nó có giá trị cao tại các độ trễ 1, 2, ..., q và giảm mạnh sau q, đồng thời dạng hàm PACF tắt lịm dần. Do đã xác định chuỗi lợi suất ban đầu có tính dừng nên d có thể nhận giá trị 0 hoặc 1. Từ giản đồ tương quan tác giá chọn p=q=1 và d=0.

Sau khi nhận diện được mô hình và tìm được mô hình phù hợp với chuỗi lợi suất chỉ số VN30, mô hình ARIMA (1,0,1) được lựa chọn và thu được kết

Hình 7: Kết quả dự báo chuỗi lợi suất chỉ số VN30



Nguồn: Tổng hợp từ Python.

quá dự báo như Hình 7.

Kết quả trên cho thấy việc sử dụng ARIMA dự báo lợi suất chuỗi VN30 là không thực sự hiệu quả. Độ chính xác của mô hình này là:

```
{'mape': 1.2665518665315432,
'me': -0.00042517128979493616,
'mae': 0.015531442382899349,
'mpe': -1.2074310864773665,
'rmse': 0.020725022070341843,
'corr': -0.02328750553628702,
'minmax': 8.329912400984576}
```

Từ phần đo lường độ tin cậy của mô hình phía trên, có thể thấy rằng chỉ số MAPE ở mức 126,65%. Theo đó, thì với các mô hình có chỉ số MAPE lớn hơn 50% sẽ được coi là mô hình dự báo không chính xác. Như vậy, đây là mô hình không được công nhận

##### 5. Thảo luận và một số khuyến nghị đối với nhà đầu tư

Từ các kết quả nghiên cứu định lượng ở trên, có thể kết luận rằng với mục đích nhằm dự báo biến động của chuỗi VN30 hoặc biến động lợi suất của chuỗi VN30 thì mô hình tự hồi quy tích hợp trung bình trượt ARIMA sẽ phù hợp hơn với việc dự báo biến động của chuỗi VN30. Hơn nữa, việc dự báo biến động của chuỗi VN30 sẽ chỉ phù hợp đối với các khoảng thời gian xem xét có ít biến động đến từ dữ liệu đóng cửa của chỉ số VN30. Trên cơ sở này, nghiên cứu đưa ra một số gợi ý nhằm ứng dụng dữ liệu chuỗi thời gian hiệu quả hơn trong thực tiễn tại Việt Nam.

Bài nghiên cứu cũng đưa ra một số khuyến nghị

dành cho các nhà đầu tư. Với các nhà đầu tư trên sàn chứng khoán HOSE (Thành phố Hồ Chí Minh), từ kết quả dự báo trong mẫu dữ liệu với biên độ của các biến động bất thường là không lớn, mô hình ARIMA biểu diễn được kết quả so với thực tế ở mức tin cậy khá cao. Ngoài ra, đối với dự báo ngoài mẫu với sai số nhỏ hơn 5% thì nhà đầu tư hoàn toàn có thể tin tưởng và sử dụng mô hình ARIMA nhằm mục đích dự báo biến động của chỉ số VN30. Cụ thể, đối với dự báo ngoài mẫu 16 tuần (khoảng 4 tháng) giao dịch, nhà đầu tư trong giai đoạn tới có thể tập trung và dồn nhiều vốn vào các nhóm mã cổ phiếu thuộc rổ VN30 bởi theo kết quả dự đoán ngoài mẫu của mô hình cho thấy chỉ số VN30 có xu hướng tăng trong khoảng thời gian dự báo là 16 tuần tiếp theo. Như vậy, trong khoảng 4 tháng tiếp theo, nhà đầu tư trên thị trường hoàn toàn có thể đầu tư một cách chủ động và tăng tỉ trọng các mã cổ phiếu lớn thuộc nhóm VN30, một số các cổ phiếu lớn đã nằm trong nhóm VN30 lâu như: VNM, VCB, VIC,...

Tuy nhiên, nhà đầu tư trên sàn chứng khoán Thành phố Hồ Chí Minh nên bình tĩnh và cẩn trọng hơn vào một số khoảng thời gian nhạy cảm đối với thị trường chứng khoán nói riêng và bộ chỉ số như VN30 nói chung. Ví dụ như tháng 1 hàng năm là thời điểm mà Sở giao dịch chứng khoán Thành phố Hồ Chí Minh sẽ cung cấp thông tin về các mã cổ phiếu mới đủ tiêu chuẩn được liệt vào rổ VN30 và loại ra các mã cổ phiếu không có đủ tiêu chuẩn.

Ngoài ra, do những cú sốc về yếu tố kinh tế vĩ mô không chỉ tại Việt Nam mà trên toàn cầu hoàn toàn có thể tạo ra những biến động lớn trên thị trường và ảnh hưởng tới chỉ số VN30, ví dụ như dịch Virus

Covid 2019 đang làm cho các chỉ số chứng khoán thể giới đi xuống với mức độ mạnh mẽ, điều này có thể ảnh hưởng lớn tới độ chính xác vì phần nghiên cứu định lượng đã cho thấy mô hình chi hoạt động tốt trong điều kiện thị trường có độ ổn định tương đối. Vì vậy, kết quả dự báo từ mô hình hoàn toàn chi mang tính chất tham khảo đối với các nhà đầu tư.

Bảng việc sử dụng mô hình tự hồi quy tích hợp trung bình trượt ARIMA(1,1,1), nghiên cứu đã chỉ ra tính ứng dụng của việc sử dụng dữ liệu chuỗi thời gian trong dự báo biến động của chỉ số chứng khoán

VN30 tại Việt Nam. Mô hình này có tính ứng dụng cao trong điều kiện dự báo biến động của chỉ số chứng khoán đóng cửa, trong điều kiện thị trường ổn định. Tuy nhiên mô hình không hoạt động tốt trong việc dự báo biến động của mức sinh lợi chỉ số VN30 cũng như không thực sự hiệu quả khi thị trường có những cú sốc khiến giá cả chứng khoán biến động mạnh. Trên cơ sở các kết quả nghiên cứu, bài báo cũng đưa ra một số khuyến nghị dành cho nhà đầu tư để có thể ứng dụng mô hình này một cách phù hợp nhất.

## Tài liệu tham khảo

- Al-Shiab, M. (2006), 'The predictability of the Amman Stock Exchange Using Univariate Autoregressive Integrated Moving Average (ARIMA) Model', *Journal of Economic & Administrative Sciences*, 22, (2), 17-35
- Ayodele A. Adebisi, Aderemi O. Adewumi & Charles K. Ayo (2014), 'Stock Price Prediction Using the ARIMA Model', *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 105-111.
- Babu, A.S. & Reddy, S.K (2015), 'Exchange Rate Forecasting using ARIMA, Neural Network and Fuzzy Neuron', *Journal of Stock & Forex*, 4 (3), retrieved on March 13<sup>th</sup> 2020, from DOI: 10.4172/2168-9458.1000155.
- Bùi Quang Trung (2010), 'Ứng dụng mô hình ARIMA để dự báo VNIndex', *Tuyển tập Báo cáo Hội nghị Sinh viên Nghiên cứu Khoa học lần thứ 7*, Đại học Đà Nẵng năm 2010, 82-86.
- Fatai A.A, Ramysamy S & Dahud K.S (2014), 'Forecasting Stock Market Series with ARIMA Model', *Journal of Statistical and Econometric Methods*, 3 (3), 65-77.
- Ince, H., & Trafalis, T. B. (2006). 'A hybrid model for exchange rate prediction', *Decision Support Systems*, 42(2), 1054-1062.
- Jung-Hua, W. & Jia-Yann, L. (1996). 'Stock market trend prediction using ARIMA-based neural networks'. *Proceedings of International Conference on Neural Networks (ICNN '96)*, 2160-2165.
- Lê Tuấn Bách (2010), 'Phân tích dự báo giá và rủi ro thị trường cổ phiếu niêm yết Việt Nam', Luận văn thạc sỹ, Đại học Kinh tế Thành phố Hồ Chí Minh.
- Onwukwe, E.K (2010), 'Forecasting Nigerian Stock Exchange Returns: Evidence from Autoregressive Integrated Moving Average (ARIMA) Model', retrieved on March 13<sup>th</sup> 2020, from <http://dx.doi.org/10.2139/ssrn.1633006>.
- Ohyver, M. & Herena. P. (2018). 'Arima model for forecasting the price of medium quality rice to anticipate price fluctuations', *3<sup>rd</sup> International Conference on Computer Science and Computer Intelligent*, 135, 707-711
- Phạm Thị Thuý Liên (2014), 'Vận dụng mô hình ARMA – GARCH trong dự báo chỉ số VNIndex', Luận văn thạc sỹ, Đại học Đà Nẵng.
- Prapanna, M., Labani, S. & Saptarsi, G. (2014). 'Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices', *International Journal of Computer Science, Engineering and Applications*, 13-29.
- Stanley J. & Mubita S (2016), 'Forecasting Inflation Rate of Zambia Using Holt's Exponential Smoothing', *Open Journal of Statistics*, 6(2), 363-372.
- Trịnh Thị Phan Lan (2018), 'Dự báo biến động tỷ giá bằng mô hình ARIMA', *Tap chí tài chính*, 2(2), 45-47