

PHÂN TÍCH VÀ DỰ BÁO CHỈ SỐ THỊ TRƯỜNG CHỨNG KHOÁN BẰNG SỬ DỤNG CHỈ SỐ BÁO TRƯỚC

Đỗ Văn Thành¹, Nguyễn Minh Hải²

¹Khoa Công nghệ thông tin, Trường Đại học Nguyễn Tất Thành

²Khoa Cơ bản, Trường Đại học Công nghiệp, Thành phố Hồ Chí Minh
dvthanh@ntt.edu.vn, , nguyenminhhaidhcn@gmail.com

TÓM TẮT — Mục đích của bài báo này là đề xuất kết hợp phương pháp chỉ số dẫn báo và hệ số tương quan giữa chỉ số thị trường chứng khoán của sàn giao dịch với các biến dữ liệu giao dịch cổ phiếu trong việc xây dựng mô hình dự báo chỉ số thị trường chứng khoán. Áp dụng phương pháp được đề xuất, bài báo sẽ thực hành xây dựng mô hình dự báo chỉ số thị trường chứng khoán VNINDEX của sàn giao dịch Thành phố Hồ Chí Minh. So sánh kết quả dự báo bằng sử dụng mô hình so với số liệu thống kê thực tế cho thấy triển vọng của phương pháp dự báo được đề xuất trong việc xây dựng mô hình dự báo chỉ số thị trường chứng khoán.

Từ khóa — Mô hình dự báo, chỉ số thị trường chứng khoán, chỉ số báo trước, hệ số tương quan.

I. GIỚI THIỆU

Trong những nền kinh tế phát triển theo cơ chế thị trường, Chứng khoán là một kênh thu hút vốn đầu tư quan trọng cho phát triển kinh tế. Dự báo thị trường chứng khoán gồm 2 nội dung là dự báo chỉ số thị trường chứng khoán và dự báo giá của các cổ phiếu trên thị trường [12]. Dự báo thị trường chứng khoán được quan tâm nghiên cứu, ứng dụng từ lâu và liên tục được nghiên cứu tìm tòi, đề xuất mới. Hiện đã có nhiều tổ chức, trung tâm nghiên cứu thực hiện dự báo về thị trường chứng khoán và cung cấp thông tin dự báo trên internet [20]. Rất nhiều phương pháp/kỹ thuật dự báo thị trường chứng khoán đã được đề xuất và thực nghiệm, nhưng các tổ chức thực hiện và cung cấp thông tin dự báo về thị trường chứng khoán đã sử dụng các phương pháp/kỹ thuật này như thế nào để dự báo thị trường chứng khoán là điều thuộc bí mật công nghệ.

G. Preethi and B. Santhi đã khảo sát các kỹ thuật dự báo thị trường chứng khoán cho đến năm 2012 [12], theo đó có thể thấy chủ yếu đó là những kỹ thuật khai phá dữ liệu sau: mạng nơtron [2, 6, 14, 16], mô hình Markov [13, 17], hệ suy luận nơtron-mờ (Neuro Fuzzy) [8-9, 15], giải thuật di truyền [5, 14], phân tích chuỗi thời gian [7], phương pháp hồi quy [1], lý thuyết tập thô [5],... cũng như sự kết hợp của một số kỹ thuật này trong việc dự báo thị trường chứng khoán [5, 9, 11, 14, 16-17]. Một số hạn chế của các kỹ thuật nêu trên cũng được giới thiệu trong [8].

Những năm gần đây người ta đề xuất mới một số kỹ thuật dự báo thị trường chứng khoán thông qua phân tích và dự báo niềm tin thị trường, sử dụng kỹ thuật phân tích thành phần chính [3, 18], hay kỹ thuật học máy véc tơ hỗ trợ [4, 19] cũng như kết hợp kỹ thuật học máy véc tơ hỗ trợ với kỹ thuật phân tích thành phần chính [3]. Nhưng các kỹ thuật này chủ yếu chỉ dự báo được xu hướng của thị trường chứng khoán mà chưa đưa ra được giá trị dự báo cụ thể.

Trong phương pháp hồi quy để dự báo thị trường chứng khoán [1], người ta thường xây dựng mô hình dự báo thị trường chứng khoán theo cách tiếp cận kinh tế lượng, đó là sử dụng một số biến kinh tế vĩ mô và biến tài chính tiền tệ mà theo lý thuyết kinh tế có tác động đến biến động của thị trường chứng khoán làm biến giải thích trong mô hình hồi quy đa biến.

Thực tế trong lĩnh vực kinh tế cho thấy tồn tại nhiều chuỗi thời gian mà những biến động của nó có quan hệ ổn định với những biến động của một hoặc một vài chuỗi thời gian khác. Khi đó các thông tin về một số chuỗi thời gian (được gọi là chỉ số tác động) có thể được sử dụng để hỗ trợ giám sát và dự báo về biến động của một số chuỗi thời gian khác.

Các chỉ số tác động được phát hiện không dựa vào lý thuyết kinh tế mà chủ yếu dựa vào dữ liệu thực tế có được và được tìm thông qua xây dựng mô hình dự báo. Các chỉ số tác động được phân làm 3 loại: chỉ số báo trước (hay dẫn báo), chỉ số báo đồng thời và chỉ số báo sau. Chỉ số báo trước cho biết trước sự biến đổi tương lai của một số chuỗi thời gian mà nó có tác động, trong khi chỉ số báo đồng thời và chỉ số báo sau cho biết thông tin biến động cùng nhịp thời gian và thông tin biến động quá khứ của những chuỗi thời gian mà nó có tác động [7]. Trong phân tích và dự báo người ta đặc biệt quan tâm đến hai loại chỉ số tác động đầu. Các chỉ số báo trước liên quan đến việc xây dựng mô hình dự báo không điều kiện trong khi các chỉ số báo đồng thời liên quan đến mô hình dự báo có điều kiện [10].

Một số câu hỏi đặt ra khi nghiên cứu phân tích và dự báo về thị trường chứng khoán: liệu có tồn tại một số biến dữ liệu giao dịch cổ phiếu là chỉ số báo trước của chỉ số thị trường chứng khoán không? Nếu có tìm bằng cách nào? Mô hình dự báo chỉ số thị trường chứng khoán được xây dựng thông qua các chỉ số báo trước nên được xây dựng thế nào? và độ chính xác dự báo theo mô hình đó ra sao?. Mục đích của bài báo này là nhằm trả lời các câu hỏi đó, cụ thể Bài báo sẽ đề xuất phương pháp xây dựng mô hình dự báo chỉ số thị trường chứng khoán bằng sử dụng các chỉ số báo trước của chỉ số thị trường chứng khoán làm biến giải thích trong mô hình hồi quy nhiều biến và áp dụng phương pháp được đề xuất để dự báo chỉ số thị trường chứng khoán Việt Nam VNINDEX.

Phần còn lại của Bài báo được cấu trúc như sau: tiếp theo mục này, mục II tiếp theo sẽ giới thiệu một số kiến thức có tính chất chuẩn bị. Mục III sẽ đề xuất phương pháp xây dựng mô hình dự báo chỉ số thị trường chứng khoán bằng sử dụng chỉ số báo trước. Mục IV sẽ trình bày kết quả xây dựng mô hình dự báo chỉ số VNINDEX theo phương pháp được đề xuất và cuối cùng là một số kết luận và hướng nghiên cứu tiếp theo.

II. MỘT SỐ KIẾN THỨC CHUẨN BỊ

Trong phần này sẽ giới thiệu lại một số kiến thức nền tảng mà Bài báo sẽ sử dụng. Giả sử X và Y đều là những chuỗi dừng, xét 2 phương trình sau đây [7]:

$$Y = \sum_{i=1}^n a_i X(-i) + \sum_{j=1}^m b_j Y(-j) + u_{1t} \quad (1)$$

$$X_t = \sum_{i=1}^p c_i X(-i) + \sum_{j=1}^q d_j Y(-j) + u_{2t} \quad (2)$$

ở đây các a_i, c_i, b_j, d_j , là các tham số; $X(-i), Y(-j)$ tương ứng là ký hiệu biến X trễ i bước và Y trễ j bước; n, m, p, q là số các biến giải thích, nó cũng chính là độ dài trễ lớn nhất của các biến X, Y và u_{it} ($i=1, 2$) là sai số được giả định là nhiễu trắng.

Ta nói tồn tại mối quan hệ nhân quả duy nhất chiều từ X đến Y nếu $\sum_{i=1}^n a_i^2 \neq 0$ và $\sum_{j=1}^m d_j^2 = 0$, và tương tự, tồn tại mối quan hệ nhân quả duy nhất chiều từ Y đến X nếu $\sum_{i=1}^n a_i^2 = 0$ và $\sum_{j=1}^m d_j^2 \neq 0$. Nếu $\sum_{i=1}^n a_i^2 \neq 0$ và $\sum_{j=1}^m d_j^2 \neq 0$ thì hai biến X và Y có liên quan trực tiếp với nhau, khi đó giữa hai biến X và Y còn được gọi là có quan hệ nhân quả hai chiều (hay quan hệ phản hồi). Các biến X và Y là độc lập nhau nếu $\sum_{i=1}^n a_i^2 = 0$ và $\sum_{j=1}^m d_j^2 = 0$.

Khi có quan hệ nhân quả chiều từ X đến Y thì X chính là nguyên nhân gây ra Y và nó cũng chính là chỉ số báo trước của Y .

Một cách tổng quát: quan hệ giữa các biến X_1, X_2, \dots, X_n và Y được gọi là quan hệ nhân quả từ các biến này đến Y nếu Y được dự báo từ quá khứ của các biến đó. Nói cách khác khi đó:

$$Y_t = \sum_{i=1}^n a_i X_1(-i) + \sum_{j=1}^p b_j X_2(-j) + \dots + \sum_{h=1}^k c_h X_n(-h) + \sum_{q=1}^m d_q Y(-q) + u_t \quad (3)$$

ở đây: $\sum_{i=1}^n a_i^2 \neq 0, \sum_{j=1}^p b_j^2 \neq 0$ và $\sum_{h=1}^k c_h^2 \neq 0$ và trong trường hợp này các biến X_1, X_2, \dots, X_n cũng là các chỉ số báo trước của biến Y .

Trong ứng dụng thực tế, để làm giảm tính phức tạp trong các công thức (1) và (2) người ta thường cho $n=m=p=q$, tức là độ trễ lớn nhất của các biến trong các công thức này là như sau [7, 21].

III. PHƯƠNG PHÁP XÂY DỰNG MÔ HÌNH DỰ BÁO

Giá trị của chỉ số thị trường chứng khoán của một sàn giao dịch được tính toán thông qua giá trị giao dịch của các cổ phiếu được niêm yết trên sàn này. Nên về nguyên tắc ta có thể dự báo chỉ số thị trường chứng khoán thông qua các biến dữ liệu giao dịch cổ phiếu.

Ký hiệu Y là biến chỉ số thị trường chứng khoán. Y là một biến véctor, $Y^T = (y_1, y_2, \dots, y_m)$, $y_i, i=1, 2, \dots, m$ là giá trị của biến Y ở ngày giao dịch thứ i , m là số các giao dịch được thực hiện.

Ký hiệu X_j ($j=1, 2, \dots, n$), là biến dữ liệu giao dịch cổ phiếu thứ j và n là số các biến dữ liệu giao dịch cổ phiếu. Ta có thể sử dụng các mã cổ phiếu để làm tên cho các biến này. Các biến X_j cũng là biến véctor, $X_j^T = (x_{j1}, x_{j2}, \dots, x_{jm})$, ở đây x_{jk} là giá trị của X_j trong ngày giao dịch thứ k , nó là tích của giá trung bình trong ngày của X_j nhân với số lượng của cổ phiếu này được giao dịch trong ngày đó.

Phương pháp xây dựng mô hình dự báo chỉ số thị trường chứng khoán bằng sử dụng chỉ số báo trước được trình bày trong Hình 1 ở dưới và được giải thích như sau:

Bước 1: Tìm các biến dữ liệu giao dịch cổ phiếu X_j là chỉ số báo trước của chỉ số thị trường chứng khoán Y

1) Kiểm tra tính dừng của các biến Y và X_j . Nếu chúng chưa dừng thì phải biến đổi các chuỗi không dừng thành chuỗi dừng [7].

2) Giả sử các biến Y và X_j đều dừng, tìm quan hệ nhân quả với trễ là 5 giữa biến X_j và biến Y . Độ dài trễ là 5 hàm ý rằng chỉ số thị trường chứng khoán cũng như các biến dữ liệu giao dịch cổ phiếu do được thống kê theo ngày trong tuần làm việc chỉ gồm 5 ngày. Đó cũng là độ dài mùa vụ của các biến chuỗi thời gian Y và X_j .

3) Lựa chọn các biến X_j thỏa mãn giả thuyết H_0 : không có quan hệ nhân quả chiều từ X_j đến Y với mức ý nghĩa dưới 10% và khi đó giả thuyết H_1 : Có quan hệ nhân quả chiều từ X_j đến Y sẽ có mức ý nghĩa trên 90%.

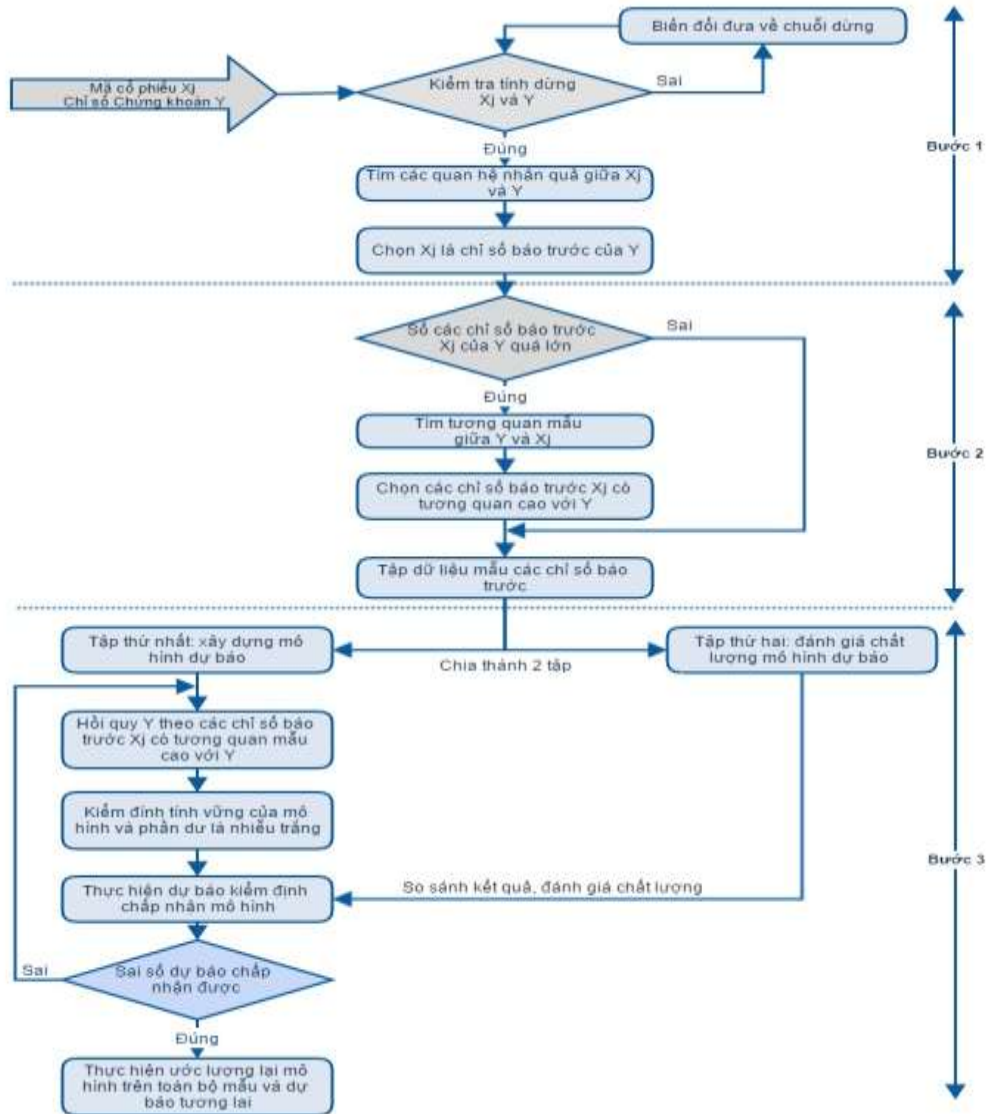
Bước 2: Lựa chọn chỉ số báo trước để xây dựng mô hình dự báo

Nếu số lượng các chỉ số báo trước của biến Y là quá lớn (do người sử dụng tự quy ước, nhưng thường là trên 10) thì:

4) Tìm tương quan mẫu giữa chỉ số thị trường chứng khoán Y và các chỉ số dẫn báo X_j ;

5) Lựa chọn các chỉ số báo trước của Y mà giữa chúng có hệ số tương quan cao về giá trị tuyệt đối (ngưỡng hệ số tương quan để lựa chọn do người dùng quyết định).

Trong trường hợp các chỉ số báo trước của Y là không quá lớn thì chuyển sang Bước 3.



Hình 1. Phương pháp dự báo chỉ số thị trường chứng khoán bằng sử dụng chỉ số báo trước

Bước 3: Xây dựng mô hình dự báo chỉ số thị trường chứng khoán Y theo các chỉ số báo trước

6) Chia tập dữ liệu thành 2 tập. Tập thứ nhất để xây dựng mô hình dự báo và tập thứ 2 để phục vụ kiểm thử mô hình dự báo.

7) Giả sử $X_{j1}, X_{j2}, \dots, X_{jk}$ là các chỉ số báo trước được chọn ở Bước 2, hồi quy theo công thức: $Y = f(X_{j1}, X_{j2}, \dots, X_{jk}) + u_t$, ở đây $f(\cdot)$ có dạng như trong công thức (3). Quá trình hồi quy được lặp đi lặp lại sao cho tất cả các tham số ước lượng trong mô hình dự báo đều có ý nghĩa thống kê, theo thông lệ thường ở mức dưới 10%, phân dư là nhiễu trắng, không có hiện tượng nội sinh của phân dư và mô hình ước lượng là vững [10].

8) Thực hiện dự báo kiểm định chấp nhận mô hình bằng cách: sử dụng mô hình được xây dựng trên tập dữ liệu thứ nhất để dự báo tập dữ liệu thứ 2 và so sánh kết quả dự báo với số liệu thực tế trong tập dữ liệu thứ 2 để đánh giá chất lượng dự báo của mô hình.

9) Nếu sai số của dự báo là chấp nhận được theo quan điểm của người dùng thì ước lượng lại mô hình trên toàn bộ tập dữ liệu và sử dụng nó để dự báo tương lai.

Để xây dựng mô hình dự báo chỉ số thị trường chứng khoán theo phương pháp vừa nêu có thể sử dụng các phần mềm công cụ sau: SAS, STATA, EVIEW hay R,... Bài báo này sử dụng phần mềm công cụ EVIEW [21].

IV. XÂY DỰNG MÔ HÌNH DỰ BÁO CHỈ SỐ VNINDEX

4.1. Tập dữ liệu được sử dụng

Dữ liệu về chỉ số thị trường chứng khoán VNINDEX và các biến dữ liệu giao dịch cổ phiếu được thu thập trên sàn giao dịch Thành phố Hồ Chí Minh từ ngày 4/01/2010 đến ngày 5/5/2016, bao gồm 278 biến dữ liệu giao dịch cổ phiếu kể cả chỉ số thị trường chứng khoán và 1574 quan sát do các ngày thứ 7, chủ nhật và các ngày nghỉ lễ, sàn giao dịch chứng khoán không hoạt động.

4.2. Xây dựng mô hình dự báo

Bước 1: Tìm các chỉ số báo trước của Δ VNINDEX

- Thực hiện kiểm định Dickey – Fuller tăng cường về nghiệm đơn vị của tất cả các biến dữ liệu giao dịch cổ phiếu và chỉ số VNINDEX ta sẽ nhận được tất cả các biến dữ liệu giao dịch cổ phiếu đều là chuỗi dừng, trong khi VNINDEX không dừng nhưng sai phân bậc 1 của nó (ký hiệu là Δ VNINDEX) là chuỗi dừng.

- Thực hiện kiểm định quan hệ nhân quả với độ dài trễ là 5 ta sẽ nhận được 38 biến dữ liệu giao dịch cổ phiếu có quan hệ nhân quả chiều từ các biến này đến Δ VNINDEX với ý nghĩa thông kê ở mức dưới 10%. Cột 2 (Bảng 1) chỉ ra danh sách 38 biến dữ liệu giao dịch cổ phiếu như vậy. Cột 3 (Bảng 1) là mức ý nghĩa của giả thuyết H_0 : “Biến dữ liệu giao dịch cổ phiếu ở cùng dòng thuộc Cột 2 không phải là nguyên nhân của Δ VNINDEX”. Các ký hiệu: *, ** và *** trong Cột 3 (Bảng 1) tương ứng là các mức ý nghĩa dưới 10%, 5% và 1%.

Bước 2: Lựa chọn chỉ số báo trước để xây dựng mô hình dự báo

- Nếu lựa chọn tất cả 38 biến dữ liệu giao dịch cổ phiếu (là chỉ số báo trước) để làm biến giải thích khi xây dựng mô hình dự báo chỉ số thị trường chứng khoán VNINDEX theo công thức (3) thì mô hình dự báo khi đó có thể có đến $39 \times 5 = 195$ biến giải thích (VNINDEX và mỗi biến dữ liệu giao dịch cổ phiếu đều có 5 biến trễ). Số lượng biến giải thích khá lớn sẽ gây khó khăn khi phải thực hiện nhiều kỹ thuật xử lý bổ sung khác để mô hình ước lượng cuối cùng thỏa mãn tất cả các yêu cầu kiểm định đã nêu ở mục III. Vì thế cần lựa chọn chỉ một số số biến trong đó làm đại diện.

- Tính hệ số tương quan mẫu giữa 38 chỉ số báo trước và Δ VNINDEX ta sẽ nhận được Cột 3 Bảng 1. Trong bài báo này các chỉ số báo trước mà hệ số tương quan mẫu của nó với Δ VNINDEX có giá trị tuyệt đối không nhỏ hơn 0.0399 sẽ được chọn. Theo tiêu chuẩn này có 6 chỉ số báo trước được khi chọn để làm biến giải thích xây dựng mô hình dự báo VNINDEX là: ANV, EIB, ITC, PXL, TIC, VCF (Cột 3, Bảng 1).

Số thứ tự	Mã cổ phiếu (1)	Mức ý nghĩa của giả thuyết H_0 (2)	Tương quan mẫu với Δ VNINDEX (3)	Mã cổ phiếu được lựa chọn	Số thứ tự	Mã cổ phiếu (1)	Mức ý nghĩa của giả thuyết H_0 (2)	Tương quan mẫu với Δ VNINDEX (3)	Mã cổ phiếu được lựa chọn
1	AGF	*	0.0239		20	PVD	*	-0.0054	
2	AGR	*	0.0204		21	PXL	*	0.0399	PXL
3	ANV	*	-0.0629	ANV	22	RIC	**	-0.0049	
4	ASM	***	0.0022		23	SAM	*	-0.0016	
5	BGM	*	0.0132		24	SFI	*	-0.0048	
6	D2D	*	-0.0151		25	SRF	**	-0.0258	
7	DPR	**	-0.0191		26	TCL	*	0.0099	
8	EIB	**	0.0436	EIB	27	TDC	**	-0.0043	
9	FMC	**	-0.0361		28	TIC	**	-0.0666	TIC
10	HU1	*	0.0175		29	TLH	*	0.0122	
11	ITC	***	-0.0534	ITC	30	TMP	***	-0.0284	
12	KSA	*	0.0017		31	TNC	*	-0.0132	
13	LSS	*	0.0238		32	TRA	*	0.0316	
14	NBB	*	0.0032		33	TYA	**	0.0349	
15	NTL	**	0.0048		34	VCF	***	-0.0876	VCF
16	PJT	*	-0.0096		35	VFG	*	0.0047	
17	POM	*	-0.012		36	VID	*	-0.0178	
18	PPI	***	-0.0206		37	VNA	**	0.002	
19	PTC	**	-0.0256		38	VRC	*	0.0164	

Bảng 1. Chỉ số báo trước, hệ số tương quan mẫu và biến dữ liệu giao dịch cổ phiếu được chọn

Bước 3: Xây dựng mô hình dự báo chỉ số thị trường chứng khoán

- Tập số liệu đầu vào được chia thành hai tập. Tập thứ nhất gồm các ngày từ 4/01/2010 đến ngày 22/4/2016 được sử dụng để xây dựng mô hình dự báo, tập thứ 2 gồm các ngày từ 25/4/2016 đến ngày 5/5/2016 (gồm 7 ngày giao dịch do các ngày từ 30/4/2016 đến 3/5 là những ngày nghỉ lễ, sàn giao dịch không làm việc) được sử dụng để kiểm định mô hình.

- Do $\Delta \text{VNINDEX}$ là dừng nên $\Delta \text{Ln}(\text{VNINDEX})$ cũng là chuỗi dừng, hơn nữa $\Delta \text{Ln}(X)$ chính là tốc độ thay đổi của biến X . Vì thế thay cho biến $\Delta(\text{VNINDEX})$, bài báo này sẽ thực hiện hồi quy $\Delta \text{Ln}(\text{VNINDEX})$ theo công thức (3) đối với các chỉ số báo trước ANV, EIB, ITC, PXL, TIC, VCF. Kết quả ta nhận được Bảng 2.

- Thực hiện lặp phép hồi quy theo cách loại trừ dần từng biến theo thứ tự của mức ý nghĩa thống kê của hệ số các biến sao cho phương trình ước lượng với các tham số đều có ý nghĩa thống kê ở mức dưới 10% và bằng cách bổ sung thêm bình phương và tích chéo của các biến trễ của các biến giải thích vào tập các biến giải thích ta nhận được mô hình dự báo được xác định như trong Bảng 2.

Biến phụ thuộc: DLOG(VNINDEX) (trong EVIEW: DLOG là ΔLn)

Số quan sát: 1560 sau khi điều chỉnh;

Biến	Hệ số	Sai số chuẩn	Mức ý nghĩa
DLOG(VNINDEX(-1))	0.251563	0.026508	***
DLOG(VNINDEX(-2))	-0.070045	0.025714	***
DLOG(VNINDEX(-3))	0.116759	0.031631	***
DLOG(VNINDEX(-4))	-0.070953	0.025003	***
ITC(-1)	-5.84E-11	2.28E-11	**
ITC(-4)	1.02E-10	2.27E-11	***
ITC(-5)	-5.55E-11	2.21E-11	**
TIC(-1)	-4.12E-09	1.22E-09	***
VCF(-2)	-1.09E-09	3.06E-10	***
C	0.000946	0.000341	***
DLOG(VNINDEX(-1))*VCF(-4)	4.63E-08	1.24E-08	***
DLOG(VNINDEX(-3))*DLOG(VNINDEX(-4))	6.940.499	1.831.745	***
DLOG(VNINDEX(-3))*EIB(-2)	-2.63E-09	9.16E-10	***
DLOG(VNINDEX(-3))*VCF(-5)	3.20E-08	1.43E-08	**
ANV(-5)*PXL(-3)	-7.47E-17	2.67E-17	***
EIB(-2)*VCF(-5)	-6.57E-17	1.64E-17	***
EIB(-4)*PXL(-1)	-1.21E-17	4.19E-18	***
EIB(-4)*PXL(-2)	1.48E-17	3.91E-18	***
EIB(-4)*VCF(-5)	4.96E-17	1.46E-17	***
ITC(-1)*VCF(-4)	-6.25E-17	3.15E-17	**
ITC(-4)*VCF(-5)	-1.49E-16	4.73E-17	***
PXL(-1)*VCF(-2)	2.37E-16	8.37E-17	***
VCF(-4)*VCF(-5)	4.39E-16	8.11E-17	***

$R^2 = 0.164$; Thống kê DW = 1.994797.

Bảng 2. Bảng ước lượng $\Delta \text{Ln}(\text{VNINDEX})$

Phương trình ước lượng được nêu trong Bảng 2 là vững theo kiểm định Ramsey, phần dư có kỳ vọng bằng 0, phần dư có phân phối chuẩn theo kiểm định Jarque-Bera, không có hiện tượng nội sinh phần dư và phần dư không tự tương quan theo kiểm định Breusch-Godfrey và phần dư có phương sai không đổi theo kiểm định White [10, 21]. Phân tích chéo của các biến trễ trong phương trình ước lượng ở Bảng 2 thực chất nhằm khắc phục tình trạng phần dư có phương sai thay đổi.

Như vậy mô hình dự báo chỉ số VNINDEX sử dụng chỉ số báo trước được nêu ở Bảng 2 có các tham số đều có ý nghĩa thống kê, các kiểm định về tính vững và kiểm định về phần dư đều đáp ứng yêu cầu để mô hình ước lượng là không chệch, vững và tốt nhất theo phương pháp hồi quy bình phương tối thiểu.

Nếu chú ý rằng DLOG(VNINDEX) chính là tốc độ thay đổi của chỉ số thị trường chứng khoán VNINDEX thì Phương trình ước lượng (Bảng 2) cung cấp nhiều thông tin có giá trị. Chẳng hạn từ phương trình này có thể rút ra một số kết luận kiểu như: nếu hôm nay chỉ số thị trường chứng khoán VNINDEX tăng/giảm 1% thì ngày mai chỉ số này sẽ

tăng/giảm 0.25% nhưng đến ngày kia thì chỉ số VNINDEX giảm/tăng 0.07%,... Còn tác động của các biến dữ liệu giao dịch cổ phiếu đến chỉ số VNINDEX được diễn giải khác một chút cụ thể nếu hôm nay, khối lượng giao dịch của mã chứng khoán ITC tăng lên 1 đơn vị thì chỉ số thị trường chứng khoán VNINDEX ngày mai sẽ giảm $5.84e-11$ % nhưng đến 4 ngày sau đó nó sẽ tăng lên $1.02e-10$ %, ...

Để đánh giá chất lượng dự báo của mô hình, ta sử dụng mô hình để dự báo cho 7 phiên giao dịch gồm các ngày từ 25/4/2016 đến 29/4/2016 và 2 ngày 4, 5 tháng 5 năm 2016. Bảng 3 ở dưới so sánh % sai số giữa kết quả dự báo bằng mô hình so với giá trị thống kê thực tế. Sai số lớn nhất rơi vào ngày đầu tiên của dự báo mà theo lẽ thông thường nó cần cho kết quả dự báo tốt nhất. Nhưng nếu để ý rằng đó là ngày thứ 2 đầu tuần và ở thứ 6 tuần trước, ngày 22/4/2016 chỉ số thị trường chứng khoán VNINDEX khi chốt phiên là 583.55 điểm thì việc thứ 2 ngày 25/4/2016 VNINDEX đạt 596.6 điểm có thể do trong 2 ngày cuối tuần đã có những thông tin tích cực đối với thị trường chứng khoán làm các nhà đầu tư lạc quan, đẩy mạnh tham gia thị trường. Mô hình dự báo có thể chưa nắm bắt kịp và đưa được những thông tin này vào mô hình dự báo.

Thứ	Ngày	VNINDEX	VNINDEXF	% sai số
Thứ 2	25/04/2016	596.6	587.1673	-1.58
Thứ 3	26/04/2016	594.35	600.2165	0.99
Thứ 4	27/04/2016	596.6	594.815	-0.30
Thứ 5	28/04/2016	593.65	597.9075	0.72
Thứ 6	29/04/2016	595.45	591.7211	-0.63
Thứ 4	04/05/2016	597.75	596.4889	-0.21
Thứ 5	05/05/2016	602.15	597.9781	-0.69

Bảng 3. So sánh VNINDEX thực tế và VNINDEXF được dự báo bằng mô hình

V. KẾT LUẬN

- Mô hình dự báo chỉ số thị trường chứng khoán VNINDEX là mô hình dự báo không điều kiện. Ta có thể tính được giá trị của VNINDEX chỉ bằng cách dựa vào quá khứ của nó và của các chỉ số báo trước. Sai số dự báo của mô hình nói chung là có thể chấp nhận được.

- Để dự báo được chính xác hơn ngoài những thông tin định lượng nhận được từ các chỉ số báo trước đã được đưa vào mô hình dự báo cần phải thu thập và phân tích một số thông tin bổ sung khác, nhất là các thông tin kinh tế và tài chính có tác động đến biến động của thị trường chứng khoán. Nói cách khác cần kết hợp dự báo bằng mô hình định lượng với dự báo định tính (thu thập và phân tích thông tin có tác động đến chỉ số thị trường chứng khoán). Chính yêu cầu này làm cho việc dự báo nói chung và dự báo chứng khoán nói riêng khó khăn lên rất nhiều. Không chỉ đơn thuần là khoa học, để dự báo chính xác còn cần nhiều kinh nghiệm, vốn sống và văn hóa của người làm công tác dự báo.

- Như đã biết các quan hệ nhân quả thường không ổn định bởi vậy phương pháp dự báo chỉ số thị trường chứng khoán sử dụng chỉ số dẫn báo chỉ áp dụng cho các dự báo ngắn hạn. Hơn nữa khi hoạt động của sàn giao dịch có dấu hiệu diễn ra bất thường thì một mặt phải xác định lại các chỉ số báo trước, đồng thời cần xem xét phân tích thêm thông tin về những yếu tố có tác động đến hoạt động giao dịch trên sàn cũng như thông tin có ảnh hưởng đến giao dịch của các mã cổ phiếu là chỉ số báo trước được lựa chọn để xây dựng mô hình.

- Thông thường, khi nền kinh tế trong nước cũng như thế giới có những diễn biến rất bất thường, hoạt động giao dịch trên sàn chứng khoán bất ổn thì xu hướng chung là sử dụng mô hình dự báo có điều kiện trong dự báo chỉ số thị trường chứng khoán. Xây dựng mô hình có điều kiện để dự báo chỉ số thị trường chứng khoán như thế nào đang được các tác giả bài báo này nghiên cứu và sẽ được trình bày trong nghiên cứu khác.

TÀI LIỆU THAM KHẢO

- [1] Abdulsalam S. O., Adewole K., Jimoh R.G., “Stock Trend Prediction using Regression Analysis – A Data Mining Approach”, *AJSS journal*, ISSN 2222-9833, 2010.
- [2] Akinwale A.T., Arogundade O.T. and Adekoya A.F., “Translated Nigeria stock market price using artificial neural network for effective prediction. *Journal of theoretical and Applied Information technology*, 2009.
- [3] Carol A.H., Chandrika K.M. (2015), “The Selection of Winning Stocks Using Principal Component Analysis” *American Journal of Marketing Research*, Vol. 1, No. 3, 2015, pp. 183-188.
- [4] Chandrika K.M., Carol A.H., “Stock Trading Using Analytics”, *American Journal of Marketing Research*, Vol. 2, No. 2, 2016, pp. 27-37.
- [5] Cheng C.-H., Chen T.-L., Wei L. Y., “A hybrid model based on rough set theory and genetic algorithms for stock price forecasting”, 2010, pp. 1610-1629.
- [6] Dase R.K. and Pawar D.D., “Application of Artificial Neural Network for stock market predictions: A review of literature” *International Journal of Machine Intelligence*, ISSN: 0975-2927, Volume 2, Issue 2, 2010, pp-14-17.
- [7] Enders, W., *Applied Econometric Time Series*, Wiley: USA, 2004.

- [9] George S. A. and Kimon P.V., “Forecasting stock market short-term trends using a neuro-fuzzy based methodology”, *Expert systems with Application*, 2009, pp. 10696-10707.
- [10] George S. A., Emmanouil M.D. and Constantinos D. Z., “Elliot Wave Theory and neurofuzzy systems, in stock market predictions: The WASP system”, *Expert systems with application*, 2011, pp. 9196-9206.
- [11] Graham E., Granger C.W.J., Timmerman A. (2006), *Hanbook of Economic Forecasting*, Volume 1, Elsevier BV, 2006, 933 p.
- [12] Kuang Y. H., Jane C.-J., “A hybrid model stock market forecasting and portfolio selection based on ARX, grey system and RS theories”, *Expert systems with Applications*, 2009, pp. 5387-5392.
- [13] Preethi G. and Santhi B., “Stock Market Forecasting Techniques: A Survey”, *Journal of theoretical and Applied Information technology*, Vol 46, No 1, 2012, pp. 24-30.
- [14] Rafiul H.Md., Baikunth N., “Stock Market forecasting using Hidden Markov Model: A New Approach”, *Proceeding of the 2005 5th international conference on intelligent Systems Design and Application 0- 7695-2286-06/05, IEEE 2005*.
- [15] Rafiul H.Md., Baikunth N. and Michael K., “ A fusion model of HMM, ANN and GA for stock market forecasting”, *Expert systems with Applications*, 2007, pp. 171-180.
- [16] Samarth A., Manoj J. and Pillai G.N., “Preduction using Adaptive NeuroFuzzy Inference System (ANFIS)”, *proceeding of the international Multiconference of engineers and computer scientists*, 2010, Vol I.
- [17] Suresh B.M., Geethanjali N. and Sathyanarayana B., “ Forecasting of Indian Stock Market Index Using Data Mining & Artificial Neural Nework”, *International journal of advance engineering & application*, 2011.
- [18] Wang, Y.-F., Shihmin C. and Hsu M-H., “Incorporating the Markov chain concepts into fuzzy stochastic prediction of stock indexes”, *Applied Soft Computing*, 2010, pp.613-617.
- [19] Wang, Z., Sun, Y., Stockli, P. (2014). “Functional Principal Components Analysis of Shanghai Stock Exchange 50 Index”. *Discrete Dynamics in Nature and Society* Volume 2014 (2014), Article ID 365204, 7 pages.
- [20] Wang, Y., Choi, I-C., (2013).” Market Index and Stock Price direction prediction using Machine Learning Techniques: An empirical study on the KOSPI and HSI”. *Science Direct*. Pages 1-13.
- [21] <http://www.forecasts.org/stock-index/index.htm>.
- [22] <http://www.eviews.com>.

ANALYZING AND FORECASTING A STOCK MARKET INDEX BY USING LEADING INDICATORS

Thanh Do Van, Hai Nguyen Minh

ABSTRACT— *The goal of this paper is to propose a methodology of combining the leading indicator method and correlation coefficients between the stock market index of a stock exchange and the stock transaction data variables to build a forecast model of the stock market index. Applying the proposed methodology, the paper practised to build a model forecasting the stock market index of Ho Chi Minh City’s stock exchange: VNINDEX. Comparing the forecasted results by using the built model with real statistical data shows good prospects of the proposed methodology for building forecast models of stock market index.*