



Thống kê Bayes và ứng dụng trong dự báo giá chứng khoán của các ngân hàng và công ty tài chính ở Việt Nam*

NGUYỄN PHÁT ĐẠT ^{a,**}, LÊ THANH HOA ^b, PHẠM THẾ BẢO ^c

^a Trường Đại học Kinh tế TP. Hồ Chí Minh

^b Trường Đại học Kinh tế - Luật, Đại học Quốc gia TP. Hồ Chí Minh

^c Trường Đại học Sài Gòn

THÔNG TIN	TÓM TẮT
Ngày nhận: 10/06/2020 Ngày nhận lại: 02/09/2020 Duyệt đăng: 09/10/2020 Mã phân loại JEL: C11; B26 Từ khóa: Thống kê Bayes; Dự báo giá chứng khoán; Mã chứng khoán ngân hàng; Công ty tài chính. Keywords: Bayesian statistics; Stock price forecasts;	Dự báo là một vấn đề rất được quan tâm, đặc biệt trong lĩnh vực kinh tế - tài chính. Một khi dự báo tốt, kết quả dự báo sẽ rất hữu ích cho các doanh nghiệp trong việc đưa ra các chính sách quản lý, cũng như giúp các nhà đầu tư tối đa hóa lợi nhuận. Bên cạnh đó, theo Quyết định số 242/QĐ-TTg ngày 28/2/2019 của Thủ tướng Chính phủ về phê duyệt đề án “cơ cấu lại thị trường chứng khoán và thị trường bảo hiểm đến năm 2020 và định hướng đến năm 2025” đòi hỏi các công ty phải chuẩn hóa hoạt động nghiệp vụ kinh doanh, báo cáo chính xác và minh bạch theo tiêu chuẩn quốc tế. Nhờ các thông tin chính xác và minh bạch, các nhà đầu tư sẽ có đầy đủ cơ sở giúp đưa ra các quyết định đầu tư, góp vốn hiệu quả nhất. Thống kê Bayes sử dụng phối hợp thông tin hàm hợp lý từ các giá trị quá khứ và thông tin tiên nghiệm sẽ đưa ra kết quả dự báo phù hợp với thực tế hơn nên chính xác hơn. Các bài toán dự báo ngắn hạn thường tại thời điểm t sẽ dự báo thời điểm ngay sau đó là $(t + 1)$; trong khi đó, đối với thị trường chứng khoán Việt Nam, nhà đầu tư cần phải dự báo đến thời điểm $(t + 3)$ – đây là thời điểm nhà đầu tư mới có thể giao dịch cổ phiếu mới mua, tức là nhà đầu tư cần mở rộng hơn kết quả dự báo nhằm lường trước các rủi ro có thể xảy ra. Hơn thế nữa, thị trường chứng khoán liên tục có các tin tức cập nhật nên các tham số dự báo được coi như là hằng số sẽ không còn phù hợp. Chính vì vậy, nhóm tác giả thực hiện nghiên cứu ứng dụng thống

* Bài viết này thuộc chuyên san Toán kinh tế do PGS.TS. Lê Xuân Trường (lxuantruong@ueh.edu.vn) phụ trách.

** Tác giả liên hệ: Lê Thanh Hoa.

Email: datnguyen60.ck28@st.ueh.edu.vn (Nguyễn Phát Đạt), hoalt@uel.edu.vn (Lê Thanh Hoa), ptbao@sgu.edu.vn (Phạm Thế Bảo).
Trích dẫn bài viết: Nguyễn Phát Đạt, Lê Thanh Hoa, & Phạm Thế Bảo. (2020). Thống kê Bayes và ứng dụng trong dự báo giá chứng khoán của các ngân hàng và công ty tài chính ở Việt Nam. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 31(4), 40–63.

Bank securities code;
Financial companies.

kê Bayes trong bài toán dự báo giá chứng khoán của các ngân hàng thương mại và công ty tài chính đến thời điểm $(t + 3)$. Ứng dụng cho kết quả dự báo khoảng đúng khá cao, đạt 89,02% (dự báo khoảng với xác suất 68%).

Abstract

Forecasting has gained much attention in recent years, especially in economics and finance sectors. Once the forecast is good enough, the result will be helpful for enterprises in making decisions as well as for investors in maximizing profit. According to Decision No. 242/QĐ-TTg dated February 28th, 2019 of the Vietnamese Prime Minister approves the project "Restructuring the securities market and insurance market to 2020 and orientation to 2025", it requires companies to standardize their business operations, report accurately and transparently according to international standards. Thanks to transparent and accurate information, investors have sufficient information for making the most significant capital investment decisions. The Bayesian statistics combine past information and prior knowledge for making accurate forecasts. Short-term forecasting problems usually use the information at time t to predict the one at $t + 1$. However, for the Vietnamese stock market, investors need a prediction at $t + 3$, when it is possible for trading recently purchased stocks. Under these circumstances, investors could minimize potential risks. Moreover, when the stock market data and information are frequently updated, leading to the forecasting parameter usage as constant is no longer suitable. Therefore, in this paper, we apply the Bayesian statistic to predict the stock price of banking and financial companies, up to $t + 3$, with a forecasting performance of 89% and the prediction interval of 68%.

1. Giới thiệu

Trong thống kê tần suất, các tham số dự báo cũng là những hằng số cố định, nhưng không được biết. Tuy nhiên, khi dữ liệu được cập nhật theo thời gian, các tham số dưới dạng hằng số không còn biểu diễn được sự thay đổi này. Vì vậy, cần các biểu diễn tham số dưới dạng mới thông qua dạng biến ngẫu nhiên theo cách tiếp cận thống kê Bayes (Bolstad & Curran, 2016).

Hơn thế nữa, trong thống kê Bayes, ngoài các thông tin từ dữ liệu thông qua hàm hợp lý, còn bổ sung thông tin tiên nghiệm từ các nghiên cứu trước đó (Robert, 2017); do đó, thống kê Bayes sử dụng được đa dạng nguồn thông tin nên được kỳ vọng sẽ dự báo chính xác hơn so với thống kê tần suất. Việc áp dụng thống kê Bayes trong dự báo giá chứng khoán là một vấn đề cần thiết trong phát triển các mô hình dự báo tốt hơn. Dự báo là một vấn đề có tính ứng dụng cao trong các doanh nghiệp; một kết quả dự báo tốt sẽ giúp doanh nghiệp điều hành, kiểm soát và đặt ra các phương án hành động với từng tình huống cụ thể giúp doanh nghiệp phát triển một cách ổn định.

Trong dự báo, thông thường có các phương pháp dự báo là dự báo ngắn hạn, trung hạn và dài hạn. Cụ thể:

- Dự báo trung và dài hạn thông qua phương pháp dự báo định tính, đây là cách tiếp cận của các chuyên gia rất am hiểu về đặc trưng của từng ngành. Dưới góc độ rất am hiểu các ngành, nhận xét của các chuyên gia phản ánh bản thân nội lực của ngành, công ty trong bối cảnh quốc gia và khu vực, do đó sẽ có các nhận xét, dự đoán về xu thế chung của vận động giá chứng khoán và không phản ánh hết sự lên xuống của giá chứng khoán thực tế theo từng ngày.

- Dự báo ngắn hạn thông qua phương pháp dự báo định lượng, để dữ liệu nói lên kết quả, với một số chỉ số nghiên cứu phổ biến như: GDP, lạm phát, mức tăng trưởng hoặc giá cả thị trường... và một số mô hình dự báo phổ biến như: ARIMA, SARIMA, ARCH, GARCH... Tuy nhiên, các mô hình này khó thể hiện được xu hướng vận động của thị trường khi thị trường thay đổi đột ngột (Guha & Bandyopadhyay, 2016). Hạn chế của các mô hình này là do các tham số của mô hình chỉ là các hằng số nên không cập nhật kịp sự biến đổi của dữ liệu.

Đối với dự báo bằng thống kê Bayes, khi dữ liệu được cập nhật, các tham số được cập nhật lại thông qua dạng biến ngẫu nhiên nên phản ánh được sự vận động của thị trường một cách chính xác hơn. Đặc biệt, theo đề án “Cơ cấu lại thị trường chứng khoán và thị trường bảo hiểm đến năm 2020 và định hướng đến năm 2025” thì sẽ có 100% ngân hàng thương mại niêm yết, đăng ký giao dịch trên thị trường chính thức. Thông qua đề án này, các ngân hàng thương mại cần chuẩn hóa hoạt động nghiệp vụ kinh doanh, báo cáo chính xác và minh bạch theo tiêu chuẩn quốc tế giúp nhà đầu tư có nhiều thông tin rõ ràng khi đưa ra các quyết định đầu tư, góp vốn... Vì thế, việc dự báo tốt giá chứng khoán của các ngân hàng thương mại không chỉ giúp cho các ngân hàng kịp thời đưa ra các quyết định điều hành hợp lý mà còn giúp các nhà đầu tư đưa ra quyết định đầu tư đúng đắn.

Bên cạnh đó, một số nghiên cứu về thống kê Bayes trong dự báo cũng đạt được kết quả khả quan như: Tìm khoảng dự báo dựa vào hàm mật độ hậu nghiệm Bayes mờ (Lê Thanh Hoa và cộng sự, 2017a), thống kê Bayes mờ trong dự báo giá chứng khoán (Phạm và cộng sự, 2017), dự báo tỷ giá và chỉ số kinh tế (Lê Thanh Hoa và cộng sự, 2017b).

Trong bài báo này, nhóm tác giả ứng dụng thống kê Bayes để dự báo giá chứng khoán của các ngân hàng thương mại và công ty tài chính thông qua dự báo ngắn hạn. Thông thường, trong các mô hình dự báo giá chứng khoán ngắn hạn, tại thời điểm t sẽ dự báo giá tại thời điểm $(t + 1)$. Tuy nhiên, ở thị trường chứng khoán Việt Nam, nhà đầu tư cần phải có đến ba ngày thì cổ phiếu mới về đến tài khoản để có thể giao dịch, do đó, để chủ động cho nhà đầu tư thì tại thời điểm t cần dự báo giá tại các thời điểm $(t + 1)$, $(t + 2)$, $(t + 3)$. Đây chính là lý do nhóm tác giả đề xuất mô hình nghiên cứu mới về dự báo giá chứng khoán tới thời điểm $(t + 3)$ theo thống kê Bayes. Kết quả ứng dụng về dự báo các mã chứng khoán ngân hàng và công ty tài chính có kết quả khả quan hơn so với trường hợp chỉ dự báo dựa vào mỗi thời điểm $(t + 1)$. Tất nhiên, nếu dựa vào giá trị dự báo $(t + 1)$ dự báo tiếp $(t + 2)$, dựa vào giá trị dự báo $(t + 2)$ dự báo tiếp $(t + 3)$ thì kết quả dự báo đúng thấp hơn rất nhiều so với cách dự báo của nhóm tác giả.

2. Thống kê Bayes

2.1. Công thức xác suất đầy đủ và định lý Bayes

Giả sử tập các biến cố $B_1, B_2, \dots, B_n$ thỏa mãn hai điều kiện:

(1) $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$, tức là hợp tất cả các biến cố chính là không gian mẫu.

(2) Các biến cố đôi một xung khắc với nhau:

$$B_i \cap B_j = \emptyset, \forall i = \overline{1, n}, j = \overline{1, n}, i \neq j$$

Khi đó, với một biến cố A bất kỳ, xác suất của A được biểu diễn qua tổng các xác suất:

$$P(A) = \sum_{j=1}^n P(A \cap B_j) = \sum_{j=1}^n P(A|B_j) \times P(B_j)$$

Định lý Bayes về tính xác suất có điều kiện được biểu diễn theo công thức của Bolstad và Curran (2016) như sau:

$$P(B_i|A) = \frac{P(A|B_i) \times P(B_i)}{\sum_{j=1}^n P(A|B_j) \times P(B_j)}$$

Định lý Bayes chính là chìa khóa của thống kê Bayes, trong đó, $B_1, B_2, \dots, B_n$ được xem như các biến cố không quan sát được, khi đó, xác suất $P(B_i)$ là xác suất tiên nghiệm cho các biến cố $B_i, \forall i = \overline{1, n}$. Phân phối tiên nghiệm được dựa trên các xác suất tiên nghiệm có trọng số (Bolstad & Curran, 2016).

Hàm hợp lý của các biến cố không quan sát được $B_1, B_2, \dots, B_n$ được xác định bởi $P(A|B_i)$. Do đó, $P(B_i|A)$ là xác suất hậu nghiệm của biến cố B_i với điều kiện biến cố A đã xảy ra. Các biến cố không quan sát được $B_1, B_2, \dots, B_n$ trùng với các giá trị tham số θ , biến cố đã xảy ra A trùng với thông tin từ dữ liệu.

2.2. Thống kê Bayes

Giả sử hàm mật độ xác suất mẫu (hàm mật độ xác suất của dữ liệu) y là $p(y|\theta)$, hàm mật độ xác suất tiên nghiệm là $p(\theta)$. Khi đó, hàm mật độ xác suất hậu nghiệm được xác định theo công thức của Gelman và cộng sự (2013) như sau:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Trong đó, $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$ trong trường hợp θ tuân theo biến ngẫu nhiên rời rạc, hoặc $p(y) = \int_{\theta} p(\theta)p(y|\theta)d\theta$ trong trường hợp θ tuân theo biến ngẫu nhiên liên tục. Một cách tương đương, công thức về mối quan hệ giữa hàm mật độ xác suất hậu nghiệm, tiên nghiệm và hàm hợp lý được biểu diễn dưới dạng:

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

Giả sử mẫu ngẫu nhiên $y_1, y_2, \dots, y_n$ bao gồm các quan sát độc lập và cùng tuân theo một phân phối xác suất với hàm mật độ xác suất phụ thuộc vào tham số $f(y|\theta)$. Khi đó, hàm hợp lý của mẫu ngẫu nhiên có dạng:

$$p(y|\theta) = p(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n f(y_i|\theta)$$

Phân phối tiên nghiệm cho tham số θ bao gồm cả trường hợp θ tuân theo biến ngẫu nhiên rời rạc hoặc biến ngẫu nhiên liên tục. Thông thường, có ba cách lựa chọn phân phối tiên nghiệm: (1) Tiên nghiệm phi thông tin (tiên nghiệm đều), (2) tiên nghiệm liên hợp, và (3) tiên nghiệm dựa vào Maximum Entropy. Nếu dựa vào tiên nghiệm đều thì phân phối hậu nghiệm trùng với hàm hợp lý,

nghĩa là không đưa thêm thông tin ngoài dữ liệu vào các nghiên cứu được. Chính vì vậy, nhóm tác giả sẽ hạn chế lựa chọn phương pháp tiên nghiệm này. Ngoài ra, nếu lựa chọn dựa vào Maximum Entropy thì cần rất nhiều các tính toán phức tạp dựa vào Entropy. Do đó, để vừa đảm bảo cập nhật thêm thông tin ngoài dữ liệu, vừa không quá phức tạp trong tính toán, nhóm tác giả sử dụng tiên nghiệm liên hợp.

Tiên nghiệm liên hợp là các phân phối xác suất được chọn sao cho phân phối hậu nghiệm được tính toán từ công thức Bayes cũng có cùng dạng phân phối xác suất với phân phối tiên nghiệm. Một số tiên nghiệm liên hợp thường được chọn là trung bình của phân phối chuẩn, với giả định đã biết phương sai của tổng thể có tiên nghiệm liên hợp là phân phối chuẩn, xác suất thành công trong phân phối nhị thức có tiên nghiệm liên hợp là phân phối Beta, phương sai của phân phối chuẩn đã biết trung bình của tổng thể có tiên nghiệm liên hợp là phân phối Gamma ngược (Bolstad & Curran, 2016).

2.3. Các suy luận trong thống kê Bayes

Suy luận đầu tiên thường được quan tâm về các quan sát chưa biết chính là suy luận dự báo. Với thống kê Bayes, dự báo dựa vào hàm mật độ dự báo hậu nghiệm được xác định theo công thức của Gelman và cộng sự (2013) như sau:

$$p(\tilde{y}|y) = \int_{\theta} p(\tilde{y}, \theta|y) d\theta = \int_{\theta} p(\tilde{y}|\theta, y) p(\theta|y) d\theta = \int_{\theta} p(\tilde{y}|\theta) p(\theta|y) d\theta$$

Trong công thức trên, đẳng thức cuối cùng thể hiện hàm mật độ dự báo hậu nghiệm giống như là trung bình của hàm dự báo có điều kiện theo hàm phân phối hậu nghiệm của θ .

Giả sử mẫu ngẫu nhiên $\{y_1, y_2, \dots, y_n\}$, các quan sát trong đó cùng tuân theo phân phối chuẩn với trung bình mẫu $\bar{y}$. Khi đó, hàm hợp lý được xác định theo công thức của Bolstad và Curran (2016) như sau:

$$p(y_1, y_2, \dots, y_n | \theta = \mu) \propto \exp\left(-\frac{1}{\frac{2\sigma^2}{n}} (\bar{y} - \theta)^2\right)$$

Chúng ta nhận thấy hàm hợp lý có dạng phân phối chuẩn, với kỳ vọng $\theta = \mu$ và phương sai $\frac{\sigma^2}{n}$. Để đơn giản trong các tính toán, nhóm tác giả giả sử phân phối tiên nghiệm của tham số $\theta = \mu$ là phân phối dạng liên hợp, tức là phân phối của tham số $\theta = \mu$ cũng là phân phối chuẩn với kỳ vọng là m và phương sai là s^2 , tức là:

$$p(\theta) \propto \exp\left(-\frac{1}{2s^2} (\theta - m)^2\right)$$

Khi đó, phân phối hậu nghiệm $p(\theta|y_1, y_2, \dots, y_n)$ cũng tuân theo phân phối chuẩn với kỳ vọng m' và phương sai s'^2 được biểu diễn theo công thức của Bolstad và Curran (2016) như sau:

$$\begin{cases} m' = \frac{\frac{1}{s^2}}{\frac{n}{\sigma^2} + \frac{1}{s^2}} m + \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{s^2}} \bar{y} \\ \frac{1}{(s')^2} = \frac{1}{s^2} + \frac{n}{\sigma^2} \end{cases}$$

Khi đó, phân phối dự báo hậu nghiệm có dạng phân phối chuẩn với kỳ vọng m' và phương sai $(s'^2 + \sigma^2)$. Từ đây suy ra dự báo điểm cho quan sát kế tiếp y_{n+1} với tổn thất dạng bình phương chính

là trung bình của phân phối dự báo hậu nghiệm m' (Robert, 2007), dự báo khoảng cho quan sát kế tiếp dựa vào miền mật độ cao nhất (Le và cộng sự, 2020).

3. Ứng dụng vào dữ liệu

3.1. Mô tả dữ liệu

Nhóm tác giả thu thập dữ liệu của 25 ngân hàng thương mại và công ty tài chính trên thị trường chứng khoán Việt Nam của Thomson Reuters tại Trung tâm Nghiên cứu Kinh tế và Tài chính - Trường Đại học Kinh tế - Luật, Đại học Quốc gia TP.HCM từ ngày bắt đầu giao dịch đầu tiên của các mã chứng khoán đến ngày 17/03/2020 theo chi tiết mô tả tại Bảng 1.

Bảng 1.

Danh sách các mã chứng khoán ngân hàng và công ty tài chính

STT	Mã chứng khoán	Tên công ty	Ngày giao dịch đầu tiên
1	ABI	Công ty Cổ phần Bảo hiểm Ngân hàng Nông nghiệp (ABIC)	16/07/2009
2	ACB	Ngân hàng Thương mại Cổ phần Á Châu	21/11/2006
3	BAB	Ngân hàng Thương mại Cổ phần Bắc Á	02/01/2018
4	BIC	Tổng Công ty Cổ phần Bảo hiểm Ngân hàng Đầu tư và Phát triển Việt Nam	06/09/2011
5	BID	Ngân hàng Thương mại Cổ phần Đầu tư và Phát triển Việt Nam	24/01/2014
6	BMI	Tổng Công ty Cổ phần Bảo Minh	28/11/2006
7	BVH	Tập đoàn Bảo Việt	26/06/2009
8	CTG	Ngân hàng Thương mại Cổ phần Công Thương Việt Nam	16/07/2009
9	EIB	Ngân hàng Thương mại Cổ phần Xuất nhập khẩu Việt Nam	27/10/2009
10	HDB	Ngân hàng Thương mại Cổ phần Phát triển thành phố Hồ Chí Minh	05/01/2018
11	LPB	Ngân hàng Thương mại Cổ phần Bưu điện Liên Việt	05/10/2017
12	MBB	Ngân hàng Thương mại Cổ phần Quân đội	01/11/2011
13	MIG	Tổng Công ty Cổ phần Bảo hiểm Quân Đội	05/05/2017
14	NVB	Ngân hàng Thương mại Cổ phần Quốc Dân	13/09/2010
15	PGI	Tổng Công ty Cổ phần Bảo hiểm Petrolimex	29/07/2011
16	PTI	Tổng Công ty Cổ phần Bảo hiểm Bưu điện	02/03/2011
17	PVI	Công ty Cổ phần PVI	10/08/2007
18	SHB	Ngân hàng Thương mại Cổ phần Sài Gòn - Hà Nội	20/04/2009
19	STB	Ngân hàng Thương mại Cổ phần Sài Gòn Thương Tín	12/07/2006
20	TCB	Ngân hàng Thương mại Cổ phần Kỹ Thương Việt Nam	04/06/2018

STT	Mã chứng khoán	Tên công ty	Ngày giao dịch đầu tiên
21	TPB	Ngân hàng Thương mại Cổ phần Tiên Phong	19/04/2018
22	VCB	Ngân hàng Thương mại Cổ phần Ngoại thương Việt Nam	30/06/2009
23	VIB	Ngân hàng Thương mại Cổ phần Quốc tế Việt Nam	09/01/2017
24	VNR	Tổng công ty Cổ phần Bảo hiểm Quốc gia Việt Nam	12/03/2006
25	VPB	Ngân hàng Thương mại Cổ phần Việt Nam Thịnh Vượng	17/08/2017

Để biến đổi dữ liệu về dạng phân phối xác suất chuẩn, nhóm tác giả biến đổi giá gốc về tỷ suất sinh lợi theo công thức:

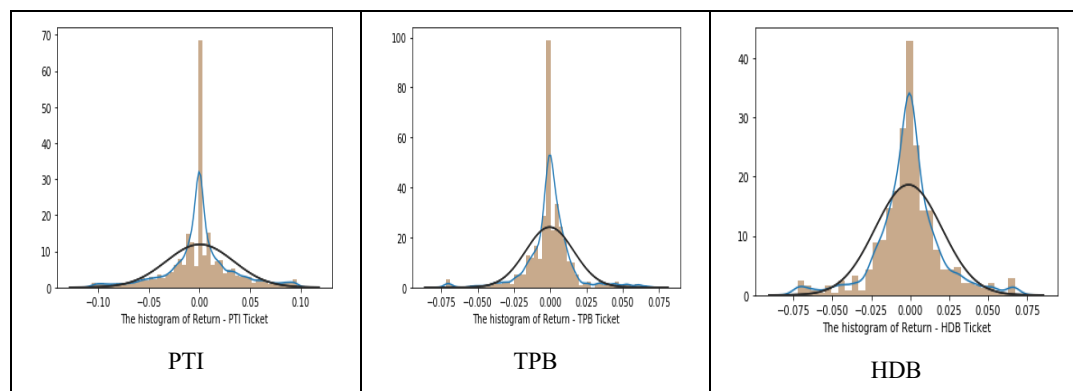
$$return_t = y_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$$

Trong đó:

p_t : Giá đóng cửa chứng khoán tại thời điểm ngày (t);

p_{t-1} : Giá đóng cửa tại thời điểm ngày trước đó, tức là ngày ($t - 1$).

Nhóm tác giả minh họa về dạng phân phối của suất sinh lợi của các mã chứng khoán trong ba trường hợp: Xấp xỉ tề nhất, xấp xỉ trung bình và xấp xỉ tốt nhất khi so sánh với phân phối chuẩn theo Hình 1.



Hình 1. Biểu đồ Histogram biểu diễn tỷ suất sinh lợi của ba mã chứng khoán đại diện là PTI, TPB và HDB

Đối với trường hợp xấp xỉ tề nhất theo phân phối chuẩn là mã chứng khoán PTI có 1.869 quan sát với tỷ suất sinh lợi trung bình là 0,00045, thấp nhất là −0,106322, cao nhất là 0,096264 theo Hình 1. Hình 1 cho thấy mặc dù dữ liệu đã dao động quanh giá trị 0, tuy nhiên, giá trị 0 có tần suất khá cao, vượt xa các giá trị khác. Đối với trường hợp xấp xỉ trung bình theo phân phối chuẩn, với đại diện là mã chứng khoán TPB có 1.869 quan sát với tỷ suất sinh lợi trung bình là −0,000394, thấp nhất là −0,072176, cao nhất là 0,066842. Đối với trường hợp xấp xỉ tốt nhất theo phân phối chuẩn, với đại diện là mã chứng khoán HDB có 543 quan sát, trong đó: Suất sinh lợi trung bình là −0,001084, cao nhất là 0,067325 và thấp nhất là −0,072502. Phân phối xác suất của suất sinh lợi khá gần với phân phối chuẩn.

Từ biểu đồ Histogram của 3 mã chứng khoán đại diện, nhóm tác giả nhận thấy khi biến đổi giá chứng khoán gốc về suất sinh lợi, phân phối xác suất của suất sinh lợi khá gần với phân phối chuẩn. Bên cạnh đó, một số kiểm định dữ liệu tuân theo phân phối chuẩn có thể sử dụng như: Kiểm định χ^2 về dạng phân phối xác suất, kiểm định Jarque Bera... để xem xét dữ liệu có tuân theo phân phối chuẩn hay không. Tuy nhiên, các kiểm định dữ liệu có tuân theo phân phối chuẩn hay không đều dựa vào giá trị p-value; giá trị này đang gây rất nhiều tranh cãi trong các kết luận của bài toán kiểm định giả thuyết do cần phải bổ sung thông tin mới quyết định được chấp nhận hay bác bỏ giả thuyết H_0 . (Tran và cộng sự, 2017). Tất nhiên, chúng ta có thể xem xét đến các phương pháp lựa chọn dạng phân phối xác suất khác của dữ liệu thông qua lựa chọn mô hình như: Sử dụng giá trị AIC cực tiểu, BIC cực tiểu, Relative Entropy cực đại... Trong khuôn khổ bài báo này, nhóm tác giả không đặt nặng lựa chọn dạng phân phối xác suất chính xác nhất của dữ liệu, mà nhóm tác giả chỉ xem xét các dạng phân phối xác suất có thể xấp xỉ được của dữ liệu.

Chính vì vậy, trong bài báo này, nhóm tác giả sử dụng các đặc trưng Moment của phân phối chuẩn để xem xét bộ dữ liệu có xấp xỉ phân phối chuẩn hay không. Cụ thể, nhóm tác giả xem xét các chỉ số độ lệch Skewness = 0 và độ nhọn Kurtosis = 3. Các chỉ số độ lệch và độ nhọn của các bộ dữ liệu suất sinh lợi được biểu diễn theo Bảng 2 dưới đây.

Bảng 2.

Các chỉ số độ lệch Skewness và độ nhọn Kurtosis của suất sinh lợi toàn bộ dữ liệu, dữ liệu đã chia với 25 mã chứng khoán

STT	Mã chứng khoán	Toàn bộ dữ liệu		Dữ liệu đã chia 1, bao gồm các quan sát ở vị trí $\{1, 4, \dots, t+1, t+4, \dots\}$		Dữ liệu đã chia 2, bao gồm các quan sát ở vị trí $\{2, 5, \dots, t+2, t+5, \dots\}$		Dữ liệu đã chia 3, bao gồm các quan sát ở vị trí $\{3, 6, \dots, t, t+3, \dots\}$	
		Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
1	ABI	-0,1463	0,8801	0,0280	2,8076	-0,1704	2,7488	-0,1745	2,7069
2	ACB	0,0537	4,2213	0,1650	4,5117	0,1130	3,9208	-0,0954	4,8920
3	BAB	-4,6713	66,6986	-1,6184	27,5629	-4,5651	37,9547	-4,1920	41,9051
4	BIC	0,1357	0,9747	0,6461	2,5570	0,6449	2,6042	0,5460	2,9905
5	BID	0,0491	1,7898	-0,0550	3,4498	0,0644	3,2902	0,0895	2,5922
6	BMI	-0,0529	3,7761	0,3649	3,7013	0,3209	3,2224	0,3311	4,7761
7	BVH	0,0241	0,1687	0,0402	1,5905	-0,0136	2,0243	0,0856	1,7095
8	CTG	0,1446	1,6878	0,2245	2,2443	0,0496	2,1573	0,1617	1,9251
9	EIB	0,2720	3,0786	0,1349	4,4595	0,4066	5,1950	0,4791	3,7042
10	HDB	-0,1135	2,4676	0,1697	2,4793	-0,0876	2,7336	-0,3796	3,2695
11	LPB	-0,1316	6,9929	0,2827	3,1913	0,3541	9,0920	0,2630	2,3916
12	MBB	0,0570	2,9188	-0,0244	4,6426	-0,1094	3,2500	0,0985	3,1297
13	MIG	0,0888	5,1289	-0,0496	4,9381	-0,3950	2,5484	-0,0064	3,2764
14	NVB	0,0960	1,5877	0,5056	4,7162	0,5165	5,8549	0,7287	5,6511
15	PGI	-0,0151	0,3856	0,2186	2,1436	0,1262	1,2347	0,0129	1,8155

STT	Mã chứng khoán	Toàn bộ dữ liệu		Dữ liệu đã chia 1, bao gồm các quan sát ở vị trí $\{1, 4, \dots, t+1, t+4, \dots\}$		Dữ liệu đã chia 2, bao gồm các quan sát ở vị trí $\{2, 5, \dots, t+2, t+5, \dots\}$		Dữ liệu đã chia 3, bao gồm các quan sát ở vị trí $\{3, 6, \dots, t, t+3, \dots\}$	
		Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
16	PTI	-0,0747	1,5902	0,1465	2,8752	0,4645	3,6585	0,2189	2,6640
17	PVI	0,1411	2,2565	0,0782	3,8773	0,1148	3,9441	0,2909	4,3392
18	SHB	0,2282	2,3606	0,6941	4,6064	0,7153	3,9914	0,6974	5,0622
19	STB	0,1714	0,8764	0,1923	2,0223	0,4487	2,4724	0,1438	1,9122
20	TCB	-0,5647	3,1650	-0,9524	2,5390	-0,6130	2,0980	-0,0050	3,3688
21	TPB	-0,1834	5,2991	-0,4878	1,6787	0,0600	1,7723	0,0004	1,9156
22	VCB	0,0435	1,2995	0,2764	1,6945	-0,0891	2,1661	0,2200	1,4615
23	VIB	-0,6121	6,6010	0,1875	5,1226	-0,3668	5,6573	-0,1577	3,9244
24	VNR	-0,1229	1,1728	0,1467	3,1692	-0,0273	2,9050	0,1667	3,6144
25	VPB	-0,1400	1,8977	-0,2136	1,9458	-0,2450	2,1522	-0,1566	1,3418

Thông qua các chỉ số về độ lệch Skewness xấp xỉ giá trị 0 và độ nhọn Kurtosis xấp xỉ giá trị 3, đặc biệt là các bộ dữ liệu đã chia nhằm xấp xỉ phân phối chuẩn trong kết quả nghiên cứu của nhóm tác giả. Do đó, các bộ dữ liệu suất sinh lợi của dữ liệu giá chứng khoán có thể xấp xỉ phân phối chuẩn.

Dựa vào các phân tích về dạng phân phối xác suất, nhóm tác giả áp dụng thống kê Bayes cho bài toán dự báo đối với dữ liệu tuân theo phân phối chuẩn; từ đây, dự báo được suất sinh lợi, nhân ngược trở lại suất sinh lợi dự báo với giá chứng khoán ở thời điểm (t) sẽ có giá dự báo ở thời điểm tiếp theo.

3.2. Đề xuất mô hình dự báo

Tại thời điểm t , nhóm tác giả có bộ dữ liệu bao gồm các quan sát $return_1 = y_1, return_2 = y_2, \dots, return_t = y_t$. Theo thống kê Bayes đã trình bày được trong mục 2.2, nhóm tác giả sẽ tính được phân phối dự báo hậu nghiệm $N(m', s'^2 + \sigma^2)$. Do đó, các dự báo suất sinh lợi cho quan sát kế tiếp y_{t+1} bao gồm: Dự báo điểm $y_{t+1} = m'$, dự báo khoảng dựa vào miền mật độ xác suất cao nhất với xác suất $(1 - \alpha)$ là $(m' \pm z_{\frac{\alpha}{2}} \sqrt{s'^2 + \sigma^2})$. Từ đây suy ra, các dự báo giá chứng khoán cho quan sát kế tiếp bao gồm dự báo điểm $p_{t+1} = m'p_t$ và dự báo khoảng với xác suất bằng $(1 - \alpha)$ là $(m' \pm z_{\frac{\alpha}{2}} \sqrt{s'^2 + \sigma^2}) p_t$.

Dựa vào bản chất của thị trường chứng khoán Việt Nam với giao dịch $(t + 3)$ và kết quả thực nghiệm của chính nhóm nghiên cứu, nhóm tác giả đề xuất tách bộ dữ liệu gốc ra làm ba bộ dữ liệu, trong đó, mỗi bộ dữ liệu cách nhau ba đơn vị thời gian, tức là ba bộ dữ liệu bao gồm: $data_1 = \{1, 4, \dots, t+1, t+4, \dots\}, data_2 = \{2, 5, \dots, t+2, t+5, \dots\}, data_3 = \{3, 6, \dots, t+3, t+6, \dots\}$.

Khi đó, tại thời điểm s bất kỳ, nhóm tác giả sẽ dự báo được giá trị tại thời điểm $(s + 1), (s + 2), (s + 3)$ theo ba bộ dữ liệu trên. Thật vậy, theo công thức thống kê Bayes nêu trên,

tại thời điểm s bất kỳ, nhóm tác giả sẽ dự báo được thời điểm $(s + 3)$, tương ứng với một trong ba bộ dữ liệu $data_1, data_2, data_3$ ở trên. Do đó, đã có giá trị tại thời điểm $(s - 1)$ nên dự báo được giá trị tại thời điểm $(s + 2)$. Tương tự, đã có giá trị tại thời điểm $(s - 2)$ nên dự báo được giá trị tại thời điểm $(s + 1)$. Như vậy, tại thời điểm t , nhóm tác giả sẽ dự báo được đến thời điểm $(t + 3)$.

Tất nhiên, cũng có thể áp dụng đơn thuần thống kê Bayes với thời điểm t dự báo $(t + 1)$, sau đó lấy giá trị dự báo $(t + 1)$ dự báo tiếp $(t + 2)$, lấy giá trị $(t + 2)$ dự báo $(t + 3)$. Nhóm nghiên cứu cũng đã kiểm tra kết quả thực nghiệm trong các trường hợp như vậy, sau mỗi lần dự báo của dự báo thì tỷ lệ đúng giảm khá nhiều.

Thông thường, để đạt được hiệu quả trong dự báo, nhóm tác giả sử dụng dự báo khoảng, trong đó, các dự báo theo các phân tích học thuật với xác suất các khoảng là 68%, 90% và 95%. Thông thường, trong thống kê, chúng ta thường quan tâm các khoảng tập trung xung quanh giá trị trung bình của tổng thể μ là $\mu \pm 1\sigma$ (tương ứng xác suất 68%), $\mu \pm 2\sigma$ (tương ứng xác suất 95%) và $\mu \pm 3\sigma$ (tương ứng xác suất 99%). Tuy nhiên, càng mở rộng khoảng dự báo thì xác suất đúng càng cao nên nhóm tác giả nghiên cứu thêm khoảng dự báo 90%, cũng là một trong những khoảng tin cậy thường được quan tâm. Hơn nữa, do hạn chế về số lượng bảng trình bày đầy đủ nhất các giá trị để so sánh nên nhóm tác giả chỉ chọn ba mức xác suất để nghiên cứu là 68%, 90% và 95%. Chi tiết các kết quả dự báo đúng được biểu diễn qua Bảng 3 dưới đây.

Bảng 3.

Tỷ lệ dự báo đúng tới thời điểm $(t + 3)$ nếu sử dụng mô hình thống kê Bayes với thời điểm t dự báo $(t + 1)$, sử dụng $(t + 1)$ đã dự báo nhằm dự báo $(t + 2)$, sử dụng $(t + 2)$ đã dự báo nhằm dự báo $(t + 3)$ với khoảng dự báo xác suất 68%, 90%, 95%.

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng dự báo xác suất 68%			Khoảng dự báo xác suất 90%			Khoảng dự báo xác suất 95%		
			$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$
1	ABI	1.975	77,1%	73,4%	68,9%	96,1%	90,5%	89,2%	97,6%	93,7%	92,2%
2	ACB	3.229	85,2%	74,7%	68,6%	94,6%	88,2%	83,2%	96,8%	91,6%	87,1%
3	BAB	459	96,7%	95,0%	95,4%	99,1%	97,6%	97,2%	99,1%	98,7%	97,8%
4	BIC	2.000	74,4%	64,3%	57,1%	89,7%	82,7%	77,5%	93,3%	87,3%	82,1%
5	BID	1.437	73,6%	60,5%	52,5%	88,9%	80,1%	71,5%	91,9%	83,9%	76,8%
6	BMI	3.206	80,5%	71,9%	64,1%	96,0%	87,6%	83,4%	97,4%	90,7%	87,7%
7	BVH	2.586	73,4%	63,9%	55,6%	93,1%	81,8%	75,1%	96,4%	85,5%	80,3%
8	CTG	2.571	77,4%	64,5%	55,4%	89,8%	81,9%	74,6%	92,7%	86,0%	80,1%
9	EIB	2.499	76,1%	65,3%	56,6%	90,2%	83,0%	77,4%	92,2%	85,9%	81,8%
10	HDB	454	84,1%	71,1%	62,3%	94,1%	87,7%	81,5%	95,8%	90,3%	86,8%
11	LPB	521	82,5%	71,2%	60,7%	94,0%	87,5%	81,0%	96,4%	90,2%	85,8%
12	MBB	1.997	79,6%	68,4%	60,0%	91,3%	83,4%	77,2%	93,7%	87,5%	82,1%
13	MIG	610	83,3%	76,4%	72,3%	94,3%	91,1%	88,0%	95,6%	93,6%	90,8%

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng dự báo xác suất 68%			Khoảng dự báo xác suất 90%			Khoảng dự báo xác suất 95%		
			$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$
14	NVB	2.105	76,4%	70,7%	66,9%	87,3%	84,9%	83,0%	90,8%	88,5%	86,7%
15	PGI	1.870	72,6%	64,4%	59,4%	88,8%	83,1%	77,5%	93,3%	88,5%	85,3%
16	PTI	1.779	72,1%	63,5%	57,4%	85,7%	80,0%	74,6%	88,5%	83,4%	79,5%
17	PVI	3.042	82,5%	72,9%	66,5%	94,2%	87,6%	82,4%	96,9%	90,6%	86,5%
18	SHB	2.634	80,1%	69,7%	61,0%	94,0%	88,1%	81,4%	96,1%	91,0%	85,5%
19	STB	3.317	76,1%	64,4%	57,0%	90,4%	81,7%	75,8%	95,8%	85,7%	79,9%
20	TCB	358	85,8%	69,8%	57,5%	96,6%	90,8%	83,2%	98,3%	94,4%	89,7%
21	TPB	387	87,9%	75,2%	67,7%	93,5%	87,6%	84,2%	94,8%	89,9%	85,5%
22	VCB	2.581	76,4%	63,7%	55,0%	91,0%	82,3%	75,6%	93,8%	86,0%	80,7%
23	VIB	704	84,7%	71,7%	62,8%	94,3%	87,5%	81,1%	95,7%	89,5%	83,9%
24	VNR	2.986	77,0%	68,1%	62,3%	91,4%	84,8%	80,2%	95,2%	89,1%	84,8%
25	VPB	553	78,7%	65,3%	59,0%	90,8%	82,6%	75,8%	92,8%	86,6%	79,2%
Trung bình của xác suất dự báo đúng			79,8%	69,6%	62,5%	92,4%	85,8%	80,5%	94,8%	89,1%	84,7%
Độ lệch chuẩn			5,7%	6,8%	8,6%	3,2%	4,1%	5,6%	2,5%	3,6%	4,8%

Kết quả Bảng 3 cho thấy với khoảng dự báo xác suất 68%, nếu sử dụng giá tại thời điểm t nhằm dự báo khoảng cho thời điểm $(t + 1)$ với tỷ lệ dự báo đúng là 79,8%; giả định sử dụng giá tại thời điểm $(t + 1)$ đã dự báo nhằm dự báo khoảng $(t + 2)$ thì tỷ lệ dự báo giảm còn 69,6%; tương tự, sử dụng giá tại thời điểm $(t + 2)$ đã dự báo nhằm dự báo khoảng $(t + 3)$ thì tỷ lệ dự báo đúng giảm xuống còn 62,5%. Các kết quả tương tự đối với dự báo khoảng xác suất 90% thì tỷ lệ dự báo đúng giảm từ tỷ lệ dự báo đúng $(t + 1)$ là 92,4% xuống còn tỷ lệ dự báo đúng ở $(t + 3)$ là 80,5%. Kết quả hoàn toàn tương tự đối với khoảng dự báo xác suất 95% từ 94,8% xuống còn 84,7%.

Bên cạnh đó, nhóm tác giả cũng sử dụng các miền biến động của khoảng dự báo, dựa vào đặc trưng của thị trường chứng khoán Việt Nam biến động tùy vào từng giai đoạn bao gồm $\pm 5\%$ (tương ứng miền biến động 10%), $\pm 7\%$ (tương ứng miền biến động 14%), $\pm 7,5\%$ (tương ứng miền biến động 15%) và $\pm 10\%$ (tương ứng miền biến động 20%). Do đó, nhóm tác giả dự báo miền biến động thu hẹp hơn so với miền biến động thực tiễn của thị trường chứng khoán Việt Nam nhằm cung cấp miền nhỏ nhất có thể và độ chính xác cao. Chi tiết các kết quả về tỷ lệ dự báo đúng với miền biến động thu hẹp được trình bày trong Bảng 4 như sau:

Bảng 4.

Tỷ lệ dự báo đúng tới thời điểm $(t + 3)$ nếu sử dụng mô hình thống kê Bayes với thời điểm t dự báo $(t + 1)$, sử dụng $(t + 1)$ đã dự báo nhằm dự báo $(t + 2)$, sử dụng $(t + 2)$ đã dự báo nhằm dự báo $(t + 3)$ với khoảng biến động $\pm 2,5\%$, $\pm 3,5\%$, $\pm 5\%$.

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng biến động $\pm 2,5\%$			Khoảng biến động $\pm 3,5\%$			Khoảng biến động $\pm 5\%$		
			$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$
1	ABI	1.975	46,0%	43,4%	39,2%	55,7%	52,0%	48,4%	67,0%	64,3%	59,0%
2	ACB	3.229	82,1%	72,6%	66,2%	89,7%	81,3%	76,0%	94,9%	89,0%	84,9%
3	BAB	459	96,1%	94,1%	94,6%	97,6%	95,9%	95,6%	98,0%	96,7%	96,3%
4	BIC	2.000	72,8%	62,6%	55,4%	83,4%	75,5%	69,6%	93,0%	86,9%	81,7%
5	BID	1.437	80,6%	67,6%	60,4%	88,4%	79,4%	72,2%	93,7%	88,7%	83,2%
6	BMI	3.206	69,7%	59,9%	51,9%	80,2%	71,4%	64,2%	93,7%	82,6%	77,0%
7	BVH	2.586	68,4%	58,5%	51,2%	78,8%	70,6%	62,7%	95,3%	82,7%	76,6%
8	CTG	2.571	81,7%	70,5%	62,7%	89,1%	81,2%	74,4%	97,2%	90,5%	86,3%
9	EIB	2.499	87,5%	80,0%	73,7%	93,1%	87,7%	83,8%	97,7%	93,8%	91,0%
10	HDB	454	82,8%	69,4%	60,6%	88,8%	78,6%	71,1%	93,6%	88,5%	84,4%
11	LPB	521	81,6%	71,2%	60,7%	90,4%	82,0%	74,9%	96,0%	89,8%	84,6%
12	MBB	1.997	88,9%	80,3%	73,5%	94,4%	88,8%	83,9%	98,2%	95,1%	92,1%
13	MIG	610	80,7%	73,6%	69,0%	87,2%	80,7%	77,9%	94,1%	91,1%	87,7%
14	NVB	2.105	69,6%	62,2%	57,2%	78,3%	73,3%	69,8%	85,5%	83,5%	80,6%
15	PGI	1.870	70,5%	61,7%	56,8%	79,4%	73,0%	68,1%	92,2%	86,7%	82,9%
16	PTI	1.779	68,3%	59,1%	52,3%	76,3%	69,3%	63,8%	85,6%	80,9%	75,9%
17	PVI	3.042	75,7%	65,8%	59,4%	84,6%	75,9%	70,3%	91,8%	84,7%	79,9%
18	SHB	2.634	76,5%	65,6%	56,8%	86,0%	78,7%	71,3%	93,4%	88,3%	82,5%
19	STB	3.317	77,3%	65,9%	58,8%	86,6%	77,1%	71,9%	97,5%	86,8%	81,6%
20	TCB	358	86,9%	73,7%	63,4%	93,0%	87,4%	76,8%	97,5%	93,0%	88,5%
21	TPB	387	90,2%	81,1%	74,9%	92,8%	87,3%	84,5%	96,1%	92,5%	90,7%
22	VCB	2.581	81,2%	69,4%	61,0%	89,8%	80,6%	74,2%	97,5%	90,7%	85,9%
23	VIB	704	84,9%	72,4%	62,9%	91,8%	83,2%	74,9%	95,5%	90,5%	85,1%
24	VNR	2.986	64,3%	55,1%	49,6%	73,8%	65,1%	59,7%	84,1%	76,9%	71,6%
25	VPB	553	81,4%	68,9%	60,8%	88,1%	79,4%	73,1%	93,9%	87,2%	81,6%

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng biến động $\pm 2,5\%$			Khoảng biến động $\pm 3,5\%$			Khoảng biến động $\pm 5\%$		
			$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$	$t + 1$	$t + 2$	$t + 3$
Trung bình của xác suất dự báo đúng			77,8%	68,2%	61,3%	85,5%	78,2%	72,5%	92,9%	87,3%	82,9%
Độ lệch chuẩn			10,3%	10,0%	10,6%	8,7%	8,8%	9,2%	6,%	6,6%	7,4%

Kết quả Bảng 4 cho thấy tỷ lệ dự báo đúng với khoảng biến động $\pm 2,5\%$ giảm từ 77,8% (nếu dự báo $t + 1$) xuống còn 61,3% (nếu dự báo đến $t + 3$). Kết quả hoàn toàn tương tự, với khoảng biến động $\pm 3,5\%$ giảm từ 85,5% xuống còn 72,5%, với khoảng biến động $\pm 5\%$ giảm từ 92,9% xuống còn 82,9%.

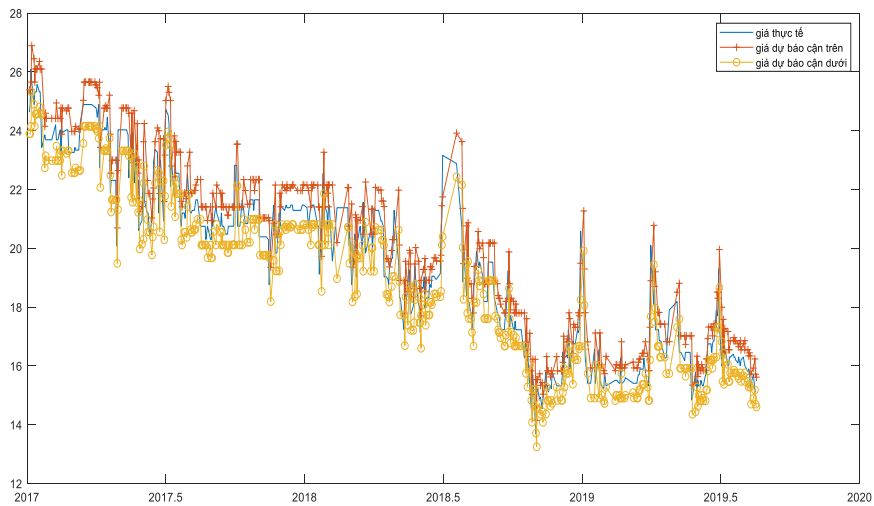
Về cách xác định tỷ lệ dự báo đúng, nhóm tác giả đánh giá một kết quả dự báo được tính là dự báo đúng nếu giá thực tế thuộc vào khoảng dự báo, ngược lại, kết quả dự báo sai.

3.3. Kết quả nghiên cứu

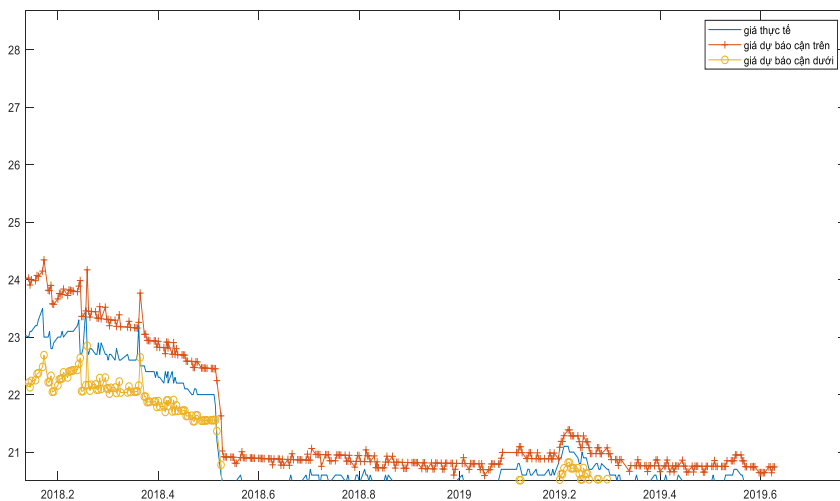
Để so sánh giữa hai phương pháp dự báo, nhóm tác giả phân tích thực nghiệm kết quả dự báo của 2 mô hình như sau:

- **MHI:** Dự báo giá chứng khoán của ngày hôm sau ($t + 1$):

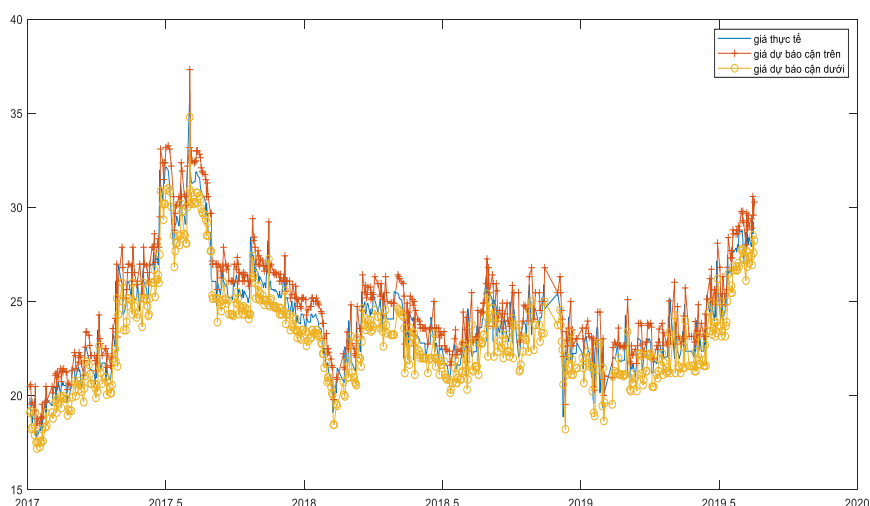
Bắt đầu tại thời điểm t , nhóm tác giả sẽ tính toán suất sinh lợi của giá chứng khoán. Từ đây, ước lượng các tham số trung bình và độ lệch chuẩn khi xấp xỉ suất sinh lợi theo phân phối chuẩn, sau đó tìm ra được hàm hợp lý của dữ liệu suất sinh lợi. Lựa chọn phân phối tiên nghiệm cho trung bình của tổng thể dạng tiên nghiệm liên hợp. Áp dụng công thức Bayes, tính ra phân phối hậu nghiệm cho trung bình của phân phối chuẩn và phân phối dự báo hậu nghiệm. Phân phối dự báo hậu nghiệm cũng là phân phối chuẩn, có trung bình và độ lệch chuẩn như đã tính trong mục 2. Từ đây, dự báo khoảng được suất sinh lợi tại thời điểm ($t + 1$), nhân với giá chứng khoán tại thời điểm t thì được giá dự báo tại thời điểm ($t + 1$). Chi tiết các kết quả giá trị dự báo tại thời điểm ($t + 1$) của các mã chứng khoán có tỷ lệ dự báo đúng cao nhất và thấp nhất được nhóm tác giả minh họa qua các trường hợp dựa vào dự báo theo xác suất và theo miền biến động thông qua các hình, từ Hình 2 đến Hình 5.



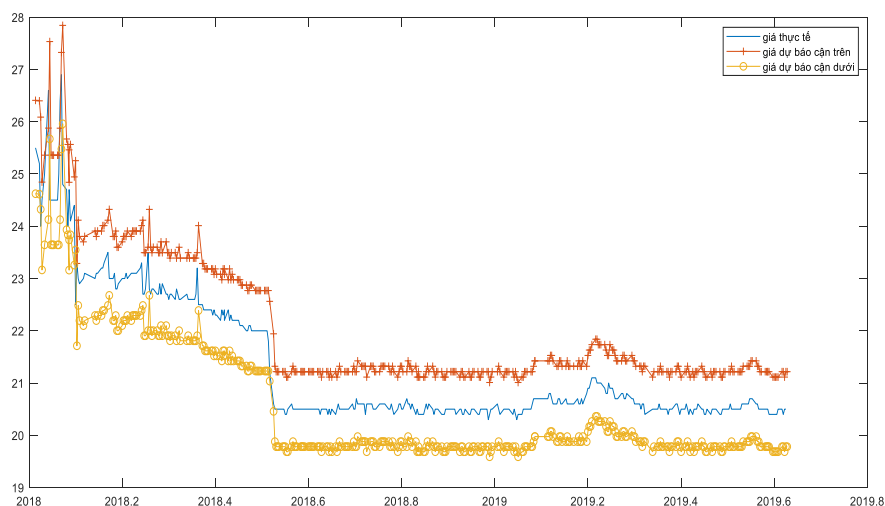
Hình 2. Đồ thị về dự báo mã cổ phiếu PTI cho thời điểm $(t + 1)$ với khoảng dự báo xác suất 68%, có tỷ lệ dự báo đúng thấp là 72,1%.



Hình 3. Dự báo mã cổ phiếu BAB cho thời điểm $(t + 1)$ với khoảng dự báo xác suất 68%, có tỷ lệ dự báo đúng cao nhất là 96,7%.



Hình 4. Dự báo mã cổ phiếu ABI cho thời điểm $(t + 1)$ với khoảng dự báo miền biến động $\pm 3,5\%$ (tương ứng miền biến động 7%), có tỷ lệ dự báo đúng thấp nhất là 55,7%.

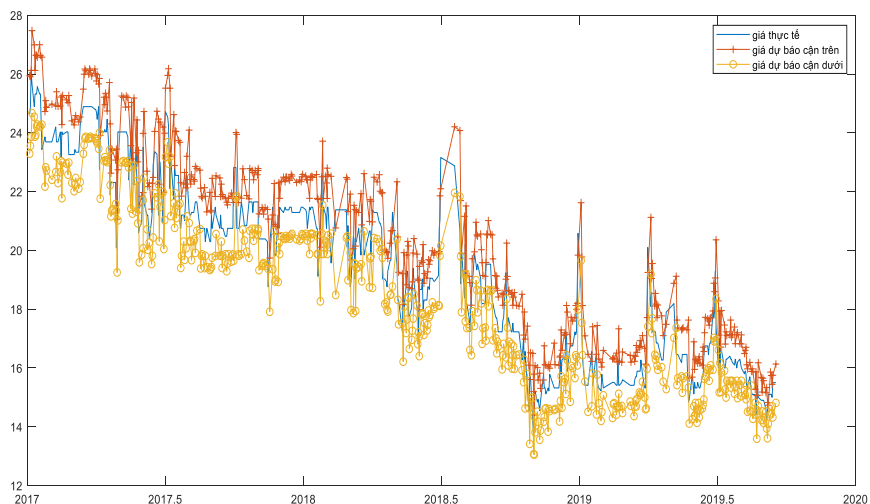


Hình 5. Dự báo mã cổ phiếu BAB cho thời điểm $(t + 1)$ với khoảng dự báo miền biến động $\pm 3,5\%$ (tương ứng miền biến động 7%), có tỷ lệ dự báo đúng cao nhất là 97,6%.

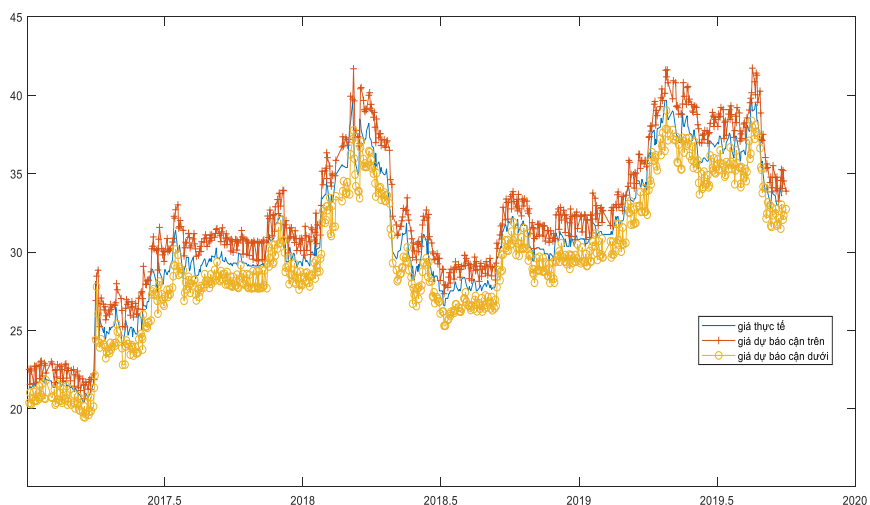
- **MH2:** MH đề xuất dự báo giá chứng khoán của 3 ngày tiếp theo $(t + 3)$

Bắt đầu từ thời điểm t , nhóm tác giả chia bộ dữ liệu ban đầu thành ba bộ dữ liệu con, mỗi bộ có đơn vị thời gian cách nhau ba ngày, tức là bộ dữ liệu có dạng $\{t, t + 3, t + 6, \dots\} = \{t, t + 3 \times 1, t + 3 \times 2, \dots\} = \{s, s + 1, s + 2, \dots\}$. Khi đó, áp dụng các kết quả tính toán như MH1 cho bộ dữ liệu $\{s, s + 1, s + 2, \dots\}$. Từ đây, tại thời điểm s dự báo được $(s + 1)$, tương ứng với tại thời điểm t dự

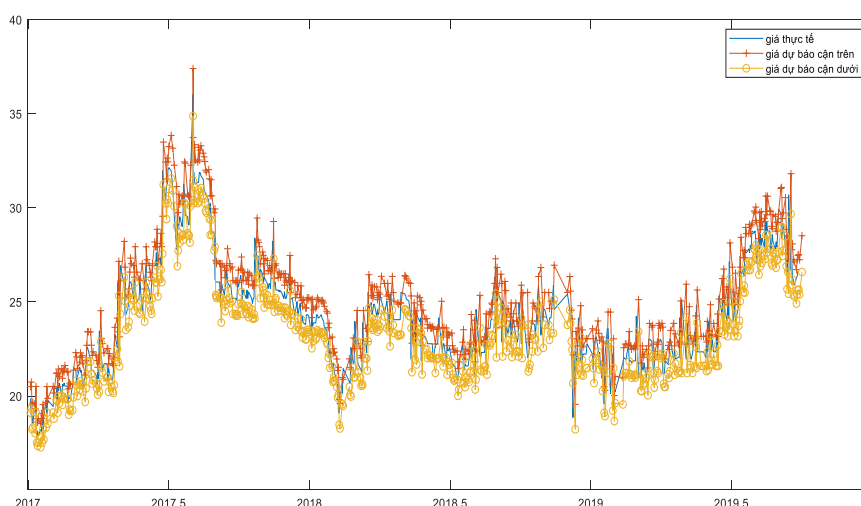
báo được ($t + 3$). Chi tiết các kết quả giá trị dự báo tại thời điểm ($t + 3$) của các mã chứng khoán có tỷ lệ dự báo đúng cao nhất và thấp nhất được nhóm tác giả minh họa qua các trường hợp dựa vào dự báo theo xác suất và theo miền biến động thông qua các hình, từ Hình 6 đến Hình 9.



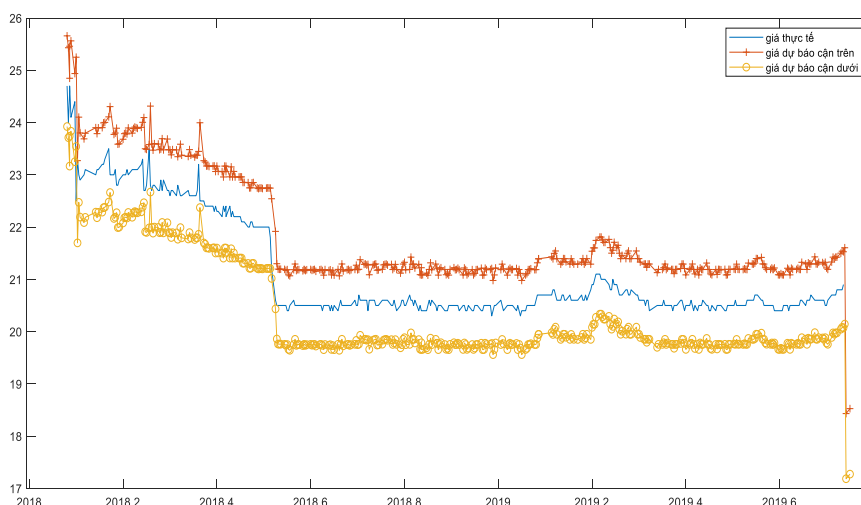
Hình 6. Dự báo mã cổ phiếu PTI cho thời điểm ($t + 3$) với khoảng dự báo xác suất 68%, có tỷ lệ dự báo đúng thấp nhất là 80,9%.



Hình 7. Dự báo mã cổ phiếu PVI cho thời điểm ($t + 3$) với khoảng dự báo xác suất 68%, có tỷ lệ dự báo đúng cao nhất là 93,3%.



Hình 8. Dự báo mã cổ phiếu ABI cho thời điểm $(t + 3)$ với khoảng dự báo miền biến động $\pm 3,5\%$ (tương ứng miền 7%), có tỷ lệ dự báo đúng thấp nhất là 55,9%.



Hình 9. Dự báo mã cổ phiếu BAB cho thời điểm $(t + 3)$ với khoảng dự báo miền biến động $\pm 3,5\%$ (tương ứng miền 7%), có tỷ lệ dự báo đúng cao nhất là 96,08%.

Như kết quả đã phân tích ở trên, nếu sử dụng MH1, ở thời điểm t dự báo thời điểm $(t + 1)$, sau đó sử dụng $(t + 1)$ đã dự báo nhằm dự báo $(t + 2)$, tiếp tục sử dụng $(t + 2)$ đã dự báo nhằm dự báo $(t + 3)$ thì tỷ lệ đúng giảm đi rất nhiều. Do đó, để đánh giá tính hiệu quả của MH1 – mô hình thông thường và MH2 – mô hình đề xuất, nhóm tác giả sẽ lần lượt so sánh các tỷ lệ dự báo đúng của hai mô

hình: Xuất phát từ thời điểm t , MH1 dự báo đến thời điểm $(t + 1)$, còn MH2 dự báo đến thời điểm $(t + 3)$.

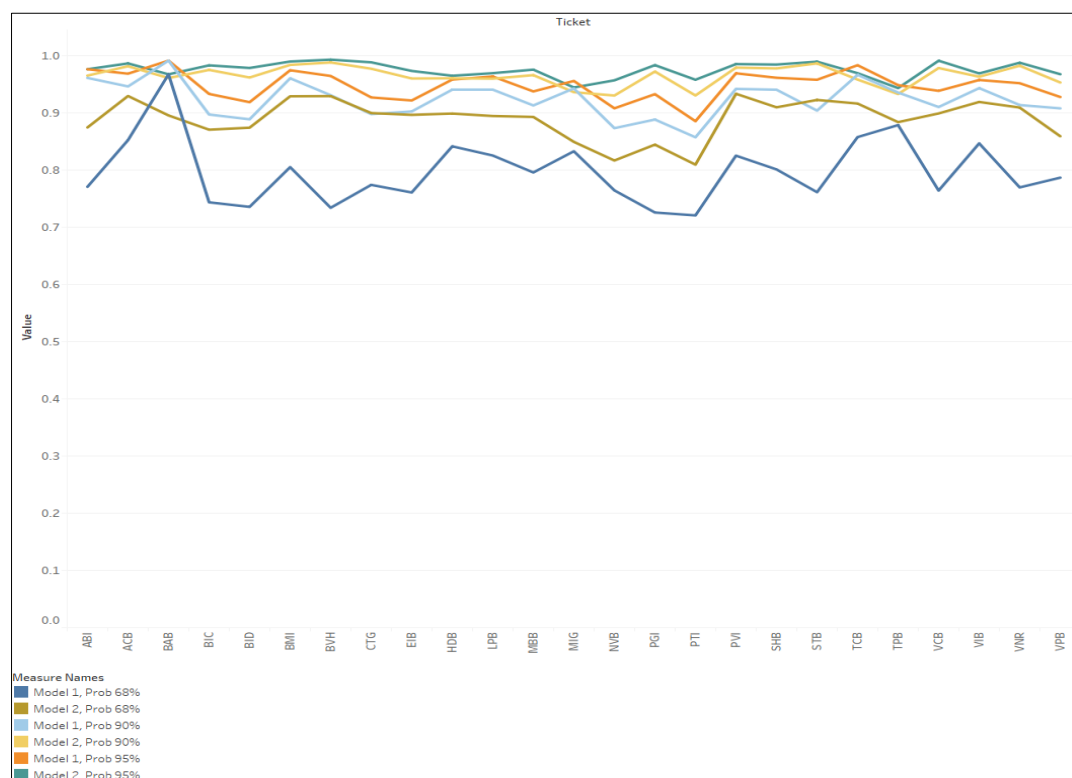
Trường hợp đầu tiên, nhóm tác giả so sánh độ chính xác của kết quả dự báo dựa vào hai mô hình 1 và MH2 với các xác suất khoảng dự báo lần lượt là 68%, 90% và 95% thông qua kết quả chi tiết ở Bảng 5 và Hình 10.

Bảng 5.

Tỷ lệ dự báo đúng đến thời điểm $(t + 1)$ đối với hai MH1, thời điểm $(t + 3)$ đối với MH2, các dự báo tương ứng là dự báo khoảng với các xác suất lần lượt là 68%, 90% và 95%

STT	Mã chứng khoán	Tổng số quan sát dự báo	Xác suất 68%			Xác suất 90%			Xác suất 95%		
			MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1
1	ABI	1.975	77,1%	87,4%	10,4%	96,1%	96,5%	0,4%	97,6%	97,6%	0,0%
2	ACB	3.229	85,2%	92,9%	7,7%	94,6%	98,1%	3,5%	96,8%	98,6%	1,8%
3	BAB	459	96,7%	89,5%	-7,2%	99,1%	96,1%	-3,0%	99,1%	96,7%	-2,4%
4	BIC	2.000	74,4%	87,1%	12,7%	89,7%	97,5%	7,8%	93,3%	98,3%	5,0%
5	BID	1.437	73,6%	87,4%	13,8%	88,9%	96,2%	7,3%	91,9%	97,8%	6,0%
6	BMI	3.206	80,5%	92,9%	12,4%	96,0%	98,4%	2,3%	97,4%	99,0%	1,5%
7	BVH	2.586	73,4%	92,9%	19,5%	93,1%	98,8%	5,7%	96,4%	99,3%	2,9%
8	CTG	2.571	77,4%	90,0%	12,6%	89,8%	97,7%	7,9%	92,7%	98,8%	6,1%
9	EIB	2.499	76,1%	89,6%	13,6%	90,2%	96,0%	5,8%	92,2%	97,3%	5,2%
10	HDB	454	84,1%	89,9%	5,7%	94,1%	96,0%	2,0%	95,8%	96,5%	0,7%
11	LPB	521	82,5%	89,4%	6,9%	94,0%	96,0%	1,9%	96,4%	96,9%	0,6%
12	MBB	1.997	79,6%	89,3%	9,7%	91,3%	96,6%	5,3%	93,7%	97,6%	3,8%
13	MIG	610	83,3%	84,9%	1,6%	94,3%	93,6%	-0,7%	95,6%	94,4%	-1,1%
14	NVB	2.105	76,4%	81,7%	5,2%	87,3%	93,0%	5,7%	90,8%	95,7%	4,9%
15	PGI	1.870	72,6%	84,4%	11,9%	88,8%	97,2%	8,4%	93,3%	98,3%	5,1%
16	PTI	1.779	72,1%	80,9%	8,9%	85,7%	93,0%	7,3%	88,5%	95,8%	7,2%
17	PVI	3.042	82,5%	93,3%	10,8%	94,2%	97,9%	3,7%	96,9%	98,5%	1,6%
18	SHB	2.634	80,1%	91,0%	10,9%	94,0%	97,8%	3,7%	96,1%	98,4%	2,3%
19	STB	3.317	76,1%	92,3%	16,1%	90,4%	98,6%	8,3%	95,8%	98,9%	3,2%
20	TCB	358	85,8%	91,6%	5,9%	96,6%	95,8%	-0,8%	98,3%	96,9%	-1,4%
21	TPB	387	87,9%	88,4%	0,5%	93,5%	93,3%	-0,3%	94,8%	94,3%	-0,5%
22	VCB	2.581	76,4%	89,9%	13,5%	91,0%	97,8%	6,8%	93,8%	99,1%	5,3%
23	VIB	704	84,7%	91,9%	7,2%	94,3%	96,3%	2,0%	95,7%	96,9%	1,1%
24	VNR	2.986	77,0%	90,9%	14,0%	91,4%	98,2%	6,9%	95,2%	98,8%	3,6%
25	VPB	553	78,7%	85,9%	7,2%	90,8%	95,3%	4,5%	92,8%	96,8%	4,0%

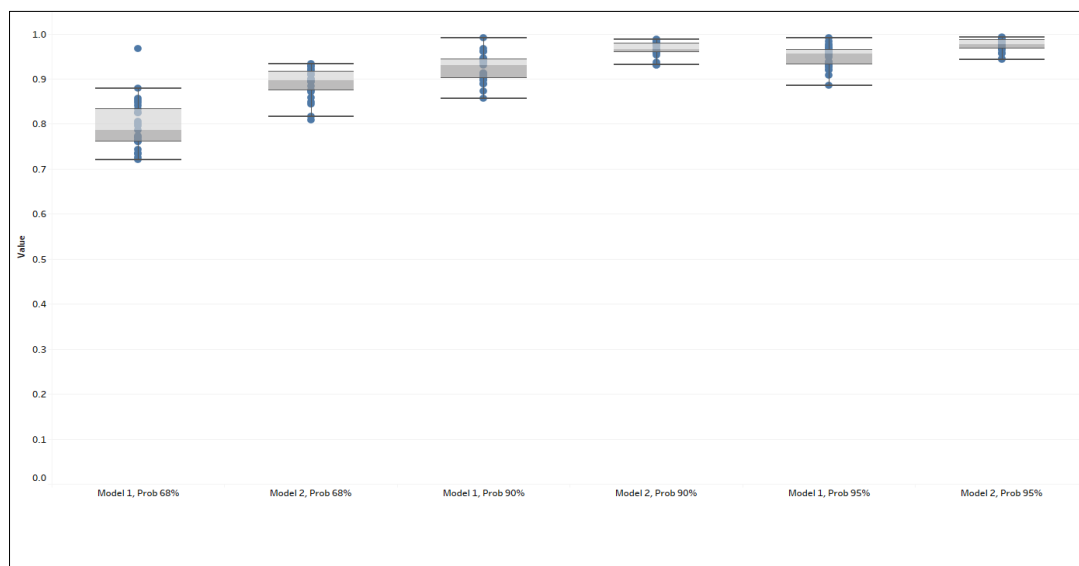
STT	Mã chứng khoán	Tổng số quan sát dự báo	Xác suất 68%			Xác suất 90%			Xác suất 95%		
			MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1
Trung bình của xác suất dự báo đúng			79,8 %	89,0%	9,3%	92,4%	96,5%	4,1%	94,8%	97,5%	2,7%
Độ lệch chuẩn của xác suất dự báo đúng			5,7%	3,4%	5,6%	3,2%	1,7%	3,3%	2,5%	1,4%	2,6%



Hình 10. Đồ thị dạng đường của các xác suất dự báo đúng đối với 25 mã cổ phiếu lần lượt dựa vào xác suất 68%, 90% và 95% theo MH1 và MH2

Rõ ràng, dựa vào kết quả Bảng 5 và Hình 10 cho thấy trong hầu hết các trường hợp, MH2 có tỷ lệ chính xác cao hơn. Cụ thể, trong trường hợp khoảng dự báo nhỏ nhất với xác suất $(1 - \alpha) = 68\%$, số lượng dự báo đúng của MH2 cao hơn hẳn so với MH1. Tỷ lệ dự báo đúng trung bình của MH2 là 89,02%, tỷ lệ dự báo đúng này cao hơn MH1 là 9,3% (với kết quả dự báo đúng của MH1 là 79,8%). Vì vậy, MH2 cho kết quả dự báo chính xác và tin cậy hơn; đáng chú ý, kết quả dự báo nhiều mã của MH2 cao hơn nhiều so với MH1: Mã BVH tăng 19,5% số quan sát dự báo đúng, mã STB tăng 16,1% số quan sát dự báo đúng.

Nếu nhìn về tổng quan theo biểu đồ Box-Plot trong Hình 11, kết quả cho thấy MH2 đều có tỷ lệ dự báo đúng cao hơn MH1.



Hình 11. Đồ thị Box - Plot cho MH1 và MH2 tương ứng với các xác suất khoảng ước lượng

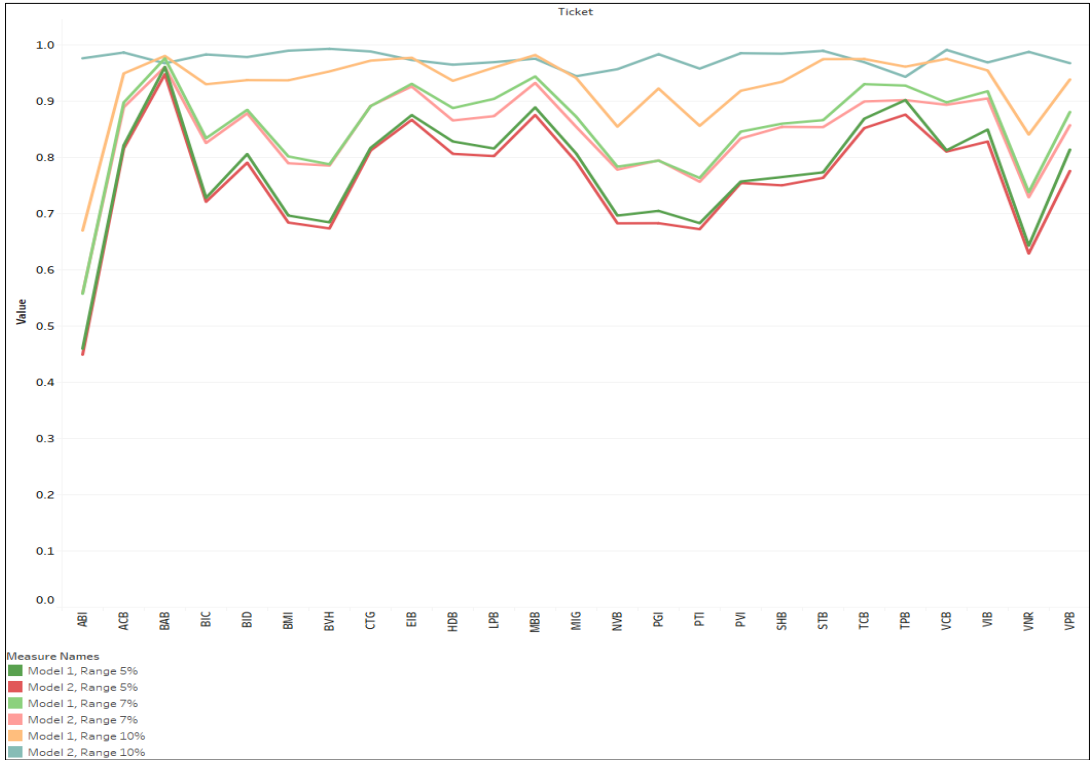
Thêm vào đó, các nhà đầu tư sẽ quan tâm đến dự báo khoảng biến động, vì nếu chỉ dựa vào khoảng dự báo dựa vào xác suất sẽ có khác biệt lớn so với các mức biến động thực của thị trường, như của chỉ số VN-Index là $\pm 7\%$, chỉ số HNX là $\pm 10\%$. Do đó, nhóm tác giả thu hẹp miền biến động lần lượt theo các khoảng biến động $\pm 2,5\%$, khoảng biến động $\pm 3,5\%$, khoảng biến động $\pm 5\%$. Kết quả chi tiết được biểu diễn qua Bảng 6 và các Hình 12, Hình 13 như sau:

Bảng 6.

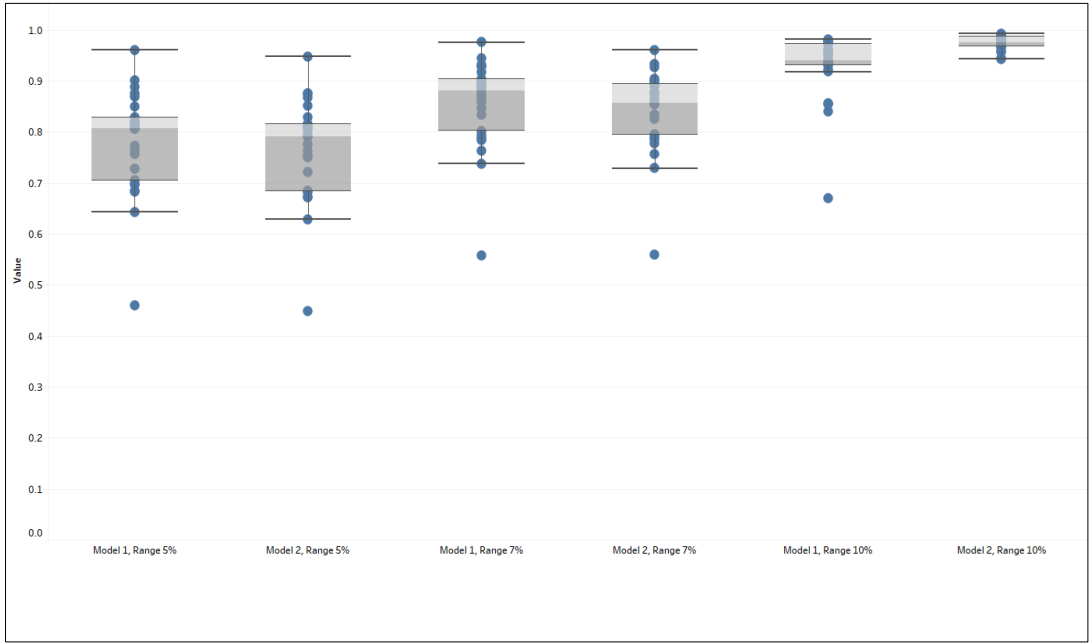
Tỷ lệ dự báo đúng đến thời điểm $(t + 1)$ đối với MH1 và đến thời điểm $(t + 3)$ đối với MH2, các dự báo tương ứng là dự báo khoảng với các miền biến động tương ứng là $\pm 2,5\%$, $\pm 3,5\%$ và $\pm 5\%$

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng biến động $\pm 2,5\%$			Khoảng biến động $\pm 3,5\%$			Khoảng biến động $\pm 5\%$		
			MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1
1	ABI	1.975	46,0%	44,9%	-1,1%	55,7%	55,9%	0,1%	67,0%	97,6%	30,6%
2	ACB	3.229	82,1%	81,6%	-0,6%	89,7%	88,9%	-0,8%	94,9%	98,6%	3,7%
3	BAB	459	96,1%	94,8%	-1,3%	97,6%	96,1%	-1,5%	98,0%	96,7%	-1,3%
4	BIC	2.000	72,8%	72,1%	-0,7%	83,4%	82,6%	-0,8%	93,0%	98,3%	5,3%
5	BID	1.437	80,6%	79,1%	-1,5%	88,4%	87,8%	-0,6%	93,7%	97,8%	4,1%
6	BMI	3.206	69,7%	68,4%	-1,3%	80,2%	79,0%	-1,2%	93,7%	99,0%	5,3%
7	BVH	2.586	68,4%	67,4%	-1,1%	78,8%	78,5%	-0,2%	95,3%	99,3%	4,0%
8	CTG	2.571	81,7%	81,2%	-0,5%	89,1%	89,2%	0,0%	97,2%	98,8%	1,6%

STT	Mã chứng khoán	Tổng số quan sát dự báo	Khoảng biến động $\pm 2,5\%$			Khoảng biến động $\pm 3,5\%$			Khoảng biến động $\pm 5\%$		
			MH1	MH 2	MH2 - MH1	MH1	MH2	MH2 - MH1	MH1	MH2	MH2 - MH1
9	EIB	2.499	87,5%	86,7%	-0,8%	93,1%	92,6%	-0,5%	97,7%	97,3%	-0,4%
10	HDB	454	82,8%	80,6%	-2,2%	88,8%	86,6%	-2,2%	93,6%	96,5%	2,9%
11	LPB	521	81,6%	80,2%	-1,3%	90,4%	87,3%	-3,1%	96,0%	96,9%	1,0%
12	MBB	1.997	88,9%	87,5%	-1,4%	94,4%	93,2%	-1,2%	98,2%	97,6%	-0,6%
13	MIG	610	80,7%	79,2%	-1,5%	87,2%	85,4%	-1,8%	94,1%	94,4%	0,3%
14	NVB	2.105	69,6%	68,3%	-1,4%	78,3%	77,8%	-0,5%	85,5%	95,7%	10,2%
15	PGI	1.870	70,5%	68,3%	-2,2%	79,4%	79,5%	0,1%	92,2%	98,3%	6,1%
16	PTI	1.779	68,3%	67,2%	-1,1%	76,3%	75,7%	-0,7%	85,6%	95,8%	10,2%
17	PVI	3.042	75,7%	75,4%	-0,3%	84,6%	83,4%	-1,2%	91,8%	98,5%	6,7%
18	SHB	2.634	76,5%	75,0%	-1,5%	86,0%	85,4%	-0,6%	93,4%	98,4%	5,0%
19	STB	3.317	77,3%	76,4%	-1,0%	86,6%	85,4%	-1,2%	97,5%	98,9%	1,5%
20	TCB	358	86,9%	85,2%	-1,7%	93,0%	89,9%	-3,1%	97,5%	96,9%	-0,6%
21	TPB	387	90,2%	87,6%	-2,6%	92,8%	90,2%	-2,6%	96,1%	94,3%	-1,8%
22	VCB	2.581	81,2%	81,0%	-0,2%	89,8%	89,4%	-0,4%	97,5%	99,1%	1,6%
23	VIB	704	84,9%	82,8%	-2,1%	91,8%	90,5%	-1,3%	95,5%	96,9%	1,4%
24	VNR	2.986	64,3%	62,9%	-1,4%	73,8%	72,9%	-0,9%	84,1%	98,8%	14,7%
25	VPB	553	81,4%	77,6%	-3,8%	88,1%	85,7%	-2,4%	93,9%	96,8%	2,9%
Trung bình của xác suất dự báo đúng			77,8%	76,5%	-1,4%	85,5%	84,3%	-1,1%	92,9%	97,5%	4,6%
Độ lệch chuẩn của xác suất dự báo đúng			10,3%	10,2%	0,8%	8,7%	8,3%	0,9%	6,6%	1,4%	6,7%



Hình 12. Đồ thị dạng đường của các xác suất dự báo đúng đối với 25 mã cổ phiếu lần lượt dựa vào khoảng biến động 2,5%, 3,5% và 5% theo MH1 và MH2



Hình 13. Đồ thị Box-Plot cho MH1 và MH2 dựa vào thu hẹp miền biến động

Kết quả thực nghiệm từ Bảng 6 cho thấy kết quả MH2 cho kết quả dự báo đa phần đều cao hơn MH1, mặc dù MH2 dự báo đến thời điểm $(t + 3)$ còn MH1 chỉ dự báo đến thời điểm $(t + 1)$. Cụ thể, các kết quả với miền biến động $\pm 2,5\%$ và $\pm 3,5\%$ có tỷ lệ dự báo đúng xấp xỉ nhau giữa hai mô hình, còn với khoảng biến động $\pm 5\%$ thì MH2 có tỷ lệ dự báo đúng cao hơn. Do đó, nếu MH1 cũng dự báo đến thời điểm $(t + 3)$ thì MH2 chắc chắn sẽ có tỷ lệ dự báo đúng cao hơn hẳn. Điều này chứng tỏ tính hiệu quả của mô hình đề xuất trong ứng dụng của thống kê Bayes trong dự báo giá chứng khoán.

4. Kết luận

Trong bài báo này, nhóm tác giả đề xuất cách dự báo giá chứng khoán trong ba ngày tiếp theo kể từ thời điểm nghiên cứu dựa vào hàm mật độ dự báo hậu nghiệm trong thống kê Bayes sẽ có kết quả cải thiện hơn so với cách dự báo chỉ dựa vào một ngày kể từ thời điểm nghiên cứu. Hướng tiếp cận được đề xuất dựa trên các phân tích hành vi của nhà đầu tư dựa trên suất sinh lợi tại thời điểm giao dịch và các phân tích này được đưa vào trong phân phối tiên nghiệm. Các kết quả được cải thiện dựa vào các độ mở trong thống kê Bayes khi các tham số được xem xét như là một biến ngẫu nhiên, đồng thời sử dụng các thông tin tiên nghiệm cho các suy luận tiếp theo thay vì chỉ sử dụng thông tin duy nhất từ dữ liệu như trong thống kê tần suất.

Minh chứng các kết quả thực nghiệm bằng phương pháp dự báo mới trên mẫu gồm 25 ngân hàng thương mại và công ty tài chính đã cho các kết quả dự báo thời điểm $(t + 3)$ cao hơn so với cách dự báo thời điểm $(t + 1)$ từ thời điểm t . Các kết quả dự báo bao gồm cả hai trường hợp: Trường hợp dựa vào lý thuyết thông qua các dự báo khoảng với xác suất 68%, 90% và 95% và trường hợp dựa vào thực tế thông qua các quyết định của thị trường như thu hẹp miền biến động còn $\pm 2,5\%$, $\pm 3,5\%$ và $\pm 5\%$. Hiệu quả của các kết quả dựa vào mô hình đề xuất đưa ra các triển vọng dự báo trong tương lai.

Thực sự việc sử dụng phân phối chuẩn để ước lượng dạng phân phối xác suất của suất sinh lợi trong các áp dụng của thống kê Bayes vẫn còn cần được cải thiện nhằm tìm ra phân phối xác suất phù hợp hơn. Trong thời gian tới, nhóm tác giả sẽ sử dụng hỗn hợp các phân phối xác suất để cải thiện kết quả cũng như mở rộng số lượng dự báo nhiều mã chứng khoán hơn. Đây thật sự là vấn đề cần thiết giúp cho các nhà đầu tư có thêm thông tin để ra quyết định đầu tư đúng đắn cũng như các công ty, doanh nghiệp có các đối sách phù hợp trong thời đại các thay đổi diễn ra thường xuyên và liên tục ■

Tài liệu tham khảo

- Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation* (pp. 106–112). doi: 10.1109/UKSim.2014.67
- Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian Statistics*. Hoboken, NJ, USA: John Wiley & Sons.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. New York: Chapman & Hall/CRC.
- Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science*, 4(2), 117–121.

- Lê Thanh Hoa, Phạm Hoàng Uyên., & Nguyễn Đình Thiên. (2017a). Một phương pháp mới tìm khoảng mật độ hậu nghiệm cao nhất và ứng dụng. *Tạp chí Phát triển Kinh tế*, 28(10), 79–120.
- Lê Thanh Hoa, Phạm Hoàng Uyên., & Nguyễn Phúc Sơn. (2017b). Sử dụng thống kê Bayes mờ trong dự báo tỷ giá và một số chỉ số kinh tế. *Tạp chí Công nghệ Ngân hàng*, 140, 92–100.
- Le, H., Pham, U., Nguyen, P., & Pham, T. B. (2020). Improvement on Monte Carlo estimation of HPD intervals. *Communications in Statistics - Simulation and Computation*, 49(8), 2164–2180. doi: 10.1080/03610918.2018.1513141
- Pham, U., Le, H. T., & Nguyen, T. (2017). Choosing the best model in fuzzy Bayesian statistics and its application in financial analysis. *Science & Technology Development Journal - Economics - Law and Management*, 1(Q2), 144–155.
- Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Science & Business Media.
- Tran, H. D., Nguyen, S., Le, H. T., & Pham, U. (2017). An alternative to p-values in Hypothesis testing with applications in model selection of stock price data. In V. Kreinovich, S. Sriboonchitta, & V.-N. Huynh (Eds.), *Robustness in Econometrics* (pp. 305–319). Springer.