

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC HOA SEN  
KHOA KHOA HỌC VÀ CÔNG NGHỆ**

# **KHÓA LUẬN TỐT NGHIỆP**

*Tên đề tài:*

**Xây Dựng Hệ Thống Phân Tích Và Dự Báo  
Thị Trường Chứng Khoán**

**Giảng viên hướng dẫn : Nguyễn Thị Phương Liên**  
**Nhóm sinh viên thực hiện : Nguyễn Minh Huy – 081667**  
**Nguyễn Anh Khoa – 080835**  
**Lương Ngọc Lan – 080838**  
**Lớp : QL081**

**Tháng 12/2011**

## **PHIẾU GIAO ĐỀ TÀI KHÓA LUẬN TỐT NGHIỆP**

- 1. Mỗi sinh viên phải viết riêng một báo cáo**
- 2. Phiếu này phải dán ở trang đầu tiên của báo cáo**

### **1. Họ và tên sinh viên/ nhóm sinh viên được giao đề tài (sĩ số trong nhóm: 3)**

- (1) Nguyễn Minh Huy.....MSSV: 081667..... Khóa: QL081.....  
(2) Nguyễn Anh Khoa.....MSSV: 080835..... Khóa: QL081.....  
(3) Lương Ngọc Lan .....MSSV: 080838..... Khóa: QL081.....  
Chuyên ngành: Kỹ thuật phần mềm..... Khoa: Khoa học và Công nghệ.....

### **2. Tên đề tài : Xây dựng hệ thống phân tích và dự báo thị trường chứng khoán**

#### **3. Các dữ liệu ban đầu:**

- Xây dựng đề tài tiếp theo của đề án chuyên ngành.
- Có kiến thức về xác suất thống kê, kiến thức về cơ sở dữ liệu, và Timeseries database

#### **4. Các yêu cầu đặc biệt:**

- Xây dựng hệ thống phân tích rõ ràng, chi tiết, và tính hữu ích cao
- Xây dựng hệ thống dự báo có tính chính xác cao
- Hệ thống dễ sử dụng, xử lý dữ liệu tốt

#### **5. Kết quả tối thiểu phải có:**

- File báo cáo về đề tài
- Có hệ thống phân tích dữ liệu
- Hệ thống dự báo
- Xây dựng trên hệ thống trên website

Ngày giao đề tài:...../...../..... Ngày nộp báo cáo: ...../...../.....

Họ tên GV hướng dẫn 1: Nguyễn Thị Phương Liên .....Chữ ký: .....

Họ tên GV hướng dẫn 2: ..... Chữ ký: .....

*Ngày tháng năm 2011*

# TRÍCH YẾU

Trong phạm vi khóa luận, nhóm chúng tôi tìm hiểu về lĩnh vực thị trường chứng khoán. Trong đó gồm có các khái niệm cơ bản và những phương pháp phân tích, dự báo được phổ biến rộng rãi hiện nay, chẳng hạn như phân tích biểu đồ giá kỹ thuật, phân tích tài chính cơ bản và dự báo giá bằng mô hình tự hồi quy kết hợp trung bình trượt (*ARIMA*). Sau khi đã nắm vững những kiến thức nền tảng, nhóm bắt tay vào hiện thực một hệ thống phân tích và dự báo thị trường chứng khoán nhằm mục tiêu cung cấp những thông tin hữu ích và cần thiết cho các nhà đầu tư.

Hệ thống phân tích và dự báo thị trường chứng khoán được chúng tôi xây dựng như là một website để cung cấp những thông tin cần thiết và hữu ích về các loại mã chứng khoán. Dựa trên thông tin đầu vào là báo cáo tài chính, giá cổ phiếu niêm yết, hệ thống sẽ tính toán các thông số dẫn xuất thể hiện sức mạnh tài chính, khả năng sinh lợi nhuận của công ty. Xác định xu hướng giá trên thị trường và tín hiệu mua / bán từ đó thiết lập mô hình để dự báo giá cổ phiếu trong các phiên giao dịch sắp tới. Người dùng khi truy cập vào website có thể xem các thông tin chi tiết về công ty và sử dụng các công cụ được hệ thống hỗ trợ để tìm hiểu và phân tích thông tin về các mã chứng khoán. Ngoài ra, hệ thống còn cung cấp thêm bộ lọc cổ phiếu để người dùng có thể nhanh chóng tìm ra mã cổ phiếu với những thông tin nhất định và tin tức liên quan đến thị trường chứng khoán nói chung và từng mã cổ phiếu nói riêng.

Thông qua việc nghiên cứu, tìm hiểu và so sánh giữa các phương pháp phân tích, dự báo. Nhóm chúng tôi đã nhận ra rằng mặc dù giữa những phương pháp này có một số điểm bất đồng. Tuy nhiên, những sự bất đồng đó không đối nghịch, bác bỏ lẫn nhau mà ngược lại mang mối quan hệ tương hỗ, bù trừ. Vì vậy, chúng tôi cho rằng sự kết hợp giữa các phương pháp phân tích sẽ cho cái nhìn tổng quát và đáng tin cậy hơn so với việc chỉ tập trung vào một phương pháp nhất định.

# MỤC LỤC

PHIẾU GIAO ĐỀ TÀI KHÓA LUẬN TỐT NGHIỆP.....	II
TRÍCH YẾU .....	III
MỤC LỤC – BẢNG VÀ HÌNH ẢNH.....	IX
LỜI CẢM ƠN.....	X
NHẬN XÉT CỦA NGƯỜI HƯỚNG DẪN.....	XI
NHẬP ĐỀ.....	1
NỘI DUNG .....	4
<b>I. Cơ sở lý thuyết.....</b>	<b>4</b>
<b>1. Phân tích kỹ thuật .....</b>	<b>4</b>
1.1. <i>Trung bình trượt</i> .....	5
1.1.1. Giới thiệu .....	5
1.1.2. Trung bình trượt giản đơn (Simple moving average) .....	5
1.1.3. Trung bình trượt cấp số nhân (Exponential moving average) .....	6
1.1.4. Cách sử dụng.....	7
1.2. <i>Giải biên độ biến động giá Bollinger (Bollinger bands)</i> .....	7
1.2.1. Giới thiệu .....	7
1.2.2. Công thức.....	7
1.2.3. Cách sử dụng.....	8
1.3. <i>Tỉ lệ thay đổi (Rate of change)</i> .....	10
1.3.1. Giới thiệu .....	10
1.3.2. Công thức.....	10
1.3.3. Cách sử dụng.....	10
1.4. <i>Trung bình trượt hội tụ và phân kỳ (Moving average convergence and divergence - MACD)</i> .....	11
1.4.1. Giới thiệu .....	11
1.4.2. Công thức.....	11

1.4.3.	Cách sử dụng.....	12
1.5.	<i>Chỉ số sức mạnh tương đối (Relative strength index - RSI)</i> .....	12
1.5.1.	Giới thiệu .....	12
1.5.2.	Công thức.....	12
1.5.3.	Cách sử dụng.....	13
1.6.	<i>Chỉ số lưu lượng tiền tệ (Money flow index - MFI)</i> .....	14
1.6.1.	Giới thiệu .....	14
1.6.2.	Công thức.....	14
1.6.3.	Cách sử dụng.....	16
1.7.	<i>Chỉ số stochastic</i> .....	17
1.7.1.	Giới thiệu .....	17
1.7.2.	Công thức.....	17
1.7.3.	Cách sử dụng.....	18
1.8.	<i>Chỉ số Williams %R</i> .....	18
1.8.1.	Giới thiệu .....	18
1.8.2.	Công thức.....	19
1.8.3.	Cách sử dụng.....	19
1.9.	<i>Kết luận</i> .....	20
<b>2.</b>	<b>Phân tích cơ bản</b> .....	<b>20</b>
2.1.	<i>Giới thiệu</i> .....	20
2.2.	<i>Các chỉ số tài chính của công ty niêm yết</i> .....	21
2.2.1.	Chỉ số EPS (Earning per Share).....	21
2.2.2.	Chỉ số P/E (Price to Earning).....	21
2.2.3.	Book value .....	22
2.2.4.	P/B (Price to Book).....	22
2.2.5.	Beta .....	23
2.2.6.	Vốn hóa thị trường – Market Capitalization.....	24
2.2.7.	ROE (Return on Equity) .....	24

2.2.8.	ROA (Return on Assets) .....	25
2.2.9.	Đòn bẩy tài chính – Financial Leverage .....	25
2.2.10.	Tỷ số thanh khoản nhanh – Quick Ratio.....	26
2.2.11.	Hệ số biên lợi nhuận ròng – Net Profit Margin .....	26
2.3.	<i>Kết luận</i> .....	27
<b>3.</b>	<b>Mô hình ARIMA .....</b>	<b>27</b>
3.1.	<i>Cơ sở lý luận</i> .....	28
3.2.	<i>Các dạng mô hình</i> .....	29
3.2.1.	Mô hình tự hồi quy (AR) .....	29
3.2.2.	Mô hình trung bình trượt (MA) .....	30
3.2.3.	Mô hình tự hồi quy trung bình trượt (ARMA) .....	30
3.2.4.	Mô hình tự hồi quy kết hợp trung bình trượt (ARIMA) .....	30
3.3.	<i>Phương pháp lập mô hình ARIMA</i> .....	31
3.3.1.	Nhận dạng .....	31
3.3.2.	Ước lượng .....	31
3.3.3.	Kiểm định.....	32
3.3.4.	Dự báo.....	32
3.4.	<i>Xác định mô hình (Model Identification)</i> .....	32
3.4.1.	Sử dụng SACF và SPACF .....	33
3.4.2.	Độ lệch chuẩn của SACF và SPACF .....	35
3.4.3.	Một vài lưu ý quan trọng .....	35
3.4.4.	Một số công cụ xác định mô hình khác .....	36
3.4.5.	Ước lượng tham số sơ bộ.....	37
3.5.	<i>Ước lượng mô hình (Model Estimation)</i> .....	39
3.5.1.	Hàm hợp lý hóa và hàm tổng bình phương.....	39
3.5.2.	Ước lượng phi tuyến .....	41
3.5.3.	Ước lượng số học của đạo hàm.....	42
3.5.4.	Thuật toán tổng quát cho mô hình có điều kiện.....	43

3.6.	<i>Kiểm định mô hình (Model Diagnostic)</i> .....	44
3.6.1.	Kiểm định tự tương quan (Autocorrelation Check).....	44
3.6.2.	Kiểm định Portmanteau Lack-of-Fit.....	45
3.7.	<i>Dự báo (Forecasting)</i> .....	46
3.7.1.	Tính toán giá trị dự báo.....	46
3.7.2.	Khoảng tin cậy.....	48
3.8.	<i>Kết luận</i> .....	48
<b>II.</b>	<b>Phân tích - Hiện thực</b> .....	<b>49</b>
<b>1.</b>	<b>Phân tích thiết kế hệ thống</b> .....	<b>49</b>
1.1.	<i>Data Flow Diagram (DFD)</i> .....	49
1.2.	<i>Đặc tả Use case</i> .....	50
1.2.1.	Xem tin tức .....	51
1.2.2.	Xem thông tin phân tích và dự báo.....	52
1.2.3.	Phân tích kỹ thuật trên biểu đồ .....	54
1.2.4.	Xem báo cáo tài chính đầy đủ.....	55
1.2.5.	Tìm kiếm mã cổ phiếu .....	56
1.2.6.	Lọc cổ phiếu.....	57
1.2.7.	Nhập liệu.....	58
1.2.8.	Lấy tin .....	59
1.2.9.	Lập mô hình.....	60
1.3.	<i>Domain Model</i> .....	62
<b>2.</b>	<b>Cấu trúc cơ sở dữ liệu</b> .....	<b>63</b>
<b>3.</b>	<b>Thu thập và xử lý dữ liệu</b> .....	<b>69</b>
3.1.	<i>Dữ liệu tài chính</i> .....	69
3.2.	<i>Dữ liệu giá</i> .....	70
<b>III.</b>	<b>Giải pháp công nghệ</b> .....	<b>71</b>
<b>1.</b>	<b>HTML 5</b> .....	<b>71</b>
<b>2.</b>	<b>CSS 3</b> .....	<b>71</b>

3. AJAX .....	72
4. JQuery .....	72
<b>IV. Giới thiệu chương trình.....</b>	<b>73</b>
1. Cấu trúc trang web.....	73
2. Chức năng chính.....	74
3. Giao diện hoàn chỉnh .....	75
4. Hướng dẫn sử dụng.....	76
4.1. Phân tích kỹ thuật .....	76
4.1.1. Thêm chỉ số kỹ thuật.....	76
4.1.2. Bỏ một chỉ số kỹ thuật .....	76
4.2. Tìm kiếm công ty.....	76
4.3. Lọc mã cổ phiếu.....	76
4.4. Xem thông tin chi tiết công ty .....	76
<b>TỔNG KẾT .....</b>	<b>77</b>
<b>I. Đánh giá tổng quan.....</b>	<b>77</b>
<b>II. Hướng phát triển.....</b>	<b>78</b>
<b>TỪ ĐIỂN THUẬT NGỮ .....</b>	<b>79</b>
<b>PHỤ LỤC.....</b>	<b>82</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>85</b>



# MỤC LỤC – BẢNG VÀ HÌNH ẢNH

Bảng 1. Đặc điểm của biểu đồ tương quan tương ứng với các loại mô hình.....	34
Bảng 2. Công thức tính giá trị tham số sơ bộ và miền giá trị .....	34
Hình 1. Mẫu biểu đồ W-Bottom .....	9
Hình 2. Mẫu biểu đồ M-Top. ....	9
Hình 3. Failure swing trên và dưới. ....	13
Hình 4. Phân phối Chi bình phương .....	45
Hình 5. User DFD.....	49
Hình 6. Administrator DFD .....	50
Hình 7. Use Case Diagram.....	50
Hình 8. SD – Xem tin tức .....	51
Hình 9. SD – Xem thông tin và phân tích dự báo.....	53
Hình 10. SD – Phân tích kỹ thuật trên biểu đồ .....	54
Hình 11. SD – Xem báo cáo tài chính đầy đủ.....	55
Hình 12. SD – Tìm kiếm mã cổ phiếu .....	56
Hình 13. SD – Lọc cổ phiếu.....	58
Hình 14. SD – Nhập liệu.....	59
Hình 15. SD – Lấy tin .....	60
Hình 16. SD – Lập mô hình .....	61
Hình 17. Domain Model .....	62
Hình 18. Sơ đồ database .....	63
Hình 19. Cấu trúc trang web .....	73
Hình 20. Giao diện trang phân tích dự báo.....	75

# LỜI CẢM ƠN

Trong khoảng thời gian làm khóa luận tốt nghiệp, từ ngày 09/09/2011 đến ngày 18/12/2011, chúng tôi đã học được rất nhiều điều, từ kiến thức về chứng khoán, phương pháp phát triển một dự án, cách tiếp cận người dùng và cách thức giải quyết những vấn đề nảy sinh.

Nhóm chúng tôi xin chân thành gửi lời cảm ơn đến cô Nguyễn Thị Phương Liên – Giảng viên khoa Khoa Học và Công Nghệ trường Đại Học Hoa Sen. Người đã tận tâm định hướng, theo sát và hỗ trợ nhóm xuyên suốt thời gian vừa qua. Cô đã đóng vai trò là người hướng dẫn, người tư vấn, người đi trước để giúp chúng tôi hoàn tất khóa luận thành công.

Ngoài ra, nhóm cũng xin gửi lời cảm ơn đến toàn bộ tập thể giảng viên khoa Khoa Học và Công Nghệ và các bạn sinh viên cùng khóa QL081, những người đã khích lệ và đưa ra nhiều lời khuyên, lời góp ý bổ ích cho nhóm.

# NHẬN XÉT CỦA NGƯỜI HƯỚNG DẪN

Trong suốt quá trình thực hiện khóa luận tốt nghiệp, tôi đánh giá cao các thành viên trong nhóm. Các bạn làm việc với thái độ nghiêm túc, chịu khó tìm tòi, nghiên cứu, thông minh và sáng tạo. Các thành viên trong nhóm hỗ trợ và phối hợp với nhau tốt, đồng thời mỗi cá nhân có thể phát huy được khả năng và thế mạnh của mình. Đây là một khóa luận tốt nghiệp được hoàn thành xuất sắc.

*Người hướng dẫn ký tên*

*Nguyễn Thị Phương Liên*

# NHẬP ĐỀ

Đầu tư chứng khoán được xem là một hoạt động mang tính rủi ro cao. Tuy nhiên, nó vẫn có sức hấp dẫn lớn đối với giới đầu tư vì tỷ suất sinh lợi khổng lồ mà hiếm có phương án đầu tư nào có thể so sánh được nếu như biết lựa chọn thời điểm xuất nhập thị trường. Hiện nay, cùng với sự phát triển của thị trường chứng khoán tại Việt Nam, nhu cầu sử dụng công cụ phân tích dự báo thị trường để tìm hiểu giá trị thực sự của cổ phiếu, xu hướng biến động giá... cũng ngày một tăng cao. Trước thực tế đó, nhóm đã quyết định nghiên cứu và phát triển **hệ thống phân tích và dự báo thị trường chứng khoán** dựa trên nền tảng công nghệ web. Mọi dữ liệu về chứng khoán sẽ được tập trung và xử lý, người dùng chỉ cần tra cứu và xem kết quả trực tuyến.

Cấu trúc của hệ thống được phân ra làm ba thành phần chính:

- **Phân tích cơ bản:** phân tích và tính toán các chỉ số trong báo cáo tài chính.
- **Phân tích kỹ thuật:** vẽ biểu đồ giá, hỗ trợ công cụ để xác định tín hiệu mua/ bán.
- **Dự báo:** đưa ra mức giá tham khảo cho các phiên giao dịch sắp tới theo mô hình ARIMA.

Hệ thống gồm có các trang nội dung như sau:

- **Trang chủ:** thể hiện tin tức thị trường mới nhất, biến động của các sàn giao dịch và đề cử các mã cổ phiếu có tính thanh khoản cao.
- **Trang trợ giúp:** giải thích một số thuật ngữ chuyên môn và cách sử dụng những công cụ do hệ thống hỗ trợ.
- **Trang phân tích và dự báo:** thể hiện thông số tài chính, biểu đồ kỹ thuật và kết quả dự báo các phiên sắp tới.
- **Bộ lọc cổ phiếu:** hiển thị các thông số để lọc và danh sách kết quả.
- **Trang thông tin mô hình dự báo:** thể hiện các bước tính toán để lập mô hình và các trường thông số liên quan.
- Và thêm vào đó là một **Trang quản lý** dành riêng cho người điều hành để thực hiện các tác vụ như: nhập liệu, lấy tin tức, lập mô hình dự báo.

## Đối tượng sử dụng

- *Nhà đầu tư*: đây là đối tượng sử dụng chính, bao gồm những người quan tâm tới giao dịch chứng khoán. Họ sẽ truy cập vào hệ thống để xem kết quả phân tích và sử dụng các công cụ được hỗ trợ.
- *Người quản trị*: có trách nhiệm quản lý nội dung hệ thống.

## Yêu cầu chức năng

STT	Tên chức năng	Mô tả
1	Tìm kiếm	Tìm theo tên công ty và mã cổ phiếu
2	Bộ lọc cổ phiếu	Tìm kiếm nâng cao dựa trên các thông số tài chính và thông tin về giá.
3	Lấy tin tức	Lấy về RSS của các bản tin chứng khoán và lưu trữ trong cơ sở dữ liệu.
4	Nhập liệu	Tự động đưa thông tin về giá của những mã nhất định vào cơ sở dữ liệu
5	Phân tích báo cáo tài chính	Thể hiện các thông số quan trọng và đưa ra các chỉ số tài chính đã được tính toán.
6	Phân tích kỹ thuật	Chuyển dữ liệu giá thành dạng biểu đồ và cung cấp các công cụ kỹ thuật đi kèm để hỗ trợ phân tích xu hướng biến động.
7	Lập mô hình dự báo	Dựa trên dữ liệu giá chứng khoán (time series), lập mô hình dự báo chung dựa theo phương pháp ARIMA

## Yêu cầu phi chức năng

- Giao diện thân thiện, bố cục sắp xếp hợp lý. Quy trình đơn giản, dễ sử dụng.
- Hiệu năng tốt, cho phép người dùng truy cập nhanh.
- Quy trình đơn giản, dễ sử dụng.

## Bảng phân công công việc

Để có thể hoàn thành đúng mục tiêu đề ra từng tuần cũng như toàn bộ khóa luận, nhóm chúng tôi đã phân chia khối lượng công việc cho từng thành viên như sau:

Thành viên	Công việc
Nguyễn Minh Huy	<ul style="list-style-type: none"><li>• Nghiên cứu và hiện thực phân phân tích kỹ thuật</li><li>• Nghiên cứu và áp dụng công nghệ mới (HTML 5, CSS 3, jQuery)</li><li>• Thiết kế, hiện thực giao diện.</li><li>• Lập trình front-end (một phần)</li><li>• Hiện thực chức năng cung cấp tin tức, nhận định thị trường</li><li>• Hiện thực chức năng tải báo cáo tài chính</li><li>• Tích hợp hệ thống</li></ul>
Lương Ngọc Lan	<ul style="list-style-type: none"><li>• Tìm hiểu về lĩnh vực Phân tích báo cáo tài chính</li><li>• Thu thập và chuẩn hóa dữ liệu tài chính.</li><li>• Xây dựng chức năng Phân tích cơ bản</li><li>• Xây dựng module cho chức năng Filter</li><li>• Format báo cáo nhóm</li></ul>
Nguyễn Anh Khoa	<ul style="list-style-type: none"><li>• Tìm hiểu lý thuyết mô hình ARIMA</li><li>• Hiện thực chức năng lập mô hình dự báo giá</li><li>• Xây dựng chức năng nhập liệu, cập nhật giá cổ phiếu mới nhất</li><li>• Phát triển chức năng tìm kiếm theo mã cổ phiếu và tên công ty</li><li>• Thiết lập bộ lọc cổ phiếu dựa trên các tính chất cơ bản của mã cổ phiếu</li></ul>

Mỗi thành viên trong nhóm sẽ chịu trách nhiệm chính cho phân công việc của mình và thảo luận, hỗ trợ các thành viên khác khi cần thiết.

# NỘI DUNG

## I. Cơ sở lý thuyết

### 1. Phân tích kỹ thuật

Phân tích kỹ thuật là một trong hai trường phái phân tích chứng khoán. Phân tích kỹ thuật bỏ qua tất cả những dữ liệu thể hiện giá trị hoặc sức mạnh của một công ty như doanh thu, lợi nhuận, ... và chỉ tập trung vào diễn biến của giá cổ phiếu trên thị trường. Mặc dù phân tích kỹ thuật nghiên cứu nhiều phạm vi như xu hướng giá, các mẫu biểu đồ giá, ... nhưng bản chất của phân tích kỹ thuật là nghiên cứu mức cung – cầu của thị trường để có thể dự đoán xu hướng giá trong tương lai.

Nền tảng của phân tích kỹ thuật dựa trên ba giả định sau:

- Giá cổ phiếu thể hiện tất cả. Giả định này cho rằng sức mạnh của một công ty hay những diễn biến kinh tế đều được thể hiện qua giá cổ phiếu. Do đó chỉ cần phân tích diễn biến của giá là đủ.
- Giá biến động theo xu hướng. Có nghĩa là nếu trong hiện tại giá đang tăng thì trong tương lai giá cũng sẽ tăng.
- Lịch sử thường lặp lại. Các diễn biến giá thường lặp lại và khi được thể hiện trên biểu đồ sẽ tạo thành các mẫu biểu đồ. Các nhà phân tích dựa vào các mẫu biểu đồ này để dự đoán xu hướng giá trong tương lai.

Trong phân tích kỹ thuật có nhiều lĩnh vực nghiên cứu. Tuy nhiên để phù hợp với phạm vi của đề án, lĩnh vực phân tích dựa vào các chỉ số kỹ thuật được đặt làm trọng tâm.

Các chỉ số phân tích kỹ thuật là kết quả của các phép tính dựa trên giá và khối lượng giao dịch của cổ phiếu. Các chỉ số này được thể hiện trên biểu đồ để cung cấp các thông tin hữu ích về giá cổ phiếu cho người dùng.

Chỉ số phân tích kỹ thuật được chia làm hai loại: chỉ số xu hướng giá và chỉ số dao động giá

- Chỉ số xu hướng giá: như tên gọi, các chỉ số này cho biết xu hướng hiện tại của giá. Các chỉ số này dịch chuyển cùng hướng với đường giá nên thường được vẽ chung với biểu đồ giá. Có ba chỉ số xu hướng giá phổ biến:
  - Trung bình trượt giản đơn (Simple moving average).
  - Trung bình trượt cấp số nhân (Exponential moving average).
  - Giải biên độ biến động giá Bollinger (Bollinger bands).

- Chỉ số dao động giá: các chỉ số này thể hiện tốc độ hay mức độ biến động của giá. Các chỉ số này dao động quanh một mức trung tâm, thường là 0 hoặc 50 và không di chuyển cùng hướng với đường giá nên biểu đồ thường được vẽ riêng. Có sáu chỉ số dao động giá thông dụng:
  - Chỉ số lưu lượng tiền tệ (Money Flow Index).
  - Trung bình trượt hội tụ và phân kỳ (Moving Average Convergence and Divergence).
  - Chỉ số tỷ lệ thay đổi giá (Rate of change).
  - Chỉ số sức mạnh tương đối (Relative strength index).
  - Chỉ số Stochastic chậm.
  - Chỉ số Stochastic nhanh.
  - Chỉ số Williams %R.

## **1.1. Trung bình trượt**

### **1.1.1. Giới thiệu**

Trung bình trượt là chỉ số phản ánh xu hướng giá, một công cụ được sử dụng phổ biến trong phân tích kỹ thuật. Mục đích của nó là để theo dõi diễn biến của xu hướng giá. Trung bình trượt còn là một công cụ làm trơn. Bằng cách lấy trung bình của dữ liệu, đường giá sẽ trở nên mượt hơn, từ đó có thể dễ dàng quan sát xu hướng giá.

Đường trung bình trượt phản ánh xu hướng giá chậm hơn diễn biến của thị trường. Độ trễ này có thể được khắc phục bằng cách sử dụng các khung thời gian ngắn để tính trung bình trượt nhưng không thể loại bỏ hoàn toàn.

Trong phân tích kỹ thuật, hai loại trung bình trượt được sử dụng rộng rãi là trung bình trượt giản đơn và trung bình trượt cấp số nhân.

### **1.1.2. Trung bình trượt giản đơn (Simple moving average)**

Để tính trung bình trượt giản đơn của một chuỗi dữ liệu với khung thời gian là x ngày ta làm như sau:

- Tính giá trị trung bình của x ngày kể từ ngày đang tính.
- Thực hiện lần lượt như trên với tất cả các giá trị trong chuỗi dữ liệu nếu có đủ dữ liệu cần thiết.



**Ví dụ:** Với chuỗi dữ liệu Phụ lục – Bảng 1, để tính trung bình trượt ba ngày của chuỗi này ta làm như sau:

- Trung bình trượt ngày 1:  $(10 + 5 + 7) / 3 = 7.3$
- Trung bình trượt ngày 2:  $(5 + 7 + 14) / 3 = 8.6$
- Trung bình trượt ngày 3:  $(7 + 14 + 3) / 3 = 8$
- Ở các ngày 4 và 5, do không có dữ liệu của hai ngày liên tục sau nó nên không thực hiện tính trung bình trượt.

Đây là dạng trung bình trượt thông dụng nhất được sử dụng trong phân tích kỹ thuật. Tuy nhiên có một số nhà phân tích cho rằng trung bình trượt giản đơn có điểm yếu là tất cả các ngày trong khung thời gian được tính có tầm ảnh hưởng như nhau và họ tin rằng các ngày càng gần với hiện tại nên có tầm ảnh hưởng lớn hơn để phản ánh đúng diễn biến của thị trường.

### **1.1.3. Trung bình trượt cấp số nhân (Exponential moving average)**

Trung bình trượt cấp số nhân loại bỏ được khuyết điểm của của trung bình trượt giản đơn, những ngày càng gần với hiện tại thì sẽ có mức quan trọng cao hơn so với những ngày trước đó.

Để tính trung bình trượt cấp số nhân của một chuỗi dữ liệu với khung thời gian là x ngày ta dùng công thức sau:

- $Multiplier = 2 / (x + 1)$ . Đây là trọng số áp dụng lên đường trung bình trượt, các ngày càng gần với hiện tại thì ảnh hưởng của nó lên đường trung bình trượt càng lớn.
- $EMA_{\text{hôm nay}} = (\text{Giá đóng của hôm nay} - EMA_{\text{hôm qua}}) * Multiplier + EMA_{\text{hôm qua}}$

**Ví dụ:** Với dữ liệu từ Bảng 1 (Phụ lục), để tính trung bình trượt cấp số nhân ba ngày của chuỗi này ta làm như sau:

- $Multiplier = 2 / (3 + 1) = 0.5$
- Trung bình trượt ngày 3:  $(7 + 14 + 3) / 3 = 8$ . Vì trước ngày này không đủ dữ liệu để tính trung bình trượt cấp số nhân nên ta dùng trung bình trượt giản đơn.
- Trung bình trượt ngày 2:  $(5 - 8) * 0.5 + 8 = 6.5$
- Trung bình trượt ngày 1:  $(10 - 6.5) * 0.5 + 6.5 = 8.25$

#### 1.1.4. Cách sử dụng

Đường trung bình trượt khi vẽ chung với đường giá sẽ tạo ra các tín hiệu mua, bán. Khi đường giá vượt lên trên đường trung bình trượt sẽ tạo tín hiệu mua và khi đường giá nằm phía dưới đường trung bình trượt sẽ tạo tín hiệu bán.

Với những khung thời gian ngắn, đường trung bình trượt sẽ dao động nhiều và cắt đường giá thường xuyên hơn, tạo ra nhiều tín hiệu mua, bán hơn. Tuy nhiên, việc này cũng tạo ra nhiều tín hiệu giả hơn do những biến động ngẫu nhiên của giá chứng khoán. Dù vậy, lợi thế của đường trung bình trượt có khung thời gian ngắn là nó tạo ra các tín hiệu mua bán sớm và có thể có lợi cho nhà đầu tư. Ngược lại, với khung thời gian dài, đường trung bình trượt tạo ra các tín hiệu có độ trễ cao nhưng đáng tin cậy.

Theo các nhà phân tích, đường trung bình trượt với khung thời gian dài hữu dụng trong trường hợp xu hướng giá tiếp tục theo hướng hiện tại còn khi xu hướng giá đảo chiều, đường trung bình trượt sẽ không thể hiện kịp và khi đó sử dụng đường trung bình trượt với khung thời gian ngắn sẽ phù hợp hơn. Ngoài ra, các nhà phân tích cũng sử dụng cùng lúc hai hoặc ba đường trung bình trượt để tạo các tín hiệu tốt hơn.

Khung thời gian phổ biến khi sử dụng hai đường trung bình trượt là 5 và 20 ngày hoặc 10 và 50 ngày còn với ba đường trung bình trượt là 4, 9 và 18 ngày. Các tín hiệu mua được tạo ra khi đường trung bình trượt có khung thời gian ngắn vượt lên đường trung bình trượt có khung thời gian dài và tín hiệu bán được tạo ra khi đường trung bình trượt có khung thời gian ngắn nằm dưới đường trung bình trượt có khung thời gian dài.

### 1.2. Giải biên độ biến động giá Bollinger (Bollinger bands)

#### 1.2.1. Giới thiệu

Giải biên độ Bollinger được phát triển bởi John Bollinger. Nó bao gồm ba thành phần: đường trung bình trượt ở giữa, biên trên và biên dưới. Hai biên tạo thành một dải biên độ bao quanh đường giá. Có thể sử dụng dải biên độ này để ước lượng giá trị mà giá có thể đạt được.

#### 1.2.2. Công thức

Để tính giải biên độ Bollinger của một chuỗi dữ liệu với khung thời gian là  $x$  ngày và biến động với  $y$  lần độ lệch chuẩn ta làm như sau:

- Tính trung bình trượt (giản đơn hoặc cấp số nhân) của chuỗi dữ liệu với khung thời gian là  $x$  ngày. Kí hiệu  $ma$
- Tính độ lệch chuẩn của  $x$  ngày của dữ liệu giá. Kí hiệu  $dev$ .

- Biên trên =  $ma + y * dev.$
- Biên dưới =  $ma - y * dev.$

**Ví dụ:** Với dữ liệu từ Bảng 1 (Phụ lục), để tính dải biên độ Bollinger ba ngày và biên động với hai lần độ lệch chuẩn của chuỗi này ta làm như sau:

- Tính trung bình trượt của chuỗi dữ liệu:

Ngày 1: 7.3

Ngày 2: 8.6

Ngày 3: 8

- Tính độ lệch chuẩn 3 ngày của chuỗi dữ liệu:

Ngày 1: 2.05

Ngày 2: 3.85

Ngày 3: 4.54

- Dữ liệu biên trên:

Ngày 1:  $7.3 + 2 * 2.05 = 11.4$

Ngày 2:  $8.6 + 2 * 3.85 = 16.3$

Ngày 3:  $8 + 2 * 4.45 = 17$

- Dữ liệu biên dưới:

Ngày 1:  $7.3 - 2 * 2.05 = 3.2$

Ngày 2:  $8.6 - 2 * 3.85 = 0.9$

Ngày 3:  $8 - 2 * 4.45 = -1$

### 1.2.3. Cách sử dụng

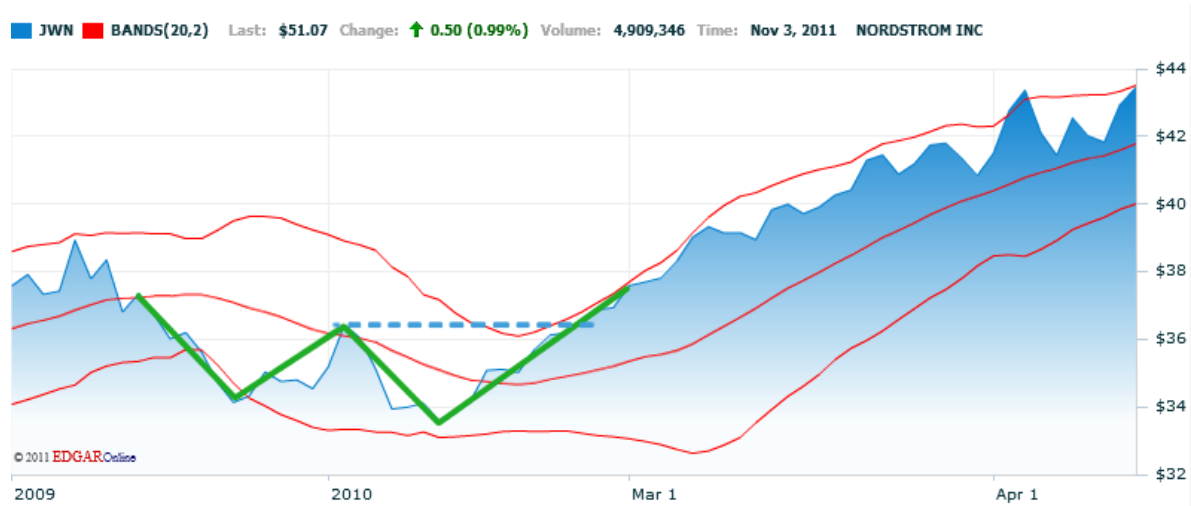
Giải biên độ Bollinger được dùng để xác định các mẫu biểu đồ W-Bottom và M-Top.

W-Bottom được hình thành khi giá đang trong xu hướng giảm. Khi nó xuất hiện báo hiệu giá sẽ đảo chiều từ xu hướng giảm sang tăng. Nó bao gồm hai đỉnh dưới. Có bốn bước để xác nhận W-Bottom:

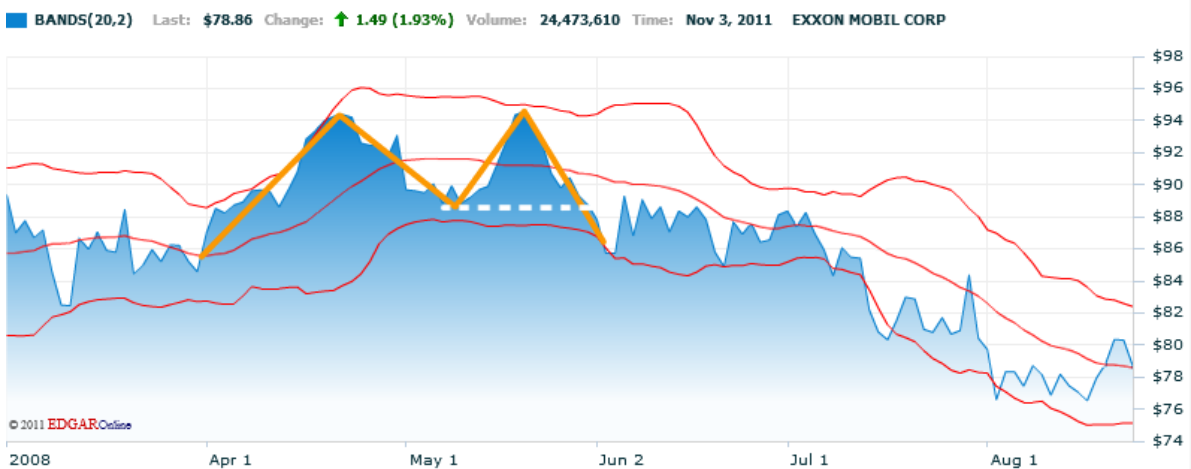
- Xác nhận có một đỉnh dưới (1) trên đường giá. Đỉnh này thường nằm dưới biên dưới của dải biên độ Bollinger.
- Đường giá sau đó sẽ vượt lên đến đường trung bình trượt ở giữa (2).

- Đường giá giảm và tạo thành đỉnh dưới thứ 2, đỉnh này thấp hơn đỉnh (1) và nằm trên biên dưới.
- Cuối cùng, đường giá vượt qua mức giá ở bước (2).

Ngược lại với W-Bottom là M-Top. Nó được hình thành khi giá đang trong xu hướng tăng và khi xuất hiện báo hiệu giá sẽ đảo chiều sang xu hướng giảm.



Hình 1. Mẫu biểu đồ W-Bottom



Hình 2. Mẫu biểu đồ M-Top.

Ngoài ra ta có thể dùng biên trên và biên dưới như là mức giá mục tiêu. Có nghĩa là sau khi giá chạm biên dưới, đang vượt lên và đã qua khỏi đường trung bình trượt ở giữa thì mức giá mà nó vươn tới là mức giá của biên trên. Điều tương tự cũng được áp dụng cho biên dưới khi giá sau khi chạm biên trên trượt xuống dưới đường trung bình trượt.

### 1.3. Tỷ lệ thay đổi (Rate of change)

#### 1.3.1. Giới thiệu

Tỷ lệ thay đổi là chỉ số cơ bản nhất trong các chỉ số dao động giá. Nó đo tỷ lệ thay đổi của giá hiện tại so với giá  $x$  ngày trước đó.

#### 1.3.2. Công thức

Để tính tỷ lệ thay đổi của giá cổ phiếu trong khoảng thời gian  $x$  ngày ta dùng công thức

$$\text{sau: } R = 100 * \frac{V - V_x}{V_x}$$

Với  $V$  là giá hiện tại của cổ phiếu và  $V_x$  là giá của cổ phiếu  $x$  ngày trước.

**Ví dụ:** Với dữ liệu từ Bảng 1 (Phụ lục), để tính tỷ lệ thay đổi của ba ngày ta làm như sau:

- Tỷ lệ thay đổi ngày 1:  $100 * (10 - 7) / 7 = 42.85$
- Tỷ lệ thay đổi ngày 2:  $100 * (5 - 14) / 14 = -64.28$
- Tỷ lệ thay đổi ngày 3:  $100 * (7 - 3) / 3 = 133.33$
- Ở các ngày 4 và 5, do không có dữ liệu của ba ngày sau đó nên không thực hiện tính tỷ lệ thay đổi.

Sau khi tính và thể hiện lên biểu đồ đường tỷ lệ thay đổi sẽ dao động quanh mức 0. Nếu giá hiện tại cao hơn giá trước đó thì đường tỷ lệ thay đổi sẽ nằm trên mức 0.

Ngược lại, nếu giá hiện tại thấp hơn giá trước đó thì đường tỷ lệ thay đổi sẽ nằm dưới mức 0.

#### 1.3.3. Cách sử dụng

Dựa vào đường tỷ lệ thay đổi ta có thể nhận định xu hướng của giá. Nếu đường tỷ lệ thay đổi luôn ở trên mức 0 có nghĩa là giá đang tăng, ngược lại, khi nó ở dưới mức 0 có nghĩa là giá đang giảm.

ROC được dùng tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức 0, xác định hiện tượng phân kỳ giữa đường ROC và đường giá.

Khi đường tỷ lệ thay đổi vượt lên mức 0 có nghĩa là xu hướng giảm đã chậm lại và chuyển sang xu hướng tăng giá, đây là báo hiệu cho tín hiệu mua. Ngược lại khi đường tỷ lệ thay đổi trượt xuống dưới mức 0 thì đó là tín hiệu bán. Với các tín hiệu mua, bán quá mức, do đường tỷ lệ thay đổi không dao động trong một khoảng cố định nên ta phải dựa vào dữ liệu

trong quá khứ để có thể đưa ra giới hạn phù hợp. Hiện tượng phân kỳ có thể sử dụng với ROC nhưng không đáng tin cậy như các chỉ số dao động giá khác.

#### 1.4. Trung bình trượt hội tụ và phân kỳ (Moving average convergence and divergence - MACD)

##### 1.4.1. Giới thiệu

MACD là một chỉ số dao động giá được phát triển bởi Gerald Appel. MACD là chỉ số dao động giá đơn giản nhất và được sử dụng rộng rãi nhất trong các chỉ số dao động giá. Dù là chỉ số dao động giá nhưng MACD cũng có một phần tính chất của các chỉ số xu hướng giá do nó là kết quả của phép trừ hai đường trung bình trượt. Ngoài ra, biểu đồ của đường MACD còn có thêm một đường tín hiệu là trung bình trượt của đường MACD để tạo các tín hiệu mua bán.

##### 1.4.2. Công thức

Công thức tính MACD nhận tham số đầu vào lần lượt là hai khung thời gian để tính trung bình trượt và khung thời gian để tính đường tín hiệu.

Giả sử tham số đầu vào theo thứ tự trên lần lượt là x, y, z.

Nếu  $x > y$ :  $MACD = EMA(y) - EMA(x)$  (\*)

Nếu  $x < y$ :  $MACD = EMA(x) - EMA(y)$  (\*)

Đường tín hiệu = trung bình trượt cấp số nhân với khung thời gian là z ngày của MACD.

(\*)  $EMA(x)$  có nghĩa là trung bình trượt cấp số nhân với khung thời gian là x ngày.

**Ví dụ:** có chuỗi dữ liệu sau:

Ngày	Giá	Ngày	Giá	Ngày	Giá
1	89.5	5	89.5	9	81.5
2	89.5	6	89.5	10	82
3	90	7	89.5	11	81
4	90	8	85.5		

MACD với khung thời gian của các đường trung bình trượt là 8, 5 và khung thời gian của đường tín hiệu là 3. Ta lần lượt tính trung bình trượt cấp số nhân 8 ngày và 5 ngày của chuỗi giá. Sau đó, lấy hiệu của trung bình trượt 5 ngày trừ đi trung bình trượt 8 ngày ta

được chỉ số MACD. Cuối cùng tính trung bình trượt 3 ngày của MACD ta được dữ liệu của đường tín hiệu (Xem Phụ lục – bảng 2).

### 1.4.3. Cách sử dụng

Một điểm đặc biệt của MACD là nó mang một phần tính chất của các chỉ số xu hướng giá nên đường sẽ thể hiện một phần xu hướng giá, nó tăng khi giá tăng và giảm khi giá giảm.

MACD được dùng tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức 0, các điểm giao với đường tín hiệu, xác định hiện tượng phân kỳ giữa đường MACD và đường giá.

Các tín hiệu mua được tạo ra khi đường MACD vượt lên mức 0 hoặc vượt lên trên đường tín hiệu và tín hiệu bán được tạo ra khi đường MACD trượt xuống dưới mức 0 hoặc trượt xuống dưới đường tín hiệu.

## 1.5. Chỉ số sức mạnh tương đối (Relative strength index - RSI)

### 1.5.1. Giới thiệu

RSI là một chỉ số dao động giá được phát triển bởi J. Welles Wilder. Nó dùng để đo vận tốc và mức độ thay đổi của giá. Nó có nhiều ưu điểm hơn các chỉ số tương tự khác như tỉ lệ thay đổi (ROC), ... Ở các chỉ số như tỉ lệ thay đổi thì biểu đồ của nó có thể dễ biến động thất thường. Ví dụ như khi tính tỉ lệ thay đổi với khung thời gian là mười ngày thì nếu giá của mười ngày trước tăng hoặc giảm đột ngột thì tỉ lệ thay đổi sẽ bị ảnh hưởng nhiều dù cho giá hiện tại ít thay đổi. Ngoài ra các chỉ số này cũng không có giới hạn giao động cụ thể. RSI giải quyết được cả hai vấn đề trên, tạo nên một biểu đồ ổn định hơn và giới hạn biên độ giao động từ 0 đến 100.

### 1.5.2. Công thức

Để tính RSI của giá cổ phiếu trong khoảng thời gian x ngày ta dùng công thức sau:

$$RSI = 100 - \frac{100}{1 + RS}$$

$$RS = \frac{\text{Trung bình các mức tăng giá trong } x \text{ ngày}}{\text{Trung bình các mức giảm giá trong } x \text{ ngày}}$$

**Ví dụ:** Với dữ liệu từ Bảng 1, để tính RSI của ba ngày ta làm như sau:

RSI ngày 1:

Mức tăng của ngày 1:  $10 - 5 = 5$ .

Mức tăng của ngày 2: 0

Mức tăng của ngày 3: 0

Trung bình các mức tăng:  $(5 + 0 + 0) / 3 = 1.67$

Mức giảm của ngày 1: 0

Mức giảm của ngày 2:  $7 - 5 = 2$

Mức giảm của ngày 3:  $14 - 7 = 7$

Trung bình các mức giảm:  $(0 + 2 + 7) / 3 = 3$

$RS = 1.67 / 3 = 0.55$ .

$$RSI = 100 - \frac{100}{1 + 0.55} = 35.4$$

Tương tự RSI cho ngày 2 và 3 là: 54.95

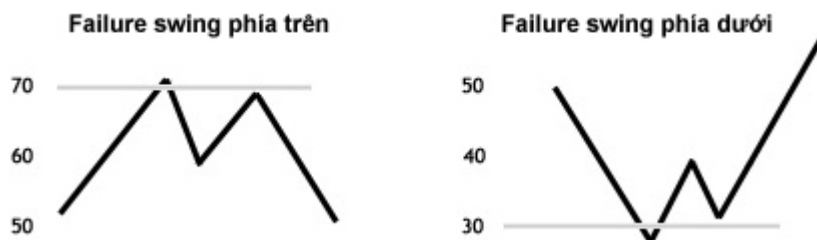
Ở các ngày 4 và 5, do không có dữ liệu của ba ngày sau nó nên không thực hiện tính RSI.

### 1.5.3. Cách sử dụng

RSI được dùng tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức 50, xác định hiện tượng phân kỳ giữa đường RSI và đường giá, và xác định failure swing.

Các điểm mua quá mức xảy ra khi đường RSI nằm trên mức 70 và các điểm bán quá mức xảy ra khi đường RSI nằm dưới mức 30. Các tín hiệu mua cũng được tạo ra khi đường RSI vượt lên mức 50 và tín hiệu bán được tạo ra khi đường RSI trượt xuống dưới mức 50.

Failure swing là một hiện tượng báo hiệu đảo chiều, nó có thể xảy ra ở biên trên hoặc biên dưới của biểu đồ RSI. Nó xảy ra ở phía trên khi đường RSI vượt qua mức 70, sau đó trượt xuống dưới mức 70 rồi tăng lên lại nhưng không vượt qua 70 và cuối cùng đảo chiều hoàn toàn. Nếu xảy ra ở biên dưới thì đầu tiên đường RSI trượt xuống dưới mức 30, sau đó vượt lên trên mức 30 rồi giảm xuống nhưng không thấp hơn mức 30 và cuối cùng là đảo chiều hoàn toàn.



Hình 3. Failure swing trên và dưới.



## 1.6. Chỉ số lưu lượng tiền tệ (Money flow index - MFI)

### 1.6.1. Giới thiệu

MFI là một chỉ số dao động giá được phát triển bởi Gene Quong và Avrum Soudack. MFI sử dụng cả giá và khối lượng giao dịch trong việc tính toán do đó ta có thể xem MFI tương đương với RSI nhưng có thêm ảnh hưởng của khối lượng giao dịch. Và các nhà phân tích nhận định rằng, khối lượng thể hiện trước xu hướng của giá, do đó đường MFI sẽ tạo các tín hiệu mua bán sớm hơn đường RSI.

### 1.6.2. Công thức

Để tính MFI của giá cổ phiếu trong khoảng thời gian x ngày ta thực hiện tính các bước sau:

- Giá trung bình =  $\frac{(\text{High} + \text{Low} + \text{Close})}{3}$
- Tiền giao dịch = Giá trung bình \* Khối lượng giao dịch
- Dòng tiền vào = Tổng lượng tiền vào trong x ngày.
- Dòng tiền ra = Tổng lượng tiền ra trong x ngày.
- $$\text{MFI} = 100 - \frac{100}{1 + \frac{\text{Dòng tiền vào}}{\text{Dòng tiền ra}}}$$

*Trong đó:*

High/ Low/ Close: lần lượt là giá cao nhất/ thấp nhất/ đóng cửa trong một ngày giao dịch của cổ phiếu.

Tiền giao dịch được thêm vào tổng lượng tiền vào khi nó lớn hơn tiền giao dịch của ngày trước đó.

Tiền giao dịch được thêm vào tổng lượng tiền ra khi nó nhỏ hơn tiền giao dịch của ngày trước đó.

Nếu tiền giao dịch bằng với ngày trước thì không được thêm vào tổng lượng tiền.

**Ví dụ:** có chuỗi dữ liệu sau:

Ngày	High	Low	Close	Khối lượng giao dịch
1	130.06	128.47	128.58	21700
2	129.12	128.06	128.60	75680
3	129.29	127.61	127.93	58800
4	128.47	127.60	128.11	80540
5	128.09	127.00	127.60	73000

Để tính MFI cho chuỗi dữ liệu trên với khung thời gian là ba ngày ta làm như sau:

Đầu tiên ta tính giá trung bình:  $\text{Giá trung bình} = \frac{(\text{High} + \text{Low} + \text{Close})}{3}$

Ngày	High	Low	Close	Giá trung bình
1	130.06	128.47	128.58	89.83
2	129.12	128.06	128.60	90
3	129.29	127.61	127.93	89.83
4	128.47	127.60	128.11	89.33
5	128.09	127.00	127.60	90

Sau đó ta tính lượng tiền giao dịch:  $\text{Tiền giao dịch} = \text{Giá trung bình} * \text{Khối lượng giao dịch}$

Ngày	Giá trung bình	Khối lượng giao dịch	Tiền giao dịch
1	89.83	21700	1949383.333
2	90	75680	6811200
3	89.83	58800	5282200
4	89.33	80540	7194906.667
5	90	73000	6570000

Tiếp theo ta tính dòng tiền vào và ra: ta xét trong ba ngày, khi giá trung bình của một ngày tăng so với ngày trước đó, ta thêm tiền giao dịch tại dòng tiền vào. Ngược lại ta thêm vào dòng tiền ra. Trong bảng dữ liệu, những ngày có giá trung bình tăng được tô màu xanh, ngược lại được tô màu đỏ.

Ngày	Giá trung bình	Tiền giao dịch	Dòng tiền vào	Dòng tiền ra
1	89.83	1949383.333	12093400	1949383.333
2	90	6811200	12093400	7194906.667
3	89.83	5282200	11852200	7194906.667
4	89.33	7194906.667	12847600	7194906.667
5	90	6570000	12847600	9749350

Cuối cùng ta tính MFI bằng công thức: 
$$MFI = 100 - \frac{100}{1 + \frac{\text{Dòng tiền vào}}{\text{Dòng tiền ra}}}$$

Ngày	Dòng tiền vào	Dòng tiền ra	MFI
1	12093400	1949383.333	86.11825528
2	12093400	7194906.667	62.69809066
3	11852200	7194906.667	62.22572387
4	12847600	7194906.667	
5	12847600	9749350	

### 1.6.3. Cách sử dụng

Vì MFI được xem như RSI có thêm ảnh hưởng của khối lượng giao dịch nên cách sử dụng MFI cũng tương tự như RSI đó là dùng để tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức 50, xác định hiện tượng phân kỳ giữa đường MFI và đường giá, và xác định failure swing.

Các điểm mua quá mức xảy ra khi đường MFI nằm trên mức 80 và các điểm bán quá mức xảy ra khi nó nằm dưới mức 20. Các tín hiệu mua cũng được tạo ra khi đường MFI vượt lên mức 50 và tín hiệu bán được tạo ra khi nó trượt xuống dưới mức 50.

## 1.7. Chỉ số stochastic

### 1.7.1. Giới thiệu

Stochastic là một chỉ số dao động giá được phát triển bởi George Lane. Nó thể hiện khoảng cách giữa mức giá hiện tại với mức giá cao và thấp nhất trong một khoảng thời gian cho trước.

### 1.7.2. Công thức

Biểu đồ của chỉ số stochastic gồm hai đường %K và %D. Mỗi đường có một khung thời gian riêng. Để tính chỉ số stochastic của giá cổ phiếu với khung thời gian %K, %D tương ứng là x, y ngày ta làm như sau.

$$\%K = \frac{\text{Giá đóng cửa} - \text{Giá high cao nhất trong } x \text{ ngày}}{\text{Giá high cao nhất trong } x \text{ ngày} - \text{Giá low cao nhất trong } x \text{ ngày}} * 100$$

%D là trung bình trượt y ngày của đường %K.

Giá high là giá cao nhất trong một ngày giao dịch của cổ phiếu.

Giá low là giá thấp nhất trong một ngày giao dịch của cổ phiếu.

**Ví dụ:** Với chuỗi dữ liệu Phụ lục – bảng 3, để tính chỉ số stochastic với khung thời gian %K, %D tương ứng là ba ngày ta làm như sau.

- Tính %K:

*Ngày 1:*

Giá high cao nhất trong ba ngày kể từ ngày 1: 129.14

Giá low thấp nhất trong ba ngày kể từ ngày 1: 126.9

Giá đóng cửa: 128.27

$$\%K = \frac{128.27 - 129.14}{129.14 - 126.9} * 100 = 61.33$$

Tương tự ta tính cho các ngày còn lại:

Ngày 2: 80

Ngày 3: 39.77

Ngày 4: 26.09

Ngày 5: 30.59

Ở các ngày còn lại do không có đủ dữ liệu nên không thực hiện tính toán.

- Tính %D: đường %D là trung bình trượt ba ngày của đường %K:

Ngày 1: 60.37

Ngày 2: 48.62

Ngày 3: 32.15

Ở các ngày còn lại do không có đủ dữ liệu nên không thực hiện tính toán.

Chỉ số stochastic được chia thành hai loại: nhanh và chậm. Công thức được trình bày ở trên là chỉ số stochastic nhanh. Đối với chỉ số stochastic chậm thì đường %K được làm trơn bằng trung bình trượt rồi mới tính đường %D như công thức trên.

### **1.7.3. Cách sử dụng**

Chỉ số Stochastic được dùng tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức 50, xác định hiện tượng phân kỳ giữa đường %D và đường giá.

Các điểm mua quá mức xảy ra khi đường Stochastic nằm trên mức 80 và các điểm bán quá mức xảy ra khi nó nằm dưới mức 20. Các tín hiệu mua cũng được tạo ra khi đường Stochastic vượt lên mức 50 và tín hiệu bán được tạo ra khi nó trượt xuống dưới mức 50.

Đối với phân kỳ, George Lane nhấn mạnh rằng hiện tượng này khi xảy ra ở đường %D và đường giá là một tín hiệu mua bán đáng tin cậy. Và ngoài ra thông thường George Lane còn xác định một loại phân kỳ khác giữa đường %K và đường giá để giúp chọn thời điểm tốt tham gia giao dịch. Hiện tượng phân kỳ của George Lane xảy ra khi giá tạo lower high còn đường stochastic tạo higher high. Khi hiện tượng này xảy ra, dù giá giảm nhưng đà tăng của nó vẫn mạnh báo hiệu giá sẽ tăng trong tương lai do đó đây là thời điểm thích hợp để tham gia giao dịch.

## **1.8. Chỉ số Williams %R**

### **1.8.1. Giới thiệu**

Williams %R là một chỉ số dao động giá được phát triển bởi Larry Williams. Chỉ số này thực ra là dạng đảo ngược của chỉ số Stochastic. Do đó đồ thị của đường Stochastic và đồ thị của đường Williams %R hoàn toàn giống nhau nhưng khác nhau về biên độ dao động: chỉ số Stochastic dao động từ 0 đến 100 còn chỉ số Williams %R dao động từ -100 đến 0.

### 1.8.2. Công thức

Để tính chỉ số Williams %R của giá cổ phiếu với khung thời gian x ngày ta làm như sau.

$$\%R = \frac{\text{Giá high cao nhất trong x ngày} - \text{Giá đóng cửa}}{\text{Giá high cao nhất trong x ngày} - \text{Giá low thấp nhất trong x ngày}} * -100$$

Giá high là giá cao nhất trong một ngày giao dịch của cổ phiếu.

Giá low là giá thấp nhất trong một ngày giao dịch của cổ phiếu.

**Ví dụ:** Với chuỗi dữ liệu từ Phụ lục – bảng 3, để tính chỉ số Williams %R với khung thời gian ba ngày ta làm như sau.

- Ngày 1:

Giá high cao nhất trong ba ngày kể từ ngày 1: 129.14

Giá low thấp nhất trong ba ngày kể từ ngày 1: 126.9

Giá đóng cửa: 128.27

$$\%R = \frac{129.14 - 128.27}{129.14 - 126.9} * -100 = -38.84$$

- Tương tự ta tính cho các ngày còn lại:

Ngày 2: -20.09

Ngày 3: -60

Ngày 4: -73.8

Ngày 5: -69.2

Ở các ngày còn lại do không có đủ dữ liệu nên không thực hiện tính toán.

### 1.8.3. Cách sử dụng

Cách sử dụng chỉ số Williams %R cũng tương tự như chỉ số Stochastic đó là nó dùng để tạo các tín hiệu mua bán bằng cách xác định các điểm mua, bán quá mức, xác định các điểm giao với mức -50, xác định hiện tượng phân kỳ giữa đường Williams %R và đường giá.

Các điểm mua quá mức xảy ra khi đường Williams %R nằm trên mức -20 và các điểm bán quá mức xảy ra nó nằm dưới mức -80. Các tín hiệu mua cũng được tạo ra khi đường Williams %R vượt lên mức -50 và tín hiệu bán được tạo ra khi nó trượt xuống dưới mức -50.

Ngoài ra, đường Williams %R còn sử dụng một tín hiệu báo hiệu đảo chiều khác. Đó là khi việc mua quá mức diễn ra liên tục trong một khoảng thời gian và nếu sau đó nó không thể vượt qua được mức -20 thì sẽ báo hiệu đảo chiều trong tương lai. Tương tự, khi việc bán quá mức diễn ra liên tục trong một khoảng thời gian và nếu sau đó đường Williams %R giữ được mức trên -80 thì sẽ báo hiệu đảo chiều trong tương lai.

## **1.9. Kết luận**

Trong các chỉ số trên, các chỉ số dao động giá đa phần giống nhau về mục tiêu và cách sử dụng, chỉ khác nhau về ý nghĩa và công thức tính. Mặc dù vậy, việc có nhiều chỉ số phân tích kỹ thuật sẽ đáp ứng được nhu cầu sử dụng của phần lớn người dùng vì mỗi nhà đầu tư sẽ có những cách hiểu và thói quen khác nhau trong phân tích kỹ thuật. Khi phân tích nên sử dụng kết hợp cả chỉ số xu hướng giá và các chỉ số dao động giá để đem lại kết quả tốt nhất.

## **2. Phân tích cơ bản**

### **2.1. Giới thiệu**

Phân tích cơ bản là một phương pháp phân tích cổ phiếu. Việc phân tích này được thực hiện dựa trên báo cáo tài chính, báo cáo lãi lỗ, cũng như thị phần cạnh tranh của công ty đó. Dựa vào đó, nhà phân tích – đầu tư có cơ sở đánh giá chất lượng và khả năng phát triển của công ty. Và thông qua đó, có thể dự đoán được những xu hướng chuyển biến giá cổ phiếu mà công ty niêm yết.

Các nhân tố cơ bản cần phân tích bao gồm:

- Thông tin cơ bản về công ty
- Báo cáo tài chính của công ty
- Hoạt động kinh doanh của công ty
- Loại ngành mà công ty đang hoạt động
- Các điều kiện kinh tế vĩ mô ảnh hưởng chung đến giá cả cổ phiếu.

Ngoài ra, việc phân tích cơ bản còn được thực hiện dựa trên giá của các phiên giao dịch hiện tại và trong quá khứ, sẽ giúp ta đánh giá giá trị của một chứng khoán (dưới / trên giá trị hiện hành). Chính vì vậy, phân tích cơ bản được xem là kim chỉ nam cho những quyết định đầu tư trong tương lai.

Trong đề tài này, nhóm chúng tôi tập trung phân tích ba nhân tố chính: thông tin cơ bản công ty, báo cáo tài chính và hoạt động kinh doanh của công ty.

## 2.2. Các chỉ số tài chính của công ty niêm yết

Tình hình tài chính, hoạt động sản xuất kinh doanh của doanh nghiệp là yếu tố quan trọng nhất quyết định giá trị của cổ phiếu. Mặc dù giá giao dịch trên thị trường đôi khi không phản ánh đúng giá trị của cổ phiếu, nhưng việc phân tích các chỉ số tài chính cơ bản cũng giúp nhà đầu tư lựa chọn được những cổ phiếu tốt và an toàn.

Các chỉ số tài chính được trình bày dưới đây, sẽ giúp các nhà phân tích cũng như nhà đầu tư có thể nắm được tình hình tài chính của công ty trong quá khứ cho đến hiện tại, từ đó, có quyết định đúng trong tương lai. Bảng 4 (Phụ lục) sẽ tóm tắt công thức tính cho từng nhóm tỷ số với các chỉ số đại diện liên quan.

### 2.2.1. Chỉ số EPS (Earning per Share)

EPS là một trong những chỉ tiêu được sử dụng thường xuyên nhất đối với các nhà đầu tư và có ảnh hưởng lớn đối với giá cổ phiếu công ty.

EPS được sử dụng như một chỉ số thể hiện khả năng kiếm lợi nhuận của doanh nghiệp, được tính bởi công thức:

$$\text{EPS} = \frac{\text{Tổng lợi nhuận sau thuế 4 quý gần nhất}}{\text{Tổng số lượng cổ phiếu lưu hành}}$$

Trong khi tính toán giá trị EPS, ta thường tính bằng phương pháp lũy kế bốn quý (tổng lãi của bốn quý trong năm). Bởi vì, lãi cả năm chưa kiểm toán so với lãi cả năm sau kiểm toán có thể xảy ra mức độ sai số lớn, nên với cách tính này, ta sẽ điều chỉnh giảm sự chênh lệch giữa báo cáo kiểm toán và trước kiểm toán của năm gần nhất.

Khi sử dụng chỉ số này, nhà đầu tư nên xem xét mức độ ảnh hưởng bởi các chính sách hoặc ước tính kế toán hay sự sai lệch trong việc báo cáo lợi nhuận của doanh nghiệp.

Tuy nhiên, nhược điểm của EPS đó chính là không phản ánh vốn tự có trong việc tạo ra mức lợi nhuận đó. Có thể hai công ty có cùng mức EPS, nhưng một công ty dùng ít vốn hơn và nhiều vốn vay hơn. Do đó, nhà đầu tư cần xem xét chỉ số này trong sự tương quan với nhiều chỉ số khác.

### 2.2.2. Chỉ số P/E (Price to Earning)

P/E là một trong những chỉ số thuộc nhóm “tỷ số giá trị thị trường”, nhóm này được dùng để đo lường kỳ vọng của nhà đầu tư dành cho cổ đông.

P/E là chỉ số phản ánh mối quan hệ giữa thị giá (giá thị trường cổ phiếu hiện hành) cổ phiếu với lợi nhuận sau thuế trong một năm của công ty.



$$P/E = \frac{\text{Giá trị thị trường}}{\text{Lợi nhuận ròng trên một cổ phần (EPS)}}$$

Chỉ số này được dùng khá phổ biến như một công cụ để nhà đầu tư xem xét mình sẽ sẵn sàng đầu tư bao nhiêu để có được một đồng lợi nhuận của công ty.

P/E càng cao nghĩa là chúng khoán mua càng cao và ngược lại. Mặt khác, nhà đầu tư có thể mua với mức P/E giá cao để hy vọng trong tương lai lợi nhuận công ty cao thì P/E lúc đó sẽ thấp.

### 2.2.3. Book value

Book value được dùng để xác định tổng số vốn cổ đông phổ thông của một công ty trên bảng cân đối kế toán. Đồng thời, chỉ số này sẽ cho ta biết giá trị tài sản thực còn lại của công ty khi công ty rút khỏi kinh doanh.

Đây là thước đo chính xác giá trị của công ty, là yếu tố tương đối ổn định, nên nó là số liệu thích hợp trợ giúp phân tích cho các nhà đầu tư (mua cổ phiếu đúng giá trị để hưởng cổ tức hàng năm) hơn là cho các nhà đầu cơ (mua cổ phiếu rồi bán lại với giá đỉnh điểm để hưởng mức chênh lệch).

Book value được tính bởi công thức sau đây:

$$\text{Book value} = \frac{\text{Vốn chủ sở hữu} - \text{Tài sản cố định vô hình}}{\text{KLCP lưu hành hiện tại}}$$

**Mã số:** Vốn chủ sở hữu: 400

Tài sản cố định vô hình: 227

### 2.2.4. P/B (Price to Book)

Tỷ số P/B được sử dụng để so sánh giá của một cổ phiếu so với giá trị book value của cổ phiếu đó. Tỷ lệ này được tính toán bằng cách lấy giá đóng cửa hiện tại của cổ phiếu chia cho giá trị book value tại quý gần nhất của cổ phiếu đó:

$$P/B = \frac{\text{Giá đóng cửa}}{\text{Book value}}$$

Đối với các nhà đầu tư, P/B là công cụ giúp họ có thể tìm kiếm được các cổ phiếu có giá thấp mà phần lớn thị trường bỏ qua. Nếu một doanh nghiệp đang bán cổ phần với mức giá thấp hơn giá trị book value của nó ( $P/B < 1$ ), khi đó có hai trường hợp sẽ xảy ra: *hoặc là* giá trị tài sản của công ty đã bị thổi phồng quá mức *hoặc là* thu nhập trên tài sản của công ty quá thấp.

Tỷ số P/B chỉ thực sự có ích khi nhà đầu tư xem xét các doanh nghiệp có mức độ tập trung vốn cao hoặc các công ty tài chính, bởi giá trị tài sản của các công ty này tương đối lớn. Vì công tác kế toán phải tuân thủ những tiêu chuẩn nghiêm ngặt, nên giá trị book value của tài sản hoàn toàn không tính tới các tài sản vô hình như thương hiệu, uy tín, các tài sản trí tuệ khác... do công ty tạo ra.

### 2.2.5. Beta

Tham số phản ánh mối quan hệ giữa độ biến động của giá chứng khoán với sự biến động của mức giá chung trên thị trường. Nó phản ánh độ nhạy cảm của chứng khoán đang xem xét với mức giá chung của thị trường. Beta được sử dụng trong mô hình định giá tài sản vốn (CAPM) để tính toán tỷ suất sinh lời kỳ vọng của một tài sản dựa vào hệ số beta của nó và tỷ suất sinh lời trên thị trường.

$\beta$  cho ta biết khuynh hướng và mức độ biến động của một chứng khoán nào đó đối với sự biến động của thị trường.

- Nếu chứng khoán có  $\beta = 1$ , có nghĩa là giá chứng khoán đó sẽ di chuyển cùng bước đi với thị trường.
- Nếu chứng khoán có  $\beta < 1$ , nghĩa là chứng khoán đó sẽ có mức thay đổi ít hơn mức thay đổi của thị trường.
- Nếu  $\beta > 1$  thì giá chứng khoán sẽ thay đổi nhiều hơn mức dao động của thị trường.
- Ví dụ, nếu một chứng khoán có  $\beta = 1,2$  thì trên lý thuyết mức biến động của chứng khoán này sẽ cao hơn mức biến động chung của thị trường là 20%.
- Qua quá trình quan sát, nhiều cổ phiếu thuộc các ngành cung cấp dịch vụ thường có  $\beta < 1$ . Ngược lại, hầu hết các cổ phiếu dựa trên kỹ thuật công nghệ cao có  $\beta > 1$ , thể hiện khả năng tạo được một tỷ suất sinh lợi cao hơn, nhưng cũng đồng thời tiềm ẩn rủi ro cao hơn.

Giá trị  $\beta$  được xác định bởi công thức sau:

$$\beta = \frac{\text{covar}(R_i, R_m)}{\text{var}(R_m)}$$

Trong đó:

$R_i$  : Tỷ suất sinh lời của chứng khoán.

$R_m$ : Tỷ suất sinh lời của thị trường (VN-Index).

$\text{Var}(R_m)$ : Phương sai của tỷ suất sinh lời của thị trường.

Covar(R<sub>i</sub>,R<sub>m</sub>): Hiệp phương sai của tỷ suất sinh lời của chứng khoán và tỷ suất sinh lời của thị trường.

Tỷ suất sinh lời được bởi:

$$R = \frac{p_1 - p_0}{p_0}$$

Với: p<sub>1</sub> là giá đóng cửa điều chỉnh phiên đang xét.

p<sub>0</sub> là giá đóng cửa điều chỉnh phiên trước đó.

### 2.2.6. Vốn hóa thị trường – Market Capitalization

Là thước đo quy mô của một doanh nghiệp, là tổng giá trị thị trường của doanh nghiệp, được xác định bằng số tiền bỏ ra để mua lại toàn bộ doanh nghiệp này trong điều kiện hiện tại.

Khi tính giá trị vốn hoá thị trường, ta chỉ sử dụng cổ phiếu phổ thông chứ không tính đến cổ phiếu ưu đãi, vì chỉ cổ phiếu phổ thông mới đem lại cho người sở hữu quyền tham gia điều hành doanh nghiệp.

$$\text{Market capitalization} = \text{Giá đóng cửa} \times \text{Tổng số lượng cổ phiếu lưu hành}$$

Quy mô và tốc độ tăng của giá trị vốn hoá thị trường là thước đo vô cùng quan trọng để đánh giá thành công hay thất bại của một doanh nghiệp. Tuy nhiên, có nhiều yếu tố khác (ví dụ như việc mua lại chính cổ phiếu của mình trên thị trường, hoặc là mua lại một doanh nghiệp khác) có thể ảnh hưởng đến sự tăng giảm của giá trị này.

Giá trị vốn hoá thị trường phản ánh giá cổ phiếu của một công ty, trong khi giá này có thể thay đổi theo kì vọng của các nhà đầu tư, vì vậy chỉ số này không phản ánh hoàn toàn chính xác giá trị thực sự của doanh nghiệp đó.

### 2.2.7. ROE (Return on Equity)

Là tỷ số lợi nhuận ròng trên vốn chủ sở hữu, một tỷ số quan trọng nhất đối với các cổ đông, tỷ số này đo lường khả năng sinh lợi trên mỗi đồng vốn của cổ đông thường, được tính bởi công thức:

$$\text{ROE} = \frac{\text{Lợi nhuận sau thuế thu nhập doanh nghiệp}}{\text{Vốn chủ sở hữu}} \times 100$$

**Mã số:** Lợi nhuận cổ đông công ty mẹ: 60.

Vốn chủ sở hữu: 400.

Chỉ số này là thước đo chính xác để đánh giá một đồng vốn bỏ ra và tích lũy tạo ra bao nhiêu đồng lời. Hệ số này thường được các nhà đầu tư phân tích để so sánh với các cổ phiếu cùng ngành trên thị trường, từ đó tham khảo khi quyết định mua cổ phiếu của công ty nào.

Tỷ lệ ROE tỷ lệ thuận với khả năng sử dụng hiệu quả đồng vốn của cổ đông, nghĩa là công ty đã cân đối một cách hài hòa giữa vốn cổ đông với vốn đi vay để khai thác lợi thế cạnh tranh của mình trong quá trình huy động vốn, mở rộng quy mô. Cho nên hệ số ROE càng cao thì các cổ phiếu càng hấp dẫn các nhà đầu tư hơn

### 2.2.8. ROA (Return on Assets)

Là tỷ số lợi nhuận ròng trên tài sản, được dùng để đo lường khả năng sinh lợi trên mỗi đồng tài sản của công ty và được tính bởi công thức sau:

$$ROA = \frac{\text{Lợi nhuận ròng}}{\text{Tổng tài sản}} \times 100$$

**Mã số:** Lợi nhuận ròng: 60

Tổng tài sản: 100 + 200

Chỉ số ROA thể hiện tính hiệu quả của quá trình tổ chức, quản lý hoạt động sản xuất kinh doanh của doanh nghiệp. Tài sản của một công ty được hình thành từ vốn vay và vốn chủ sở hữu. Cả hai nguồn vốn này được sử dụng để tài trợ cho các hoạt động của công ty. Giá trị của ROA sẽ thể hiện: bình quân cứ một đồng tài sản được sử dụng trong quá trình sản xuất kinh doanh thì tạo ra được bao nhiêu đồng lợi nhuận. ROA càng cao càng tốt vì khi đó, công ty đang kiếm được nhiều tiền hơn trên lượng đầu tư ít hơn.

### 2.2.9. Đòn bẩy tài chính – Financial Leverage

Đòn bẩy tài chính là chỉ số liên quan đến việc doanh nghiệp sử dụng nguồn tài trợ từ các khoản vay thay cho vốn cổ phần. Các phép đo đòn bẩy tài chính là công cụ để xác định xác suất doanh nghiệp mất khả năng thanh toán các hợp đồng nợ. Doanh nghiệp càng nợ nhiều thì càng có nguy cơ cao mất khả năng hoàn thành nghĩa vụ trả nợ.

$$\text{Financial Leverage} = 1 + \frac{\text{Tổng nợ}}{\text{Vốn chủ sở hữu}}$$

**Mã số:** Tổng nợ: 300 (Nợ ngắn hạn + Nợ dài hạn)

Vốn chủ sở hữu: 400

Ý nghĩa: Đối với các doanh nghiệp hoạt động kinh doanh hiệu quả, hệ số Đòn bẩy tài chính càng cao (Nợ > Vốn chủ sở hữu) thì càng làm tăng tỉ lệ sinh lời của vốn chủ sở hữu. Ngược lại, khi doanh nghiệp không hoạt động hiệu quả, mọi rủi ro sẽ dồn vào những người chủ sở hữu vốn.

#### **2.2.10. Tỷ số thanh khoản nhanh – Quick Ratio**

Quick Ratio là một trong hai chỉ số thuộc nhóm “tỷ số thanh khoản”; là tỷ lệ giữa tài sản dễ quy đổi thành tiền mặt trên nợ ngắn hạn, nó phản ánh năng lực thanh toán nhanh các khoản nợ hiện hành bằng số tài sản khả hoán (Quick Assets), mà không buộc phải bán đi hàng tồn kho để bù vào.

Chỉ số này được xác định dựa vào thông tin từ bảng cân đối tài sản nhưng không kể giá trị hàng tồn kho trong giá trị tài sản lưu động khi tính toán:

$$\text{Quick ratio} = \frac{\text{Tài sản ngắn hạn} - \text{Hàng tồn kho}}{\text{Tổng nợ ngắn hạn}}$$

**Mã số:** Tài sản ngắn hạn: 100

Hàng tồn kho: 140

Tổng nợ ngắn hạn: 310

Ý nghĩa: Nếu hệ số này quá nhỏ thì công ty sẽ bị giảm uy tín với bạn hàng, gặp khó khăn trong việc thanh toán công nợ. Nếu hệ số này quá lớn sẽ phản ánh lượng tiền tồn quỹ nhiều làm giảm hiệu quả sử dụng vốn.

Hệ số Quick ratio được xem là bình thường khi dao động trong khoảng từ 0.5 → 1.

#### **2.2.11. Hệ số biên lợi nhuận ròng – Net Profit Margin**

Là chỉ số phản ánh khoản lợi nhuận ròng (lợi nhuận sau thuế) của một doanh nghiệp so với doanh thu, được xác định bởi công thức:

$$\text{Net Profit Margin} = \frac{\text{Lợi nhuận ròng}}{\text{Doanh thu}}$$

**Mã số:** Lợi nhuận ròng : 60

Doanh thu: 1

Mức ổn định của Hệ số biên lợi nhuận ròng giữa các ngành khác nhau là khác nhau và phụ thuộc vào chu kỳ kinh tế. Đối với các doanh nghiệp có chính sách quản lý các nguồn vốn hiệu quả sẽ đạt được mức lợi nhuận ròng tương đối cao hơn so với các doanh nghiệp khác.

Những nguyên nhân khiến cho Hệ số biên lợi nhuận ròng bị giảm thường là sự tăng trưởng doanh thu quá thấp so với mức tăng chi phí, bên cạnh đó là một số nguyên nhân liên quan đến thuế, như là doanh nghiệp bắt đầu phải đóng thuế sau một số năm được miễn giảm thì Hệ số biên lợi nhuận ròng có thể bị giảm mạnh.

Doanh nghiệp với Hệ số biên lợi nhuận ròng cao hơn và (có thể) tăng liên tục so với hệ số biên lợi nhuận ròng trung bình của ngành được xem là nhóm doanh nghiệp có tình trạng kinh doanh thuận lợi.

### 2.3. Kết luận

Mục đích của phân tích cơ bản một cổ phiếu là cung cấp thông tin để đánh giá triển vọng tăng trưởng và lợi nhuận. Tuy nhiên, hiện nay, nhiều công ty không trung thực trong việc tổng kết báo cáo tài chính, do đó, khiến cho việc phân tích có độ chính xác thấp. Thay vì chỉ nhìn nhận doanh nghiệp ở một khía cạnh, chúng ta nên kết hợp nhiều tiêu chí phân tích khác nhau (lợi nhuận, định giá...), xem xét một cách toàn diện, để có những quyết định đầu tư phù hợp và chính xác.

## 3. Mô hình ARIMA

Ý tưởng mô hình hóa các hiện tượng tự nhiên thành các phương trình toán học phục vụ cho mục đích phân tích và dự báo đã được phổ biến từ nhiều thập kỷ trước. Trong khoa học, đặc biệt là bộ môn vật lý, đôi khi chúng ta có thể rút trích ra các biểu thức, mô hình dựa trên những quy luật tự nhiên, từ đó cho phép tính toán chính xác một biến số tại bất kỳ thời điểm nào bằng các thuộc tính đã biết, chẳng hạn như chúng ta có thể tính toán được quỹ đạo của tên lửa thông qua hướng đi và vận tốc đã biết, tính toán vận tốc rơi của một viên gạch biết được khối lượng và lực cản không khí... Nếu như kết quả của những phép tính đó là chính xác, chúng ta gọi những mô hình này là mô hình **xác định** (*deterministic*).

Tuy nhiên, trong ứng dụng thực tế, hầu như không có hiện tượng nào là hoàn toàn **xác định**, có rất nhiều yếu tố ảnh hưởng tới việc tính toán mà chúng ta không biết trước được chẳng hạn như sự thay đổi về hướng gió, sức gió có thể thay đổi quỹ đạo bay của tên lửa, hoặc một vật cản bất ngờ xuất hiện có thể làm cho viên gạch giảm tốc hay thậm chí là dừng hẳn... Trong nhiều vấn đề, chúng ta phải xem xét đến yếu tố thời gian của hiện tượng, chẳng hạn như doanh thu hàng tháng của một doanh nghiệp, nó phụ thuộc vào nhiều yếu tố không biết trước được và không thể đưa ra một mô hình cho kết quả tính toán chính xác doanh thu trong tương lai. Mặc dù vậy, chúng ta có thể đưa ra mô hình dùng để tính toán xác suất của giá trị tương lai trong một khoảng giới hạn xác định. Nó được sử dụng với đầu vào chuỗi thời gian và đầu ra là các giá trị dự báo trong tương lai, mô hình này được gọi là mô hình **ngẫu nhiên** (*stochastic*).

Mô hình ngẫu nhiên, được cho là mô hình sinh ra các giá trị tương ứng theo thời gian của một hiện tượng nào đó thường được biết đến với tên gọi **chuỗi thời gian** (*time series*). Hiện nay, có rất nhiều phương pháp lập mô hình chuỗi thời gian, trong đó, phương pháp tổng quát và rất phổ biến trong thực tế là **tự hồi quy kết hợp trung bình trượt** (*autoregressive integrated moving average - ARIMA*), hay còn gọi là phương pháp luận Box-Jenkins. Mô hình này đưa ra dự báo giá trị tương lai của một chuỗi thời gian dựa trên mối liên hệ và giá trị của dữ liệu hiện tại và quá khứ.

### 3.1. Cơ sở lý luận

Chuỗi thời gian là một chuỗi dữ liệu quan sát được ghi nhận theo thứ tự thời gian ví dụ như số phản ứng hóa học xảy ra trong một giờ, số liệu hàng hóa sản xuất từng ngày của một nhà máy trong một tháng, doanh thu của một doanh nghiệp trong 24 tháng... Nhìn chung, chuỗi thời gian được sử dụng như một cách thể hiện trong nhiều lĩnh vực như kinh tế, kỹ thuật, khoa học... Về cơ bản, giá trị nội tại của chuỗi thời gian thể hiện qua nhận định rằng giá trị các quan sát sẽ **phụ thuộc** vào những giá trị trước đó tùy theo nhiều mức độ khác nhau. Tính chất này được xem như là giá trị ứng dụng thực tế đáng giá nhất của chuỗi thời gian. Phân tích chuỗi thời gian (*Time Series Analysis*) là kỹ thuật phân tích sự phụ thuộc giữa các quan sát nhằm diễn suy quy luật vận động của chuỗi. Từ đó, đưa ra các mô hình ngẫu nhiên (*stochastic*) để ứng dụng vào việc đưa ra các giá trị dự báo cho tương lai.

Trong thực tế, các chuỗi thời gian có thể có tính **dừng** (*stationary*) hoặc không. Một chuỗi thời gian được xem như là **dừng** nếu như trung bình và phương sai của quá trình không thay đổi theo thời gian và giá trị hiệp phương sai giữa hai mốc thời gian chỉ phụ thuộc vào khoảng cách hay độ trễ (*lag*) của nó chứ không phụ thuộc vào thời điểm thực tế mà hiệp phương sai được tính. Hay nói một cách đơn giản, chuỗi **dừng** khi các điểm thuộc chuỗi đạt mức cân bằng xung quanh một **hằng số trung bình** (*constant mean*) với phương sai không đổi theo thời gian. Để có thể đưa ra dự báo tin cậy, thì chuỗi thời gian đó cần phải có tính **dừng** làm cơ sở để phân tích. Tuy vậy, phần lớn các chuỗi thời gian đều không có tính **dừng**, để giải quyết vấn đề này, người ta thường lấy sai phân bậc 1 hoặc bậc 2 của chuỗi thời gian để biến nó thành chuỗi có tính **dừng**. Ý nghĩa của việc lấy sai phân là cho dù các quan sát kề nhau có thể không **dừng** nhưng sự thay đổi của chúng (chênh lệch giữa hai quan sát) thông thường không mang cách biệt quá lớn và vận động theo một quy luật ổn định.

Để đạt được tính **dừng**, người ta tạo ra một chuỗi mới bằng cách lấy chênh lệch của hai quan sát kề nhau và phân tích dựa trên chuỗi đó thay vì chuỗi ban đầu.

“Mục tiêu của Box-Jenkins là xác định và ước lượng một mô hình dựa trên dữ liệu quá khứ có sẵn; hay nói một cách khác, mô hình được xác định dựa trên dữ liệu mà nó đã sinh ra

trước đó. Nếu sau đó mô hình ước lượng này được sử dụng để dự báo, ta phải đặt giả thiết rằng các đặc điểm của mô hình này không đổi theo thời gian và đặc biệt là trong các khoảng thời gian tương lai. Vậy, lý do đơn giản của việc yêu cầu dữ liệu có tính dừng là bản thân mọi mô hình suy luận từ các dữ liệu này có thể cho là ổn định, từ đó cung cấp cơ sở có giá trị cho việc dự báo.” [2]

Mô hình ARIMA có 3 biến  $p$ ,  $d$ ,  $q$  với  $p$  là bậc tự hồi quy,  $d$  là bậc sai phân và  $q$  là bậc trung bình trượt, thường được ký hiệu dưới dạng ARIMA( $p$ ,  $d$ ,  $q$ ).

### 3.2. Các dạng mô hình

#### 3.2.1. Mô hình tự hồi quy (AR)

Gọi  $z$  là chuỗi và  $t$  là đại lượng thời gian tương ứng với giá trị  $z_t$  ta lập mô hình như sau:

$$(z_t - \mu) = \Phi_1(z_{t-1} - \mu) + a_t$$

Với  $\mu$  là giá trị trung bình của chuỗi  $z$  và  $a_t$  là sai số ngẫu nhiên tại thời điểm  $t$  (sai số sẽ được tính bằng cách lấy giá trị  $z_t$  dự báo trừ cho giá trị tương ứng trong thực tế). Nếu chuỗi sai số sẽ có giá trị trung bình bằng 0 và phương sai không đổi  $\sigma^2$ , ta có thể kết luận rằng đó là chuỗi *sai số thuần túy ngẫu nhiên* (*white noise*). Nếu như  $\Phi_1 = 1$ , phương trình có dạng

$$(z_t - \mu) = 1 * (z_{t-1} - \mu) + a_t$$

Đây là lý thuyết *bước ngẫu nhiên* (*Random Walk*), cho rằng giá trị trong tương lai bằng với giá trị trước đó cộng với một sai số ngẫu nhiên bất kỳ có trung bình bằng 0 và phương sai không đổi theo thời gian.

Ta có thể viết

$$(z_t - \mu) = \Phi_1(z_{t-1} - \mu) + \Phi_2(z_{t-2} - \mu) + \dots + \Phi_p(z_{t-p} - \mu) + a_t \quad (1)$$

Hay dạng ngắn gọn là:

$$\Phi(B)(z_t - \mu) = a_t$$

Với  $B$  là toán tử lùi:  $(B)z_t = z_{t-1}$

Phương trình (1) là mô hình *tự hồi quy* bậc  $p$ : AR( $p$ ), hay ký hiệu một cách tổng quát hơn là ARIMA( $p$ , 0, 0)

Lý do gọi mô hình này là *tự hồi quy* vì như đã thấy, phương trình (1) suy ra biến phụ thuộc dựa trên các biến độc lập nên đây là một hàm hồi quy. Các biến độc lập đó chính là giá trị  $z_{t-1}$ ,  $z_{t-2}$ , ...,  $z_{t-q}$ ; mà đây là những giá trị trước đó của  $z_t$  nên có thể nói  $z_t$  *tự hồi quy* trên những giá trị của chính nó.



### 3.2.2. Mô hình trung bình trượt (MA)

Ngoài mô hình tự hồi quy, giá trị  $z_t$  còn có thể được dự báo bằng mô hình trung bình trượt như sau:

$$(z_t - \mu) = a_t + \theta_1 a_{t-1}$$

Với  $\mu$  là hằng số trung bình và giá trị  $a$  đại diện cho sai số ngẫu nhiên. Ở đây, giá trị  $z$  tại thời điểm  $t$  sẽ bằng trung bình của chuỗi  $z$  cộng với trung bình trượt của sai số hiện tại và quá khứ. Mô hình này cho rằng một giá trị của chuỗi  $z$  trong tương lai lúc nào cũng tồn tại sai số và chúng ta có thể dùng mối liên hệ giữa sai số trong tương lai với sai số hiện tại, quá khứ để đưa ra giá trị dự báo. Đây là mô hình trung bình trượt bậc nhất MA(1).

Viết một cách tổng quát hơn:

$$(z_t - \mu) = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2)$$

Hay dạng ngắn gọn là:  $(z_t - \mu) = \theta(B)a_t$

Với  $B$  là toán tử lùi.

Phương trình (2) là phương trình trung bình trượt bậc  $q$ , hay còn ký hiệu là MA(0, 0,  $q$ ).

### 3.2.3. Mô hình tự hồi quy trung bình trượt (ARMA)

Trong thực tế, có rất nhiều chuỗi thời gian chứa tính chất của cả hai mô hình vừa giới thiệu trên. Để đạt được sự linh hoạt khi mô hình hóa các chuỗi thời gian, chúng ta có thể kết hợp tự hồi quy và trung bình trượt vào cùng một mô hình:

$$(z_t - \mu) = \Phi_1(z_{t-1} - \mu) + \dots + \Phi_p(z_{t-p} - \mu) + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (3)$$

Hay dạng ngắn gọn là:

$$\Phi(B)(z_t - \mu) = \theta(B)a_t$$

Phương trình (3) là phương trình tự hồi quy trung bình trượt bậc  $p, q$  – ARMA( $p, q$ ), hay còn ký hiệu là ARIMA( $p, 0, q$ ).

### 3.2.4. Mô hình tự hồi quy kết hợp trung bình trượt (ARIMA)

Lưu ý rằng ba mô hình vừa giới thiệu đều có sự xuất hiện của hằng số trung bình của chuỗi  $z, \mu$ . Tức là chúng chỉ có thể áp dụng đối với các chuỗi có tính dừng. Còn đối với chuỗi không dừng, chúng ta phải lấy sai phân bậc  $d$  để biến chúng thành chuỗi dừng. Trong thực tế, thông thường chỉ lấy tới sai phân bậc 1 hoặc bậc 2 chúng ta sẽ có một chuỗi dừng:

- Chuỗi gốc:  $z_t$
- Chuỗi sai phân bậc 1:  $w_t = z_t - z_{t-1}$

- Chuỗi sai phân bậc 2:  $v_t = w_t - w_{t-1}$

Giả sử sau khi lấy sai phân bậc 1 ta được chuỗi dừng, chúng ta có thể ứng dụng mô hình ARMA(p, q) như sau:

$$w_t = \Phi_1 w_{t-1} + \dots + \Phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Với  $w_t = z_t - z_{t-1}$ . Đây là mô hình tự hồi quy kết hợp trung bình trượt bậc p, 1, q – ARIMA(p, 1, q). Lưu ý rằng nếu như chúng ta thay  $w_t$  ở phương trình (4) bằng  $y_t - \mu$ , khi  $d = 0$ , thì phương trình sẽ trở về dạng ARIMA(p, 0, q) hay ARMA(p, q) (3) như một dạng đặc biệt.

### 3.3. Phương pháp lập mô hình ARIMA

Để dự báo giá trị tương lai của một chuỗi thời gian theo phương pháp luận Box-Jenkins (ARIMA), chúng ta thực hiện bốn bước sau: **nhận dạng, ước lượng, kiểm định chẩn đoán, dự báo**. Với 3 bước đầu là bước xây dựng mô hình và kiểm tra tính chính xác của mô hình. Sau khi đã có mô hình ARIMA với các bậc p, d, q và các tham số tự hồi quy và trung bình trượt ước lượng đáng tin cậy, chúng ta sẽ thực hiện bước thứ tư, cũng là bước quan trọng nhất, **dự báo**.

#### 3.3.1. Nhận dạng

Trong bước này, chúng ta sẽ xác định các bậc p, d, q để xác định mô hình phù hợp. Hai công cụ chính hỗ trợ trong bước nhận dạng là **hàm tự tương quan** (Autocorrelation Function – ACF), **hàm tự tương quan riêng phần** (Partial Autocorrelation Function – PACF) và các **biểu đồ tương quan** (Correlogram) tương ứng.

Ngoài ra chúng ta cũng có thể sử dụng phương pháp kiểm định nghiệm đơn vị để xem xét một chuỗi thời gian có phải là dừng hay không. Nếu như chuỗi không dừng thì chúng ta sẽ lấy sai phân bậc d cho đến khi nó có tính dừng. Như đã nói ở trên, trong thực tế, chỉ lấy sai phân bậc 1 hoặc bậc 2 thì ta sẽ được chuỗi dừng.

Sau khi tìm được chuỗi thời gian có tính dừng, chúng ta tiếp tục tìm bậc p và q cho mô hình dựa trên các biểu đồ tương quan.

#### 3.3.2. Ước lượng

Sau khi đã nhận dạng mô hình ARIMA, tiếp theo chúng ta cần xác định các tham số tự hồi quy  $\alpha$  và các tham số trung bình trượt  $\beta$ . Các tham số có thể được suy ra từ phương pháp tính **sai số bình phương tối thiểu** (Minimum Least Square Error) đơn giản, nhưng đôi khi cần phải áp dụng các phương pháp ước lượng **tham số phi tuyến** (Non-linear Least Squares) và ước lượng **hợp lý cực đại** (Maximum Likelihood) khá phức tạp.

### 3.3.3. Kiểm định

Với mô hình ARIMA tìm được và các tham số của nó chúng ta bắt đầu thực hiện việc kiểm định tính chính xác của mô hình. Lý do của việc kiểm định là có thể có nhiều hơn một mô hình ARIMA phù hợp với chuỗi thời gian, chưa kể đến việc ước lượng các tham số tự hồi quy, tham số trung bình trượt không chính xác. Một phép kiểm định thường được dùng là tìm sai số ước lượng của mô hình khi dự báo, tính toán hàm tự tương quan (ACF) để xem chuỗi này có phải là một chuỗi ngẫu nhiên thuần túy hay không; nếu phải, chúng ta chấp nhận mô hình đang kiểm định; nếu không, ta phải lặp lại từ đầu. Có thể nói, “*phương pháp luận Box-Jenkins là một quy trình lặp*”. [2]

### 3.3.4. Dự báo

Khi đã hoàn tất các bước ở trên, chúng ta có được một mô hình ARIMA với khả năng dự báo hiệu quả. Dựa vào mô hình này, ta có thể đưa ra các dự báo đáng tin cậy với những sai số được cho là ngẫu nhiên. Sai số được tính bằng cách lấy giá trị thực tế tại thời điểm  $t$  trừ cho giá trị dự báo, các sai số này có thể sử dụng để ước lượng sai số cho giá trị tiếp theo trong tương lai từ đó đưa ra những kết quả dự báo chính xác hơn. Ngoài ra, chúng ta cũng có thể tính các khoảng tin cậy của giá trị dự báo như khoảng 50%, khoảng 95%. Trong một số trường hợp lấy sai phân (để có chuỗi dừng), thì chúng ta cần suy ngược lại để có được giá trị  $z_t$  tương ứng.

## 3.4. Xác định mô hình (Model Identification)

Xác định mô hình là phương pháp dùng để đưa ra các mô hình phù hợp để tiếp tục xem xét ở các bước tiếp theo. Mục tiêu chính ở đây là tìm ra các bậc  $p$ ,  $d$ ,  $q$  cần thiết cho mô hình ARIMA và ước lượng sơ bộ cho các tham số hồi quy và tham số trung bình trượt.

Nhìn chung, quy trình này bao gồm 2 bước như sau:

- Lấy sai phân của chuỗi thời gian cho đến khi nó trở thành một chuỗi dừng. Hay nói cách khác, đi tìm bậc sai phân  $d$ , ký hiệu là  $I(d)$ .
- Xác định bậc  $p$ ,  $q$  cho quy trình ARMA, ký hiệu ARMA( $p$ ,  $q$ ).

Hai công cụ chính được sử dụng trong quy trình này là **hàm tự tương quan mẫu SACF** (*sample autocorrelation function*) và **hàm tự tương quan mẫu riêng phần SPACF** (*sample partial autocorrelation function*).

Hàm tự tương quan mẫu:

$$r_k = \frac{c_k}{c_0}$$

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \mu)(z_{t+k} - \mu) \quad k = 0, 1, 2, \dots, K$$

$$c_0 = \frac{1}{N} \sum_{t=1}^N (z_t - \mu)^2$$

Hàm tự tương quan riêng phần (theo thuật toán đệ quy Durbin-Levinson):

$$\hat{\Phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} r(k-j)}{1 - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} r(j)}$$

$$\hat{\Phi}_{kj} = \hat{\Phi}_{k-1,j} - \hat{\Phi}_{kk} \hat{\Phi}_{k-1,k-j} \quad j = 1, 2, \dots, k-1$$

### 3.4.1. Sử dụng SACF và SPACF

#### a. Xác định bậc sai phân (d)

Giả sử có một hằng số  $\delta$  dương và gần bằng 0. Suy ra,  $G = 1 - \delta$  sẽ gần bằng 1. Khi đó, với độ trễ k:

$$r_k = r_{k-1} * (1 - k\delta)$$

Hàm tự tương quan sẽ không giảm nhanh về giá trị zero. Do đó, hàm tự tương quan giảm chậm về giá trị zero được xem như là một dấu hiệu cho thấy tham số tự hồi quy có thể gần bằng 1, hay nói cách khác chuỗi đang xét gần như là một chuỗi ngẫu nhiên và không có tính dừng. Chúng ta có thể lấy sai phân bậc d để biến chuỗi đang xét thành một chuỗi có tính dừng:

$$W_t = \Delta^d z_t = z_t - z_{t-1}$$

Khi đó, hàm tự tương quan của  $w_t$  sẽ giảm dần về 0 tương đối nhanh.

Trong thực tế, d thường có giá trị là 0, 1 hoặc 2. Và k có thể bằng một phần ba hoặc một phần tư độ lớn của mẫu.

#### b. Xác định bậc p, q cho quy trình ARMA

Sau khi đã xác định được bậc d, chúng ta tiếp tục ước lượng SACF và SPACF của chuỗi  $w$  sau khi lấy sai phân bậc dưới  $w_t = (1 - B)^d z_t$  và vẽ các biểu đồ tương quan, để có thêm thông tin cho việc xác định bậc p, q. Bằng việc xem xét các biểu đồ tương quan so với lý thuyết biểu đồ tự tương quan và lý thuyết biểu đồ tự tương quan riêng phần của các mô hình AR(p), MA(q), ARMA(p, q), chúng ta có thể xác định sơ bộ các mô hình phù hợp.

	<b>Tự hồi quy (AR)</b>	<b>Trung bình trượt (MA)</b>	<b>Tổng hợp (ARMA)</b>
<b>Tự tương quan</b>	Vô hạn (dạng số mũ và/hoặc sóng hình sin)	Hữu hạn	Vô hạn (dạng số mũ và/hoặc sóng hình sin sau độ trễ q – p đầu tiên)
	Giảm dần	Tới hạn tại độ trễ q	Giảm dần
<b>Tự tương quan riêng phần</b>	Hữu hạn	Vô hạn (chủ yếu theo dạng số mũ và/hoặc sóng hình sin)	Vô hạn (dạng số mũ và/hoặc sóng hình sin sau độ trễ p – q đầu tiên)
	Tới hạn tại độ trễ p	Giảm dần	Giảm dần

*Bảng 1. Đặc điểm của biểu đồ tương quan tương ứng với các loại mô hình*

Nhìn chung, khi hàm tự tương quan của một quy trình tự hồi quy bậc p giảm dần, hàm tự tương quan riêng phần của nó sẽ tới hạn tại độ trễ p. Ngược lại, hàm tự tương quan của một quy trình trung bình trượt bậc q tới hạn tại độ trễ q, trong khi hàm tự tương quan riêng phần của nó giảm dần. Nếu cả hàm tự tương quan và tự tương quan riêng phần đồng thời giảm dần, thì có thể nó là quy trình hồi quy trung bình trượt ARMA. Hơn nữa, hàm tự tương quan của một quy trình tổng hợp sẽ có biểu đồ hỗn hợp giữa dạng số mũ và dạng sóng hình sin sau độ trễ q – p đầu tiên. Ngược lại, hàm tự tương quan riêng phần sẽ có biểu đồ hỗn hợp giữa dạng số mũ và sóng hình sin sau độ trễ p – q đầu tiên.

Ước lượng sơ bộ và khoảng cho tham số của các mô hình được liệt kê như sau:

	<b>Order</b>				
	(1, d, 0)	(0, d, 1)	(2, d, 0)	(0, d, 2)	(1, d, 1)
<b>Ước lượng sơ bộ</b>	$\Phi_1 = r_1$	$r_1 = \frac{-\theta_1}{1 + \theta_1^2}$	$\Phi_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}$ $\Phi_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$	$r_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ $r_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$	$r_1 = \frac{(1 - \theta_1\Phi_1)(\Phi_1 - \theta_1)}{1 + \theta_1^2 - 2\Phi_1\theta_1}$ $r_2 = r_1\Phi_1$
<b>Miền giá trị</b>	$-1 < \Phi_1 < 1$	$-1 < \theta_1 < 1$	$-1 < \Phi_2 < 1$ $\Phi_2 + \Phi_1 < 1$ $\Phi_2 - \Phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$	$-1 < \Phi_1 < 1$ $-1 < \theta_1 < 1$

*Bảng 2. Công thức tính giá trị tham số sơ bộ và miền giá trị*

Trong thực tế, chúng ta dùng hàm tự tương quan mẫu và tự tương quan riêng phần mẫu đối với một số lượng quan sát hữu hạn nên chúng thường không giống với lý thuyết cơ bản. Do đó, cần kiểm định và tính toán cùng lúc nhiều mô hình để tìm ra mô hình phù hợp nhất cho việc dự báo.

### 3.4.2. Độ lệch chuẩn của SACF và SPACF

Như đã nói qua, hàm mẫu chỉ được xây dựng dựa trên 1 tập hữu hạn các quan sát nên nó sẽ có khác biệt so với hàm lý thuyết của một chuỗi vô hạn. Vì vậy, cần phải tính toán một vài chỉ số để nhận biết độ lệch của nó so với giá trị lý thuyết. Nói cụ thể hơn, chúng ta cần một giá trị làm mốc để đánh giá các giá trị tự tương quan và tự tương quan riêng phần có ý nghĩa 0 về mặt thống kê tại độ trễ q hoặc p hay không. Đối với mô hình trung bình trượt có bậc q, chúng ta có thể tính độ lệch chuẩn của hàm tự tương quan mẫu với công thức Bartlett như sau:

$$\sigma[r_k] \approx \frac{1}{\sqrt{2n}} [1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)]^{1/2} \quad k > q \quad [3]$$

Đối với mô hình tự hồi quy bậc p, độ lệch chuẩn của hàm tự tương quan riêng phần mẫu:

$$\sigma[\Phi_{kk}] \approx \frac{1}{\sqrt{2n}} \quad k > p \quad [1]$$

Khi lập mô hình người ta thường xem xét tính dừng của mô hình theo độ lệch chuẩn của hàm tự tương quan với khoảng tin cậy là 95%. Theo đó, với bất kỳ chuỗi thời gian nào có độ lớn n thì  $r_k$  nằm trong khoảng:

$$-1.96\left(\frac{1}{\sqrt{2n}}\right) < r_k < +1.96\left(\frac{1}{\sqrt{2n}}\right)$$

Nếu thỏa điều kiện trên, ta có thể suy ra rằng  $r_k$  tuân theo phân phối chuẩn và có giá trị thống kê bằng 0. Ngược lại, nếu  $r_k$  nằm ngoài khoảng trên, suy ra  $r_k$  có giá trị thống kê khác 0.

Chúng ta cũng có thể dùng khoảng tin cậy 95% để xác định bậc p, q cho mô hình dựa trên các biểu đồ tương quan.

### 3.4.3. Một vài lưu ý quan trọng

Lấy sai phân đơn giản không thể sử dụng đối với các chuỗi thời gian có thành phần mùa.

Trong các chuỗi tồn tại xu hướng, chúng ta cần thêm vào một hằng số  $\theta_0$ . Để xác định giá trị này có thực sự tồn tại hay không, có thể so sánh trung bình của chuỗi w (sai phân bậc d) với phương sai của nó.

#### 3.4.4. Một số công cụ xác định mô hình khác

Mặc dù hàm tự tương quan và hàm tự tương quan riêng phần mẫu là hai công cụ hữu ích trong quá trình xác định mô hình. Tuy nhiên, đối với các chuỗi thời gian có dạng mô hình tổng hợp (ARMA), chúng có vẻ không mấy hiệu quả khi chỉ đưa ra được các kết quả mơ hồ. Đây không phải là một vấn đề nghiêm trọng bởi vì mô hình được xác định trong bước đầu chỉ mang tính thử nghiệm và cần được kiểm định, chẩn đoán, điều chỉnh cho phù hợp nếu cần thiết. Mặc dù vậy, hiện nay người ta bắt đầu quan tâm đến những công cụ hỗ trợ để việc lập mô hình được nhanh chóng và hiệu quả hơn. Trong đó có **tiêu chuẩn chọn lựa mô hình** (*model selection criteria*) AIC được đưa ra bởi Akaike năm 1974 đang dần trở nên phổ biến vì tính hiệu quả và đơn giản của nó.

AIC có ý nghĩa như là một chỉ số đánh giá mức độ thông tin sai lệch khi dùng một mô hình toán học để suy diễn thực tế. Nó còn được xem như miêu tả của sự đánh đổi giữa **tính thiên lệch** (*bias*) và **độ biến thiên** (*variance*) trong quá trình xây dựng mô hình, hay tổng quát hơn là giữa độ chính xác và độ phức tạp của mô hình.

Theo hướng tiếp cận này, sau khi đã có những mô hình ARIMA được ước lượng bằng hàm hợp lý hóa cực đại (sử dụng trong giai đoạn ước lượng tham số), các chỉ số AIC sẽ được tính theo công thức sau:

$$AIC = 2k - 2\ln(L) \quad [5]$$

Với  $k$  là số tham số trong mô hình và  $L$  là **giá trị cực đại của hàm hợp lý hóa** (*maximized value of the likelihood function*).

AIC của mỗi mô hình sẽ được tính toán và chúng ta lấy mô hình cho giá trị AIC nhỏ nhất. Đồng nghĩa với việc mô hình đó cho thông tin sai lệch so với thực tế ít nhất.

Trong thực tế, người ta thường dùng AICc thay cho AIC để giảm thiểu tham số trong mô hình:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad [5]$$

Với  $k$  là số tham số trong mô hình và  $n$  là độ lớn của mẫu.

Một chỉ số khác cũng được sử dụng rộng rãi là **Bayesian information criteria** (*BIC*):

$$BIC = k \ln(n) + n \ln(\hat{\sigma}_a^2)$$

Chỉ số này khá tương tự như AIC vì chỉ số càng thấp chứng tỏ mô hình càng tốt và có cùng tham số  $-2\ln(L)$ . Điểm khác chính là ở tham số đầu tiên, ngoài số tham số của mô hình, nó còn phụ thuộc vào độ lớn của mẫu.

Theo như Henry [7], thì AIC thể hiện tốt khi dùng trong các mô hình có cỡ mẫu nhỏ và vừa, nhưng với cỡ mẫu lớn thì nó không nhất quán và có khuynh hướng chọn mô hình phức tạp (nhiều tham số). Ngược lại, chỉ số BIC không hiệu quả khi dùng với mô hình có cỡ mẫu nhỏ, nhưng lại nhất quán và thể hiện tốt khi mô hình có cỡ mẫu lớn, thêm vào đó, nó ưu tiên cho những mô hình dạng đơn giản (ít tham số).

### 3.4.5. Ước lượng tham số sơ bộ

#### a. Đối với mô hình trung bình trượt (MA)

Giá trị tự tương quan của độ trễ  $q$  trong mô hình khác 0 và có thể viết dựa trên các tham số của mô hình như sau:

$$r_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad k=1,2,\dots, q \quad [3]$$

Giá trị sơ bộ của các tham số  $\theta$  có thể được tính bằng cách thế giá trị  $r_k$  vào và giải phương trình phi tuyến trên. Sau đó, chúng ta có thể tính ước lượng sơ bộ cho  $\sigma_a^2$  (phương sai của sai số  $a$ ) với phương trình:

$$c_0 = \sigma_a^2(1 + \theta_1^2 + \dots + \theta_q^2) \quad [3]$$

Với  $c_0$  là phương sai mẫu.

#### b. Đối với mô hình tự hồi quy (AR)

Các tham số được tính bằng phương trình Yule-Walker như sau:

$$\Phi = R_p^{-1}r_p \quad [3]$$

Với  $R_p$  là ước lượng của ma trận  $p \times p$  của tương quan tới  $p - 1$  và  $r_p'$  là vector  $(r_1, r_2, \dots, r_p)$ . Ví dụ, nếu  $p = 3$  phương trình Yule-Walker có dạng:

$$\begin{bmatrix} \hat{\Phi}_{31} \\ \hat{\Phi}_{32} \\ \hat{\Phi}_{33} \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \quad [3]$$

Với  $\hat{\Phi}_{pj}$  là ký hiệu của tham số tự hồi quy thứ  $j$  trong quy trình bậc  $p$ .

#### Đối với mô hình tự hồi quy trung bình trượt (ARMA)

Trong ứng dụng thực tế, thông thường sau khi lấy sai phân để có được chuỗi tĩnh tại, mô hình ARMA là mô hình biểu diễn chuỗi  $w$  tốt nhất với  $w_t = \Delta^d z_t$ .

Ví dụ, đối với mô hình  $(1, d, 1)$  ta có thể dùng công thức ở bảng 6.1 để ước lượng tham số  $\Phi_1$  và  $\theta_1$ .



Có một phương pháp tổng quát để xác định tham số cho mô hình ARMA. Đầu tiên chúng ta sẽ tìm được các tham số  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p$  thông qua giải hệ phương trình tuyến tính:

$$c_{q+1} = \hat{\Phi}_1 c_q + \hat{\Phi}_2 c_{q-1} + \dots + \hat{\Phi}_p c_{q-p+1}$$

$$c_{q+2} = \hat{\Phi}_1 c_{q+1} + \hat{\Phi}_2 c_{q-1} + \dots + \hat{\Phi}_p c_{q-p+2}$$

...

$$c_{q+p} = \hat{\Phi}_1 c_{q+p-1} + \hat{\Phi}_2 c_{q+p-2} + \dots + \hat{\Phi}_p c_q$$

Sau đó, biểu diễn  $c'_j$  của  $w'_t = \theta(B)a_t$  qua hiệp phương sai mẫu  $c_j$  của  $w_t$ :

$$c'_j = \sum_{i=0}^p \Phi_i^2 c_j + \sum_{i=1}^p (\Phi_0 \Phi_i + \Phi_1 \Phi_{i+1} + \dots + \Phi_{p-i} \Phi_p) (c_{j+i} + c_{j-i})$$

Với  $j = 1, 2, \dots, q$  và  $\Phi_0 = -1$

Sau đó giải các phương trình sau để tìm các tham số  $\theta_j$ :

$$\sigma_a^2 = \frac{c'_0}{1 + \theta_1^2 + \dots + \theta_q^2}$$

$$\theta_j = - \left( \frac{c'_j}{\sigma_a^2} - \theta_1 \theta_{j+1} - \theta_2 \theta_{j+2} - \dots - \theta_{q-j} \theta_q \right)$$

### c. Ước lượng sơ bộ cho phương sai của sai số

Đối với quy trình tự hồi quy AR, ta có thể tính bằng công thức:

$$c_0 = \sigma_a^2 (1 - \hat{\Phi}_1 r_1 - \hat{\Phi}_2 r_2 - \dots - \hat{\Phi}_p r_p)$$

Với quy trình trung bình trượt MA:

$$\sigma_a^2 = \frac{c_0}{1 + \theta_1^2 + \dots + \theta_q^2}$$

Với quy trình tự hồi quy trung bình trượt ARMA:

$$\sigma_a^2 = \frac{1 - \hat{\Phi}_1^2}{1 + \theta_1^2 - 2\hat{\Phi}_1 \theta_1}$$

### 3.5. Ước lượng mô hình (Model Estimation)

Quy trình xác định mô hình cho chúng ta một số thông tin sơ bộ về các mô hình tiềm năng. Bước tiếp theo, cần tính toán những tham số phù hợp.

#### 3.5.1. Hàm hợp lý hóa và hàm tổng bình phương

##### a. Hàm hợp lý hóa

Giả sử chúng ta có một tập  $N$  quan sát  $z$ , có hàm phân phối xác suất  $p(z|\zeta)$  với  $\zeta$  là một hoặc nhiều biến bất kỳ. Có thể sử dụng vector  $\zeta$  để ký hiệu tập các tham số, trong trường hợp này, bằng  $(p + q + 1)$  tham số  $(\Phi, \theta, \sigma)$  của mô hình ARIMA.

Khi đã có một tập quan sát  $z$ , chúng ta sẽ xem xét các giá trị khác nhau cho  $\zeta$  sao cho nó là giá trị tương ứng sinh ra tập  $z$ . Công cụ được sử dụng để phục vụ cho mục đích này là **hàm hợp lý hóa** (likelihood function)  $L(\zeta|z)$ , với  $z$  là một tập quan sát cố định và  $\zeta$  là biến cần tìm. Do đó, hàm hợp lý hóa thường được xem như là **hằng số nhân tùy biến** (arbitrary multiplicative constant).

Trong thực tế, **hàm log hợp lý hóa** (log-likelihood function) thường được dùng với dạng,

$$\ln[L(\zeta|z)] = l(\zeta|z) \quad [1]$$

Và được xem như là **hằng số cộng tùy biến** (arbitrary additive constant)

Phương pháp tìm tham số  $\zeta$  cho giá trị cực đại của hàm hợp lý hóa đồng thời là giá trị cực đại của hàm log hợp lý hóa, được gọi là **ước lượng hợp lý hóa cực đại** (maximum likelihood estimates).

Khi ước lượng tham số cho mô hình ARIMA, chúng ta tính toán chủ yếu theo phương pháp hợp lý hóa vô điều kiện. Tuy nhiên, đối với chuỗi có độ lớn  $n$  vừa phải hoặc lớn, tham số có thể ước lượng theo phương pháp hợp lý hóa có điều kiện.

##### i. Hợp lý hóa có điều kiện

Giả sử rằng chúng ta có một chuỗi  $z_{-d+1}, \dots, z_0, z_1, z_2, \dots, z_N$  có độ lớn là  $N$  sinh ra từ mô hình ARIMA( $p, d, q$ ). Từ quan sát gốc  $z_t$ , có thể sinh ra chuỗi  $w_1, w_2, \dots, w_n$  với  $w_t = \Delta^d z_t$ , độ lớn  $n = N - d$ . Do đó, chúng ta chỉ cần tìm tham số  $\Phi$  và  $\theta$  cho mô hình ARMA của chuỗi dùng  $w_t$ .

$$a_t = w_t - \Phi_1 w_{t-1} - \Phi_2 w_{t-2} - \dots - \Phi_p w_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad [1]$$

Vì vậy, với mọi tham số  $(\Phi, \theta)$ , chúng ta có thể tính được chuỗi  $a_t$  với một tập  $w_t$ . Khi đó, hàm log hợp lý hóa có điều kiện trên sự lựa chọn  $p$  giá trị khởi tạo  $w_t$  (ký hiệu là  $w^*$ ) và  $q$  giá trị khởi tạo  $a_t$  (ký hiệu là  $a^*$ ) được biểu diễn như sau:

$$l^*(\Phi, \theta, \sigma_a) = -n \ln(\sigma_a) - \frac{S_*(\Phi, \theta)}{2\sigma_a^2} \quad [1]$$

Và:

$$S_*(\Phi, \theta) = \sum_{t=1}^n a_t^2(\Phi, \theta | w_*, a_*. w) \quad [1]$$

Trong biểu thức trên, dấu \* được dùng để ký hiệu rằng nó có điều kiện với sự lựa chọn giá trị khởi tạo. Và có thể thấy rằng hàm log hợp lý hóa  $l^*$  chỉ liên hệ với dữ liệu thông qua **hàm tổng bình phương** (*sum of squares function*). Lúc này, nó gần giống như phương pháp **ước lượng bình phương tối thiểu** (*least squares estimates*) với  $\sigma_a$  cố định. Hay nói cách khác,  $l^*$  là phương trình tuyến tính của  $S_*$  với các tham số được chọn sao cho giá trị  $S_*(\Phi, \theta)$  là nhỏ nhất.

## ii. Hợp lý hóa vô điều kiện

Hàm hợp lý hóa vô điều kiện được biểu diễn như sau:

$$l(\Phi, \theta, \sigma_a) = f(\Phi, \theta) - n \ln(\sigma_a) - \frac{S_*(\Phi, \theta)}{2\sigma_a^2} \quad [1]$$

Với  $f(\Phi, \theta)$  là một hàm của  $\Phi$  và  $\theta$ . Hàm **tổng bình phương vô điều kiện** (*unconditional sum of squares*) được tính theo công thức:

$$S(\Phi, \theta) = \sum_{t=1}^n [a_t | w, \Phi, \theta]^2 + [e_*]' \Omega^{-1} [e_*] \quad [1]$$

Với  $[a_t | w, \Phi, \theta] = E[a_t | w, \Phi, \theta]$  ký hiệu cho kỳ vọng của  $a_t$  theo  $w, \Phi, \theta$ .  $e_* = (w_{t-p}, \dots, w_0, a_{t-q}, \dots, a_0)$  ký hiệu vector  $q + p$  giá trị khởi tạo cho  $w_t$  và  $a_t$  trước thời gian  $t = 1$ .

Thông thường,  $f(\Phi, \theta)$  chỉ có ý nghĩa trong chuỗi có  $n$  nhỏ. Đối với  $n$  vừa và lớn,  $l(\Phi, \theta, \sigma_a)$  chủ yếu phụ thuộc vào  $S(\Phi, \theta) / 2\sigma_a^2$ . Do đó, tham số được ước lượng bằng cách tối thiểu hóa  $S(\Phi, \theta)$  (phương pháp bình phương tối thiểu) cho giá trị rất gần với phương pháp ước lượng hợp lý hóa cực đại.

### b. Phương pháp tính tổng bình phương vô điều kiện với mô hình trung bình trượt

Mô hình của  $w_t$  sau khi đã xác định có thể viết theo hai dạng:

$$[a_t] = [w_t] + \theta[a_{t-1}]$$

$$[e_t] = [w_t] + \theta[e_{t+1}]$$

Với  $a_t$  tượng trưng cho sai số hiện tại và quá khứ, còn  $e_t$  tượng trưng cho sai số hiện tại và tương lai. Phương trình ở trên là **dự báo xuôi** (*forward forecasting*), bên dưới là **dự báo ngược** (*backward forecasting*).

- Từ chuỗi gốc  $z_t$ , đưa ra chuỗi  $w_t = \Delta^d z_t$ .

- $[e_0], [e_{-1}], \dots$  có giá trị bằng 0. Chúng ta mặc định sai số trong tương lai  $e_{t+1} = 0$ .
- Giá trị  $[a_0], [a_{-1}], \dots, [a_{-q+1}]$  khác 0 và được tính bằng phương pháp dự báo ngược.

**c. Phương pháp tính tổng bình phương vô điều kiện tổng quát**

Với mô hình ARIMA(p, d, q) tổng quát, ta có thể tính  $a_t$  qua công thức:

$$a_t = w_t - \Phi_1 w_{t-1} - \Phi_2 w_{t-2} - \dots - \Phi_p w_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad [1]$$

Đầu tiên chúng ta thế giá trị khởi tạo bằng 0 để tính toán  $a_t$ . Sau đó, tính chuỗi  $u_t$  ngược lại (*backward*) theo công thức:

$$u_t = a_t^0 + \theta_1 u_{t+1} + \dots + \theta_q u_{t+q} \quad [1]$$

Với  $a_t^0$  là ký hiệu của chuỗi  $a_t$  được tính bằng giá trị khởi tạo  $a_{1-q} = 0$  và  $u_{n+1} = 0$ . Với chuỗi  $u_t$  ta có thể tính được giá trị  $[a_0]$  và  $[w_0]$  theo công thức:

$$a_0 = \frac{-u_0(1 - \theta^2)}{1 - 2^{2(n+1)}}$$

Và dùng chúng làm giá trị ban đầu để tính lại giá trị cho  $[a_t]$  và  $[w_t]$ .

Sau đó chúng ta có thể tính được  $S(\Phi, \theta)$  với biểu thức sau:

$$S(\Phi, \theta) = \sum_{t=1}^n [a_t]^2 + \frac{([w_0] - [a_0])^2}{K} \quad [1]$$

Với  $K = \sigma_a^{-2} \gamma_0 - 1 = (\Phi - \theta)^2 / (1 - \Phi^2)$

**3.5.2. Ước lượng phi tuyến**

Như đã phân tích ở trên, chúng ta có thể thấy ở hầu hết các trường hợp, ước lượng hợp lý hóa cực đại tương tự như ước lượng bình phương tối thiểu. Trong đó:

$$S(\Phi, \theta) = \sum_{t=1}^n [a_t | w, \Phi, \theta]^2 + [e_*]' \Omega^{-1} [e_*] \quad [1]$$

là nhỏ nhất. Giá trị này có thể được ước lượng bằng việc tính toán sai số của mô hình dự báo thông qua **hướng tiếp cận dự báo ngược** (*backward model approach*).

Về tổng quát, sự đơn giản hóa đáng kể trong việc tối thiểu hóa dựa trên các tham số  $\Phi, \theta$ , của hàm tổng bình phương  $\sum_{t=1}^n [f_t(\Phi, \theta)]^2$ , nếu mỗi  $f_t(\Phi, \theta)$  là một phương trình tuyến tính của các tham số  $\Phi, \theta$ . Trong mô hình tự hồi quy thuần túy  $[a_t] = \Phi(B)[w_t]$ , ngoại trừ giá trị khởi tạo,  $[a_t]$  tuyến tính đối với các tham số  $\Phi$ .

$$\frac{d[a_t]}{d\Phi_i} = -[w_{t-i}] + \Phi(B) \frac{d[w_t]}{d\Phi_i}$$

Ngược lại, trong mô hình trung bình trượt thuần túy  $[a_t] = \theta^{-1}(B)[w_t]$ , giá trị  $[a_t]$  luôn luôn phi tuyến đối với các tham số.

$$\frac{d[a_t]}{d\theta_j} = \theta^{-2}(B)[w_{t-j}] + \theta^{-1}(B) \frac{d[w_t]}{d\theta_j}$$

### Tuyến tính hóa mô hình

Để tiện cho việc biểu diễn công thức, ở đây chúng ta sẽ dùng ký hiệu  $\beta$  thay cho  $k = p + q$  tham số  $(\Phi, \theta)$ . Việc cần quan tâm ở đây là tối thiểu

$$S(\Phi, \theta) \approx \sum_{t=1-Q}^n [a_t | \mathbf{w}, \beta]^2 = \sum_{t=1-Q}^n [a_t]^2$$

Khai triển  $[a_t]$  trong một chuỗi Taylor các giá trị tương ứng với các bộ tham số dự đoán  $\beta_0 = (\beta_{1.0}, \beta_{2.0}, \dots, \beta_{k.0})$

$$[a_t] = [a_{t.0}] - \sum_{i=1}^k (\beta_i - \beta_{i.0}) x_{t,i} \quad (2.1.1)$$

Với:

$$[a_{t.0}] = [a_t | \mathbf{w}, \beta_0]$$

Và:

$$x_{t,i} = - \left. \frac{d[a_t]}{d\beta_i} \right|_{\beta = \beta_0}$$

Nếu  $X$  là  $(n + Q) * k$  ma trận  $\{x_{t,i}\}$ , biểu thức (2.1.1) có thể biểu diễn như sau:

$$[a_0] = X(\beta - \beta_0) + [a]$$

Giá trị điều chỉnh  $\beta - \beta_0$  cho giá trị  $S(\beta) = S(\Phi, \theta) = [a] [a]$  tối thiểu lúc này có thể tính bằng phương pháp bình phương tối thiểu. Bởi vì chuỗi  $[a_t]$  không hoàn toàn tuyến tính với tập tham số  $\beta$ , nên nếu chỉ điều chỉnh một lần thì sẽ không lập tức cho giá trị bình phương tối thiểu. Mà thay vào đó, phải liên tục thay đổi các điều chỉnh cho đến khi giá trị bình phương hội tụ về mức thấp nhất. Sự hội tụ sẽ diễn ra nhanh hơn khi sử dụng giá trị khởi tạo hợp lý, chẳng hạn như những giá trị sơ bộ được tính ở giai đoạn xác định mô hình. Nếu chúng ta dùng những giá trị khởi tạo bất hợp lý, quy trình có thể sẽ không bao giờ hội tụ.

### 3.5.3. Ước lượng số học của đạo hàm

Ở đây, chúng ta sử dụng phương pháp ***bình phương tối thiểu phi tuyến tổng quát*** (*general nonlinear least squares*). Nói cụ thể hơn, chúng ta thay đổi tuần tự bộ tham số bằng cách cộng vào 1 biến số  $\delta_i$  cho nó. Với một mô hình cho trước, giá trị  $[a_t | \mathbf{w}, \beta_{1.0}, \beta_{2.0}, \dots, \beta_{k.0}]$  với  $t = 1 - Q, \dots, n$  có thể được tính và sau đó tiếp tục tìm các giá trị  $[a_t | \mathbf{w}, \beta_{1.0} + \delta_1, \beta_{2.0}, \dots, \beta_{k.0}]$ ,  $[a_t | \mathbf{w}, \beta_{1.0}, \beta_{2.0} + \delta_2, \dots, \beta_{k.0}]$ . Từ đó:

$$x_{t,i} = \frac{[at|w, \beta_{1.0}, \dots, \beta_{i.0}, \dots, \beta_{k.0}] - [at|w, \beta_{1.0}, \dots, \beta_{i.0} + \delta_i, \dots, \beta_{k.0}]}{\delta_i} \quad [1]$$

Ghi chú: đầu tiên, giả sử  $a_0 = 0$ , tính chuỗi  $a_t^0 = w_t + \theta a_{t-1}^0$ , sau đó tính ngược lại chuỗi  $u_t = a_t^0 + \theta u_{t+1}$  với  $u_{n+1} = 0$ . Từ chuỗi  $u_t$ , tính được

Và dùng giá trị này để tính lại  $a_t$ .

### 3.5.4. Thuật toán tổng quát cho mô hình có điều kiện

Đôi khi với các chuỗi có cỡ mẫu lớn, chúng ta có thể đặt giả thiết các giá trị khởi tạo cho  $a$  và  $x$  tương đương 0.

Chúng ta có cách tiếp cận tổng quát với mô hình

$$a_t = \theta^{-1}(B)\Phi(B)w_t$$

Với  $w_t = \Delta^d z_t$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_i B^i - \dots - \theta_q B^q$$

$$\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_j B^j - \dots - \Phi_p B^p$$

Gọi  $\beta = (\Phi, \theta)$  là giá trị thực của tham số và  $\beta_0 = (\Phi_0, \theta_0)$

$$a_{t.0} = \theta_0^{-1}(B)\Phi_0(B)w_t$$

Và

$$-\frac{d[a_t]}{d\Phi_j} \beta_0 = u_{t.j} = u_{t-j} \quad -\frac{d[a_t]}{d\theta_i} \beta_0 = v_{t.i} = v_{t-i}$$

Với

$$u_t = \Phi_{1.0} u_{t-1} + \dots + \Phi_{p.0} u_{t-p} + a_{t.0}$$

$$v_t = \theta_{1.0} v_{t-1} + \dots + \theta_{q.0} v_{t-q} - a_{t.0}$$

Phương trình tuyến tính của  $a_{t.0}$  theo  $(\Phi_j - \Phi_{j.0})$  và  $(\theta_i - \theta_{i.0})$  có dạng

$$a_{t.0} = \sum_{j=1}^p (\Phi_j - \Phi_{j.0}) u_{t-j} + \sum_{i=1}^q (\theta_i - \theta_{i.0}) v_{t-i} + a_t$$

Khi đó hiệu giữa tham số thực và tham số khởi tạo (chênh lệch) sẽ là hệ số hồi quy trên  $u_{t-j}$  và  $v_{t-i}$ . Sau khi tìm ra chênh lệch, chúng ta sẽ điều chỉnh lại tham số khởi tạo và tiếp tục vòng lặp để tính toán cho đến khi nào tham số thực và tham số khởi tạo hội tụ.

### 3.6. Kiểm định mô hình (Model Diagnostic)

Sau khi đã xác định dạng mô hình và ước lượng tham số cần thiết, chúng ta sẽ thực hiện kiểm định tính đúng đắn của mô hình. Một phương pháp phổ biến là *phân tích sai số* (*analysis of residual*). Ý nghĩa của việc kiểm định mô hình là chứng minh chuỗi sai số của mô hình không tương quan với nhau và phát sinh một cách ngẫu nhiên.

#### 3.6.1. Kiểm định tự tương quan (Autocorrelation Check)

Giả sử có mô hình:

$$\Phi(B)w_t = \theta(B)a_t$$

Với  $w_t = \Delta^d z_t$ , với tham số  $(\Phi, \theta)$  được ước lượng bằng hàm hợp lý hóa cực đại. Sai số được tính theo công thức:

$$a_t = \theta^{-1}(B)\Phi(B)w_t$$

Hay:

$$a_t = w_t - \sum_{j=1}^p \Phi_j w_{t-j} + \sum_{j=1}^q \theta_j a_{t-j} \quad t = 1, 2, \dots, n$$

Với các tham số khởi tạo được mặc định là 0. Do đó, nếu như mô hình được lập nên tương đối chính xác,

$$a_t = \ddot{a}_t + O\left(\frac{1}{\sqrt{n}}\right)$$

Với  $\ddot{a}_t$  là sai số thực của chuỗi thời gian và  $a_t$  là sai số tính từ mô hình ước lượng. Độ dài của chuỗi càng lớn thì  $a_t$  càng tiến gần về  $\ddot{a}_t$ . Vì vậy, phân tích  $a_t$  có thể chỉ ra tính chính xác của mô hình được lập nên. Hơn nữa, một số dạng đặc biệt của hàm tự tương quan ước lượng của chuỗi  $a_t$  còn hỗ trợ việc điều chỉnh mô hình cho phù hợp.

Giả sử rằng dạng mô hình được xác định là chính xác và tham số thực sự là  $\Phi$  và  $\theta$ . giá trị tự tương quan ước lượng  $r_k(a)$  của chuỗi sai số  $a_t$  sẽ không tương quan với nhau, có phân phối chuẩn trung bình bằng zero và phương sai tương đương  $n^{-1}$ .

Trong thực tế, tham số chỉ được tính theo kiểu ước lượng và không thể biết chắc giá trị thực của nó. Do đó, chỉ có thể tính được sai số ước lượng  $a_t$  mà không tìm được sai số thực  $\ddot{a}_t$ . Hàm tự tương quan  $r_k(a)$  của chuỗi  $a$  có thể cho thấy sự thiếu chính xác của mô hình được lập nên. Box và Pierce đã chứng minh rằng mặc dù trong tất cả các trường hợp, sự suy giảm về phương sai có thể xảy ra đối với các độ trễ thấp. Những độ trễ này có thể có  $r_k(a)$  tương quan mạnh với nhau nhưng tính tương quan của chúng sẽ giảm dần nhanh chóng ở những độ trễ cao. Vì thế, sử dụng độ lệch chuẩn tương đương  $n^{-1/2}$  cho  $r_k(a)$  đồng

nghĩa với việc đánh giá thấp ý nghĩa thống kê tại độ trễ thấp, nhưng giá trị đó có thể chấp nhận tại độ trễ vừa và cao.

### 3.6.2. Kiểm định Portmanteau Lack-of-Fit

Thay vì xem xét từng giá trị  $r_k(a)$  riêng lẻ, có thể dùng một chuỗi  $r_k(a)$  liên tục để xem xét tính chính xác của mô hình:

$$Q = n \sum_{k=1}^K r_k^2(a)$$

Tương đương với phân phối  $\chi^2(K - p - q)$ , với  $K = N - d$  (lấy sai phân). Mặt khác, nếu như mô hình thiếu độ chính xác, giá trị trung bình của  $Q$  sẽ trở nên rất lớn. Do đó, **phép kiểm định tổng quát (Portmanteau)** có thể xác minh lại giả thiết về tính đúng đắn của mô hình.

Tuy nhiên, Ljung và Box lại cho rằng áp dụng giá trị  $Q$  cho phân phối chi bình phương ( $\chi^2$ ) không đủ chính xác thể hiện giá trị phân phối của trị thống kê  $Q$ , dưới **giả thuyết null (null hypothesis)**, với giá trị của  $Q$  có khuynh hướng nhỏ hơn kỳ vọng dưới phân phối chi bình phương. Và từ nhận định trên, Ljung và Box đã đưa ra **dạng điều chỉnh của trị thống kê  $\tilde{Q}$  (modified Ljung-Box-Pierce statistic)**,

$$\tilde{Q} = n(n + 2) \sum_{k=1}^K \frac{r_k^2(a)}{(n-k)}$$

Chi squared																									p value
Degrees of freedom (df)																									
25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
11.52	10.86	10.20	9.54	8.90	8.26	7.63	7.01	6.41	5.81	5.23	4.66	4.11	3.57	3.05	2.56	2.09	1.65	1.24	0.87	0.55	0.30	0.11	0.02	0.00	.99
16.47	15.66	14.85	14.04	13.24	12.44	11.65	10.86	10.09	9.31	8.55	7.79	7.04	6.30	5.58	4.87	4.17	3.49	2.83	2.20	1.61	1.06	0.58	0.21	0.02	.90
18.94	18.06	17.19	16.31	15.44	14.58	13.72	12.86	12.00	11.15	10.31	9.47	8.63	7.81	6.99	6.18	5.38	4.59	3.82	3.07	2.34	1.65	1.01	0.45	0.06	.80
20.87	19.94	19.02	18.10	17.18	16.27	15.35	14.44	13.53	12.62	11.72	10.82	9.93	9.03	8.15	7.27	6.39	5.53	4.67	3.83	3.00	2.19	1.42	0.71	0.15	.70
22.62	21.65	20.69	19.73	18.77	17.81	16.85	15.89	14.94	13.98	13.03	12.08	11.13	10.18	9.24	8.30	7.36	6.42	5.49	4.57	3.66	2.75	1.87	1.02	0.27	.60
24.34	23.34	22.34	21.34	20.34	19.34	18.34	17.34	16.34	15.34	14.34	13.34	12.34	11.34	10.34	9.34	8.34	7.34	6.35	5.35	4.35	3.36	2.37	1.39	0.45	.50
26.14	25.11	24.07	23.03	21.99	20.95	19.91	18.87	17.82	16.78	15.73	14.69	13.64	12.58	11.53	10.47	9.41	8.35	7.28	6.21	5.13	4.04	2.95	1.83	0.71	.40
28.17	27.10	26.02	24.94	23.86	22.77	21.69	20.60	19.51	18.42	17.32	16.22	15.12	14.01	12.90	11.78	10.66	9.52	8.38	7.23	6.06	4.88	3.66	2.41	1.07	.30
30.68	29.55	28.43	27.30	26.17	25.04	23.90	22.76	21.61	20.47	19.31	18.15	16.98	15.81	14.63	13.44	12.24	11.03	9.80	8.56	7.29	5.99	4.64	3.22	1.64	.20
32.28	31.13	29.98	28.82	27.66	26.50	25.33	24.16	22.98	21.79	20.60	19.41	18.20	16.99	15.77	14.53	13.29	12.03	10.75	9.45	8.12	6.74	5.32	3.79	2.07	.15
34.38	33.20	32.01	30.81	29.62	28.41	27.20	25.99	24.77	23.54	22.31	21.06	19.81	18.55	17.28	15.99	14.68	13.36	12.02	10.64	9.24	7.78	6.25	4.61	2.71	.10
34.90	33.71	32.51	31.31	30.10	28.89	27.67	26.45	25.21	23.98	22.73	21.48	20.21	18.94	17.65	16.35	15.03	13.70	12.34	10.95	9.52	8.04	6.49	4.82	2.87	.09
35.47	34.27	33.06	31.85	30.63	29.41	28.18	26.95	25.71	24.46	23.20	21.93	20.66	19.37	18.07	16.75	15.42	14.07	12.69	11.28	9.84	8.34	6.76	5.05	3.06	.08
36.11	34.89	33.68	32.45	31.22	29.99	28.75	27.50	26.25	24.99	23.72	22.44	21.15	19.85	18.53	17.20	15.85	14.48	13.09	11.66	10.19	8.67	7.06	5.32	3.28	.07
36.82	35.60	34.37	33.13	31.89	30.65	29.40	28.14	26.87	25.59	24.31	23.02	21.71	20.39	19.06	17.71	16.35	14.96	13.54	12.09	10.60	9.04	7.41	5.63	3.54	.06
37.65	36.42	35.17	33.92	32.67	31.41	30.14	28.87	27.59	26.30	25.00	23.68	22.36	21.03	19.68	18.31	16.92	15.51	14.07	12.59	11.07	9.49	7.81	5.99	3.84	.05
38.64	37.39	36.13	34.87	33.60	32.32	31.04	29.75	28.44	27.14	25.82	24.49	23.14	21.79	20.41	19.02	17.61	16.17	14.70	13.20	11.64	10.03	8.31	6.44	4.22	.04
39.88	38.61	37.33	36.05	34.76	33.46	32.16	30.84	29.52	28.19	26.85	25.49	24.12	22.74	21.34	19.92	18.48	17.01	15.51	13.97	12.37	10.71	8.95	7.01	4.71	.03
41.57	40.27	38.97	37.66	36.34	35.02	33.69	32.35	31.00	29.63	28.26	26.87	25.47	24.05	22.62	21.16	19.68	18.17	16.62	15.03	13.39	11.67	9.84	7.82	5.41	.02
44.31	42.98	41.64	40.29	38.93	37.57	36.19	34.81	33.41	32.00	30.58	29.14	27.69	26.22	24.73	23.21	21.67	20.09	18.48	16.81	15.09	13.28	11.34	9.21	6.63	.01
52.62	51.18	49.73	48.27	46.80	45.31	43.82	42.31	40.79	39.25	37.70	36.12	34.53	32.91	31.26	29.59	27.88	26.12	24.32	22.46	20.51	18.47	16.27	13.82	10.83	.001

Note. Problems with df>25 would rarely be worked by hand.

Hình 4. Phân phối Chi bình phương



Từ **độ tự do** (*degrees of freedom*) tương đương  $K - p - q$  và **giá trị p** (*p value*) ta có thể lấy ra được phân phối  $\chi^2$  tương ứng để so sánh với trị thống kê  $\hat{Q}$ . Các giá trị có ý nghĩa về mặt thống kê là từ 0.05 trở xuống. Thông thường, 0.05, 0.01 và 0.001 là 3 giá trị được sử dụng. Nếu như  $\hat{Q}$  nhỏ hơn phân phối  $\chi^2$  tại ba giá trị này, có thể đặt giả thiết rằng mô hình được xác định là đáng tin cậy và ngược lại nếu lớn hơn, chúng ta sẽ kết luận rằng mô hình thiếu chính xác. Ý nghĩa của phép so sánh này là nếu như  $\hat{Q}$  lớn hơn trị chi bình phương 0.05 (5%) đồng nghĩa với việc các sai số sinh ra từ mô hình **có ý nghĩa về mặt thống kê** (*statistically significant*), hoàn toàn không phải là ngẫu nhiên mà là do mô hình chưa phù hợp và ngược lại.

Một dạng thường gặp trong thực tế của việc lập mô hình đó là khi dạng mô hình được xác định, nó có thể cố định nhưng tham số sẽ thay đổi trong từng khoảng thời gian khác nhau.

### 3.7. Dự báo (Forecasting)

Sau khi đã có được mô hình, chúng ta bắt đầu đưa ra kết quả dự báo dựa trên mô hình xác định và các tham số ước lượng từ những bước lập mô hình.

#### 3.7.1. Tính toán giá trị dự báo

Đối với một mô hình tổng quát dạng

$$\Phi(B)\Delta^d z_t = \theta(B)a_t$$

Mô hình này có thể dùng để dự báo cho giá trị  $z_{t+l}$ , với  $l \geq 1$ , tại thời điểm hiện tại  $t$ . Dự báo này được xem như là giá trị dẫn xuất từ **thời gian gốc  $t$**  (*origin  $t$* ) cho đến **khoảng thời gian  $l$**  (*lead time  $l$* ) trong tương lai.

Có ba cách để thể hiện mô hình:

- Trực tiếp theo sai phân:

$$z_{t+1} = \Phi_1 \Delta^d z_{t+1-1} + \dots + \Phi_{p+d} \Delta^d z_{t+1-1-p-d} - \theta_1 a_{t+1-1} - \dots - \theta_q a_{t+1-q} + a_{t+1}$$

- Chuỗi sai số vô hạn

$$z_{t+1} = \sum_{j=0}^{\infty} \psi_j a_{t+l-j}$$

- Chuỗi trọng số vô hạn các quan sát trước đó cộng với một sai số  $a_t$

$$z_{t+1} = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} + a_{t+l}$$

Sai số dự báo với  $l = 1$  (one-step-ahead) là:

$$e_t(l) \approx z_{t+1} - \hat{z}_t(l) = a_{t+1}$$

Với  $\hat{z}_t(l)$  tương ứng với giá trị dự báo tại thời gian gốc  $t$  cho lead  $l = 1$ . Một mô hình có độ tin cậy cao đồng nghĩa với việc sai số dự báo mà nó sinh ra không tương quan với nhau. Nói như vậy vì giả sử rằng sai số dự báo tương quan với nhau, thì sai số  $a_{t+1}$  có thể suy ra được từ sai số  $a_t, a_{t-1}, a_{t-2} \dots$  trước đó. Gọi sai số ước lượng là  $\hat{a}_{t+1}$  ta sẽ có giá trị dự báo  $\hat{z}_t(l) + \hat{a}_{t+1}$  cho  $z_{t+1}$  chính xác hơn là  $\hat{z}_t(l)$

Thông thường, người ta hay dùng phương pháp lấy sai phân trực tiếp để đưa ra dự báo:

Ví dụ: Với mô hình ARMA(1, 1, 0)

$$(1 - 0.8B)(1 - B)z_{t+1} = a_{t+1}$$

Chuyển thành

$$(1 - 1.8B + 0.8B^2)z_{t+1} = a_{t+1}$$

Hay

$$z_{t+1} = 1.8z_{t+1-1} - 0.8z_{t+1-2} + a_{t+1}$$

Dự báo tại thời điểm gốc  $t$  sẽ được tính như sau

$$\hat{z}_t(1) = 1.8z_t - 0.8z_{t-1}$$

$$\hat{z}_t(2) = 1.8\hat{z}_t(1) - 0.8z_t$$

$$\hat{z}_t(l) = 1.8\hat{z}_t(l-1) - 0.8\hat{z}_t(l-2) \quad l = 3, 4, \dots$$

Đối với trường hợp mô hình có tính đến sai số trước đó ( $q > 0$ ), ví dụ như ARMA(0, 2, 2)

$$\Delta^2 z_{t+l} = (1 - 0.9B + 0.5B^2)a_{t+1}$$

Có thể chuyển thành

$$z_{t+l} = 2z_{t+l-1} - z_{t+l-2} + a_{t+1} - 0.9a_{t+l-1} + 0.5a_{t+l-2}$$

$$\hat{z}_t(1) = 2z_t - z_{t-1} - 0.9a_t + 0.5a_{t-1}$$

$$\hat{z}_t(2) = 2\hat{z}_t(1) - z_t + 0.5a_t$$

$$\hat{z}_t(l) = 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \quad l = 3, 4, \dots$$

Lưu ý rằng mô hình chỉ tính tới sai số của hiện tại và quá khứ, sai số tương lai (chưa biết) được xem như bằng 0.

### 3.7.2. Khoảng tin cậy

Trong thực tế, lúc nào cũng có **nhiều** (*white noise*) phát sinh trong chuỗi thời gian. Điều này đã được xác định từ trước nên có thể nói rằng dự báo sẽ không bao giờ cho kết quả chính xác tuyệt đối. Tuy nhiên, chúng ta có thể tính toán khoảng tin cậy của dự báo, nó là một đại lượng tượng trưng khoảng sai số cho phép của dự báo, được dẫn xuất dựa trên tham số của mô hình dự báo và phương sai sai số của mô hình.

Với  $\varphi = \Delta^d \Phi(B)$

Đầu tiên, tính  $\psi$  theo công thức

$$\psi_1 = \varphi_1 - \theta_1$$

$$\psi_2 = \varphi_1 \psi_1 + \varphi_2 - \theta_2$$

$$\psi_j = \varphi_1 \psi_{j-1} + \dots + \varphi_{p+d} \psi_{j-p-d} - \theta_j$$

Tiếp theo, chúng ta có thể tính được khoảng tin cậy theo công thức

$$u_{\varepsilon/2} \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} \sigma^2$$

Với khoảng tin cậy 50% và 95%, giá trị  $u_{\varepsilon/2}$  tương ứng là 0.674 và 1.96.

### 3.8. Kết luận

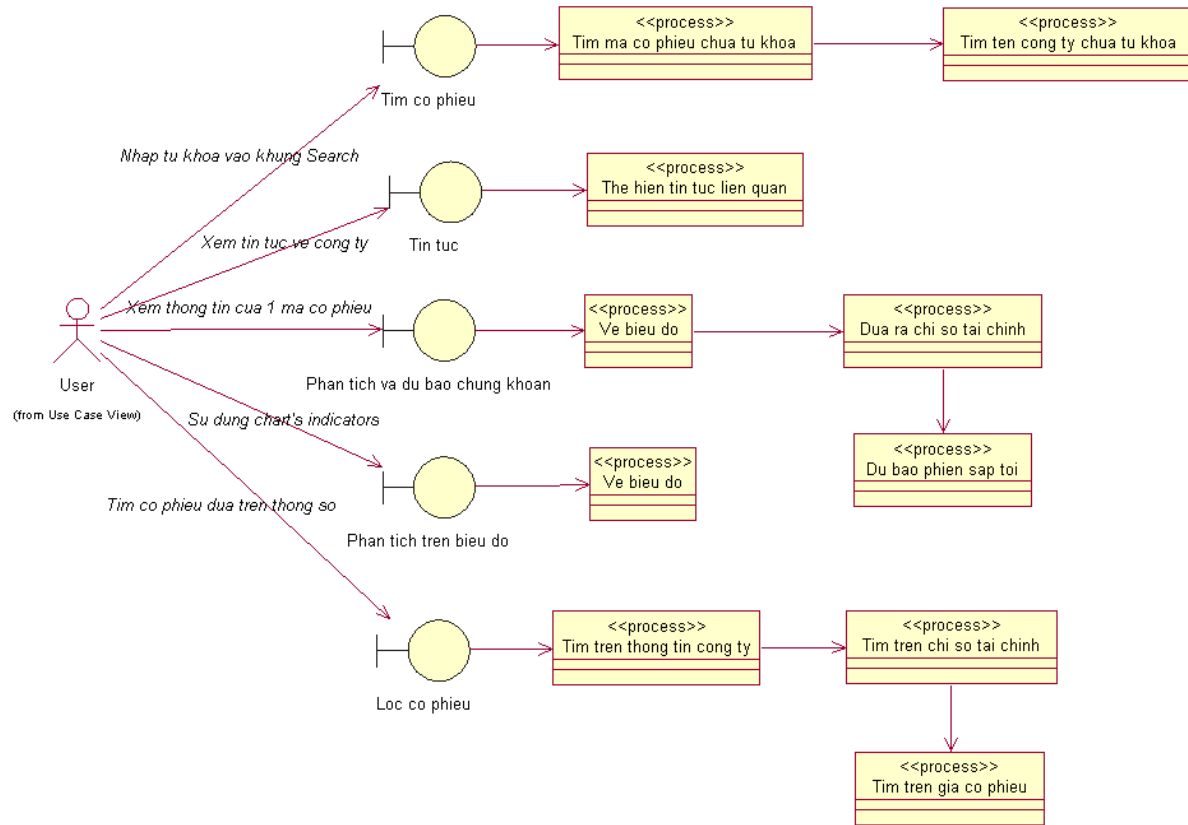
Phương pháp dự báo bằng mô hình ARIMA là một phương pháp có độ tin cậy cao và phổ biến trong các lĩnh vực kinh tế, kỹ thuật, khoa học... Ý nghĩa của phương pháp này là tìm ra quy luật vận động cơ bản của một chuỗi thời gian và đưa ra dự báo cho tương lai. Tùy theo từng chuỗi thời gian mà sai số của dự báo có thể thấp hoặc cao. Trong phạm vi nghiên cứu, nhóm chỉ quan tâm đến mô hình **đơn biến** (*univariate*) là **giá đóng cửa** (*close*) trong khi giá có thể bị ảnh hưởng bởi nhiều yếu tố khác như: khối lượng giao dịch, lợi nhuận trên một cổ phiếu, đòn bẩy tài chính... nên sai số của dự báo chưa được tối ưu.

Một điểm đặc biệt ở mô hình ARIMA đó là nó có thể cho chúng ta biết sai số cho phép của dự báo, thường được biết đến với tên gọi **Khoảng tin cậy**, chính khoảng giá trị sẽ cung cấp giá trị tham khảo rất hữu ích. Nếu như giá trị thực tế vẫn nằm trong khoảng tin cậy, ta có thể cho rằng giá cổ phiếu đang vận động một cách ổn định. Ngược lại, nếu giá trị thực tế nằm ngoài khoảng tin cậy, đó có thể là một tín hiệu cho thấy đang có những biến động bất thường.

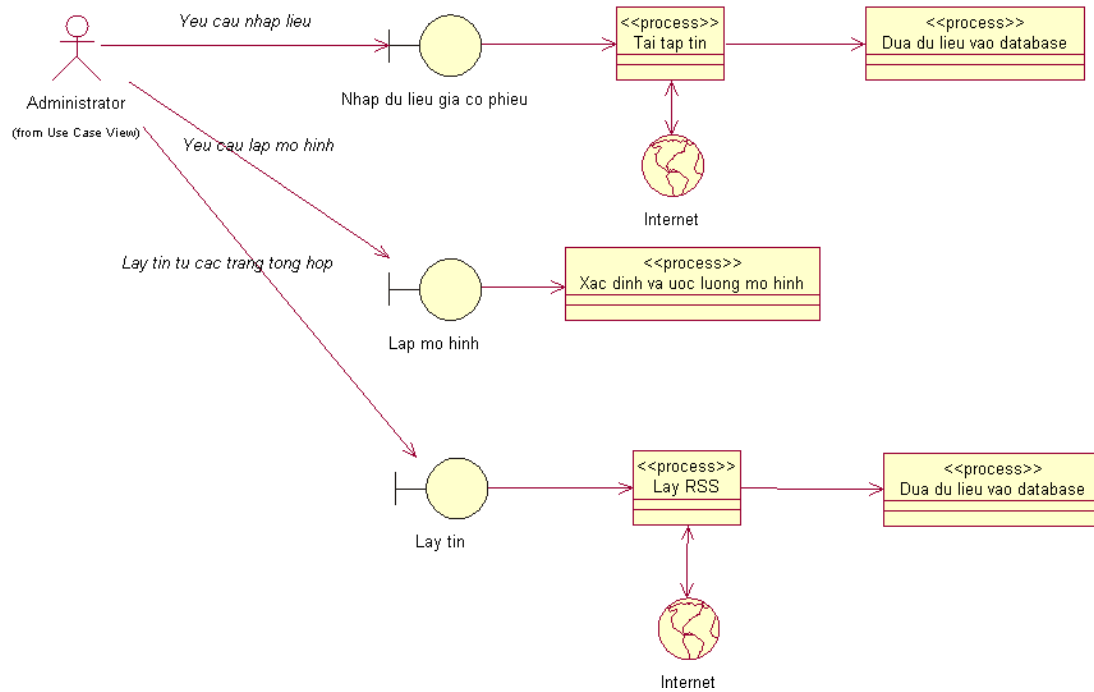
## II. Phân tích - Hiện thực

### 1. Phân tích thiết kế hệ thống

#### 1.1. Data Flow Diagram (DFD)

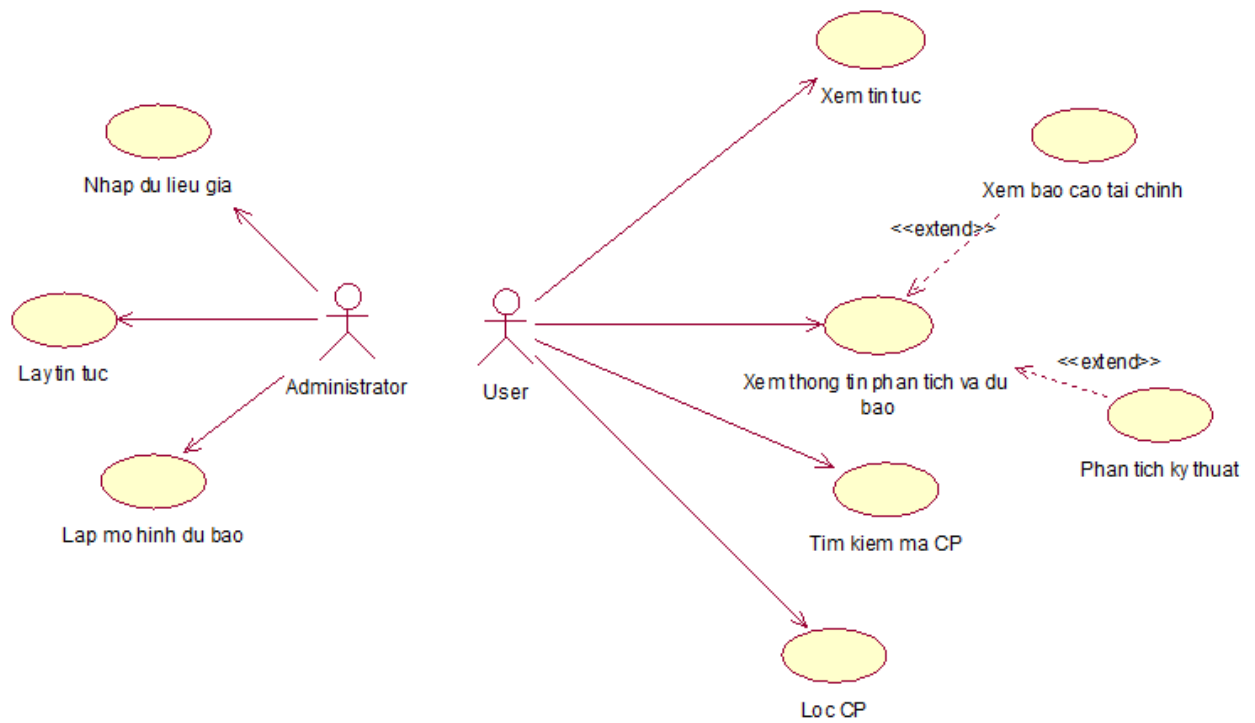


Hình 5. User DFD



Hình 6. Administrator DFD

## 1.2. Đặc tả Use case



Hình 7. Use Case Diagram

### 1.2.1. Xem tin tức

**Mô tả:** đưa người dùng đến trang lưu trữ thông tin liên quan đến thị trường chứng khoán hoặc một mã chứng khoán nhất định.

**Điều kiện bắt đầu:** đang ở trang chủ hoặc trang phân tích dự báo

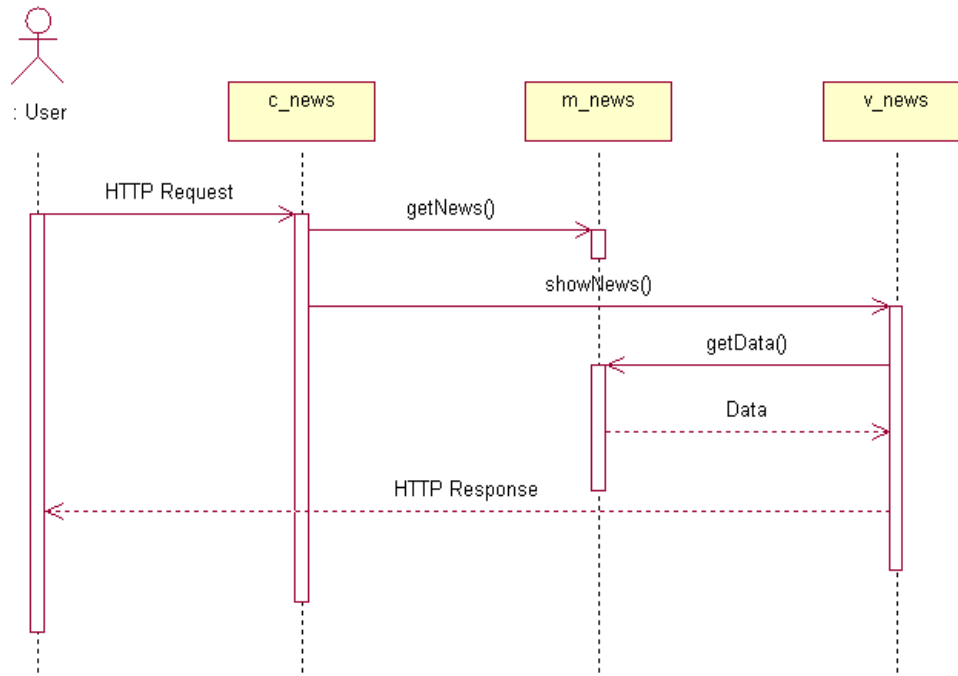
**Luồng sự kiện chính:**

Actor	System
1. Nhấn vào link “Tin tức”	2. Hiện thị danh sách những tin tức liên quan

**Luồng sự kiện phụ:** sau bước 2, nếu như người dùng nhấn vào liên kết để xem thông tin chi tiết, hệ thống sẽ tự chuyển về trang gốc của bản tin.

**Điều kiện kết thúc:** hiển thị trang danh sách tin tức liên quan

**Sequence Diagram:**



Hình 8. SD – Xem tin tức

### 1.2.2. Xem thông tin phân tích và dự báo

**Mô tả:** đưa người dùng đến trang phân tích thông tin và dự báo giá cổ phiếu.

**Điều kiện bắt đầu:** người dùng click vào liên kết mã chứng khoán bất kỳ

**Luồng sự kiện chính:**

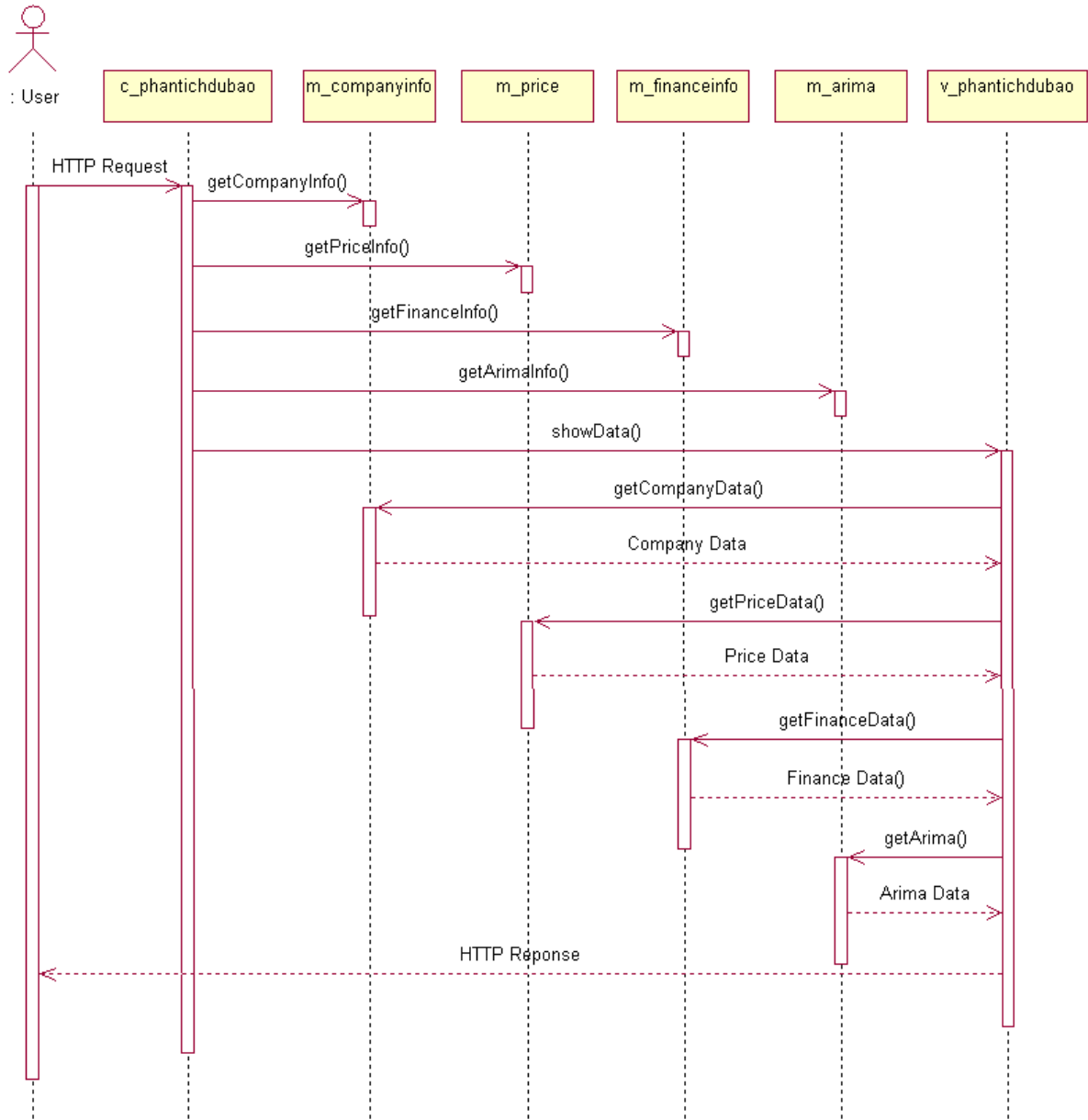
Actor	System
1. Nhấn vào link tên của một mã chứng khoán	2. Hiện thị các thông tin phân tích tài chính, biểu đồ và dự báo giá trong tương lai

**Luồng sự kiện phụ:** sau bước 2, nếu:

- Người dùng sử dụng các công cụ phân tích kỹ thuật trên biểu đồ, chuyển đến Use case 3 (*Phân tích kỹ thuật trên biểu đồ*).
- Người dùng muốn xem báo cáo tài chính chi tiết, chuyển đến Use case 4 (*Xem báo cáo tài chính đầy đủ*)

**Điều kiện kết thúc:** hiển thị các thông tin về tài chính cũng như biểu đồ giá trên thị trường, kết quả dự báo.

**Sequence Diagram:**



Hình 9. SD – Xem thông tin và phân tích dự báo



### 1.2.3. Phân tích kỹ thuật trên biểu đồ

**Mô tả:** giúp người dùng phân tích xu hướng giá cổ phiếu bằng các công cụ trên biểu đồ.

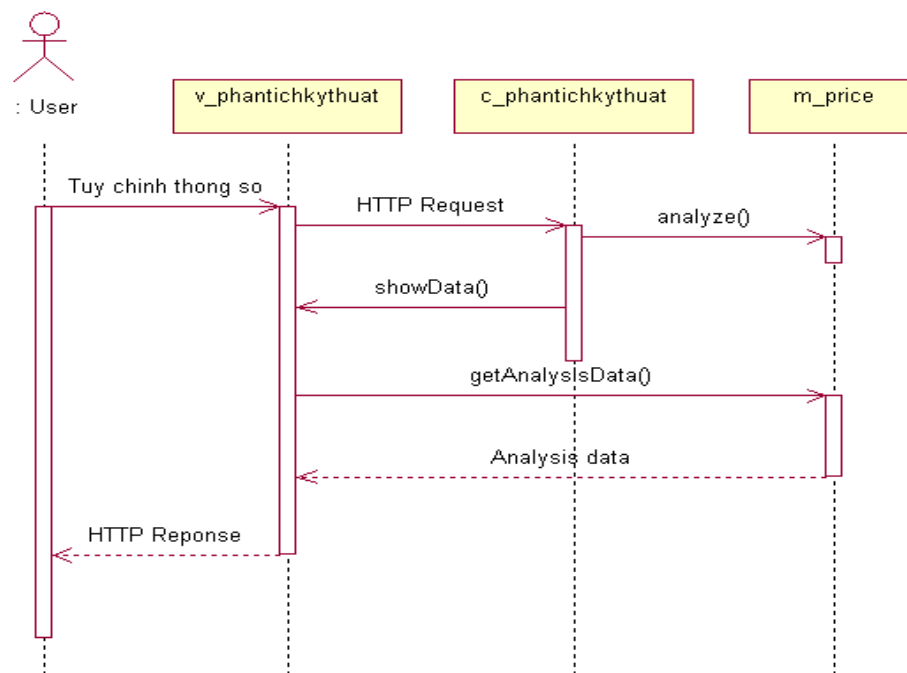
**Điều kiện bắt đầu:** đang ở trang phân tích và dự báo chứng khoán của một mã nhất định.

**Luồng sự kiện chính:**

Actor	System
1. Chọn loại công cụ	2. Hiển thị mô tả ý nghĩa và công dụng, các trường để điền thông số.
3. Điền thông số	
4. Nhấn Ok	5. Thể hiện thông số kỹ thuật trên biểu đồ.

**Điều kiện kết thúc:** thông số kỹ thuật được thể hiện trên biểu đồ.

**Sequence Diagram:**



Hình 10. SD – Phân tích kỹ thuật trên biểu đồ

### 1.2.4. Xem báo cáo tài chính đầy đủ

**Mô tả:** cho phép người dùng download bản báo cáo tài chính gốc về máy để xem.

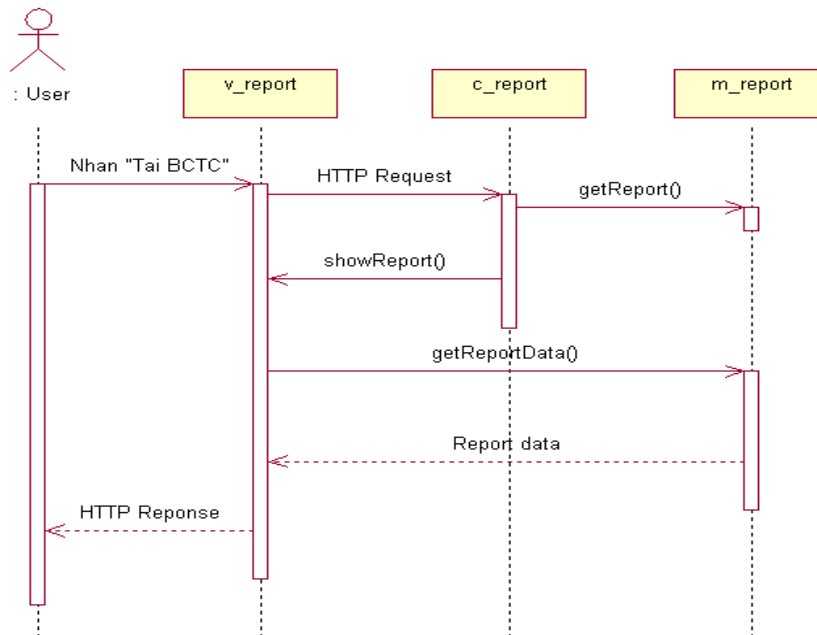
**Điều kiện bắt đầu:** đang ở trang phân tích và dự báo chứng khoán của một mã nhất định.

**Luồng sự kiện chính:**

Actor	System
1. Nhấn vào link “Tải báo cáo đầy đủ”.	2. Hiện thị danh sách liên kết đến báo cáo tài chính lưu trữ trong hệ thống.
3. Nhấn vào một link nhất định	4. Cho phép tải về máy.

**Điều kiện kết thúc:** báo cáo tài chính được tải về máy người dùng.

**Sequence Diagram:**



Hình 11. SD – Xem báo cáo tài chính đầy đủ

### 1.2.5. Tìm kiếm mã cổ phiếu

**Mô tả:** giúp người dùng tìm kiếm chứng khoán dựa trên mã hoặc tên công ty.

**Điều kiện bắt đầu:** đang ở trong hệ thống

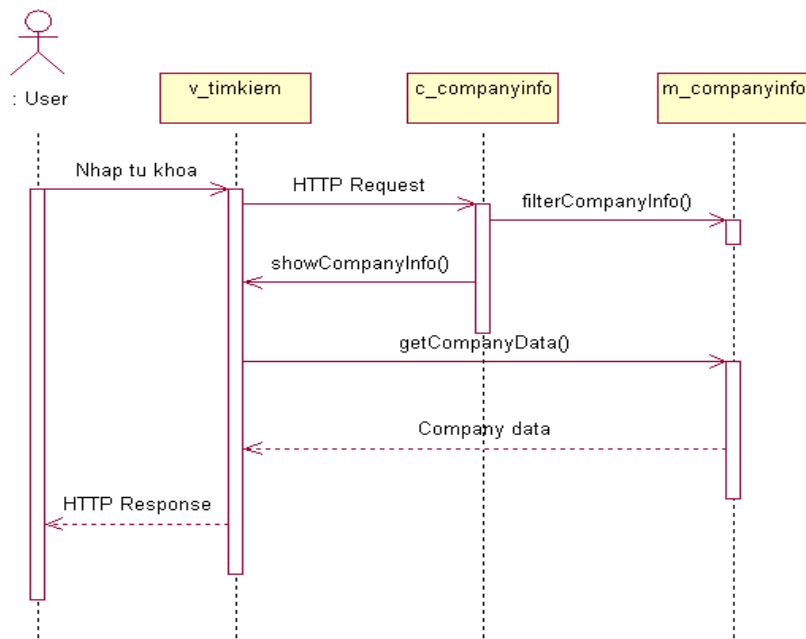
**Luồng sự kiện chính:**

Actor	System
1. Nhập thông tin ở khung tìm kiếm	2. Hiện thị danh sách kết quả phù hợp
3. Chọn mã cổ phiếu cần tìm	4. Hiện thị liên kết đến trang phân tích và dự báo cho mã cổ phiếu đó

**Luồng sự kiện phụ:** sau bước 1, nếu như danh sách kết quả là rỗng, thông báo không tìm thấy mã cổ phiếu phù hợp.

**Điều kiện kết thúc:** hiển thị liên kết đến các trang phân tích phù hợp với từ khóa.

**Sequence Diagram:**



Hình 12. SD – Tìm kiếm mã cổ phiếu

### 1.2.6. Lọc cổ phiếu

**Mô tả:** cho phép người dùng chọn lọc một bộ cổ phiếu dựa trên tập các thuộc tính nằm trong những khoảng nhất định.

**Điều kiện bắt đầu:** đang ở trong hệ thống.

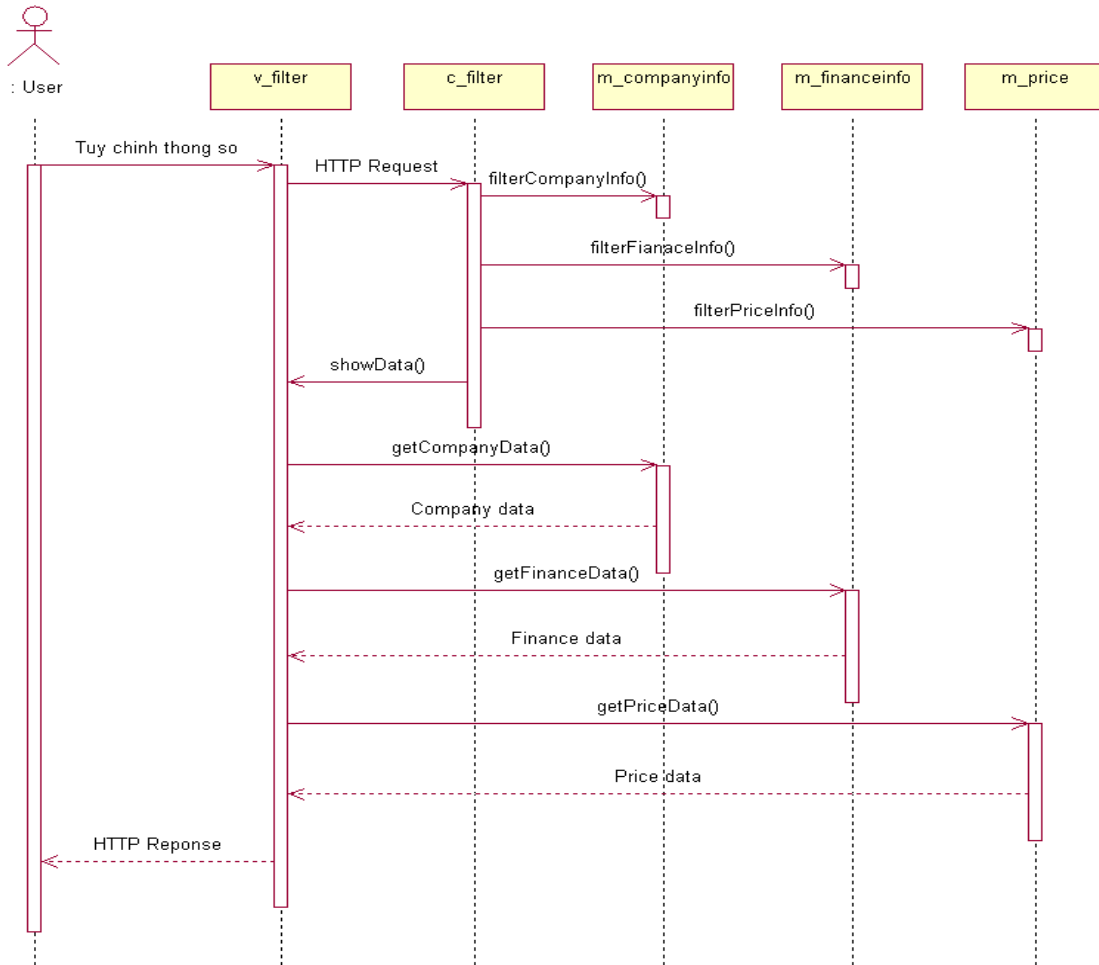
**Luồng sự kiện chính:**

Actor	System
1. Nhấn vào trang “Bộ lọc cổ phiếu”	2. Hiện thị danh sách tất cả những mã cổ phiếu trong hệ thống
3. Chọn các thuộc tính quan tâm	
4. Điều chỉnh miền giá trị cho các thuộc tính đã chọn	5. Hiện thị danh sách những mã cổ phiếu phù hợp

**Luồng sự kiện phụ:** sau bước 4, nếu như không có kết quả nào thỏa điều kiện lọc, hệ thống sẽ thông báo không tìm thấy kết quả phù hợp.

**Điều kiện kết thúc:** hiện thị danh sách những cổ phiếu có thuộc tính thỏa mãn yêu cầu lọc.

**Sequence Diagram:**



Hình 13. SD – Lọc cổ phiếu

### 1.2.7. Nhập liệu

**Mô tả:** đưa thông tin về giá của một mã cổ phiếu nhất định.

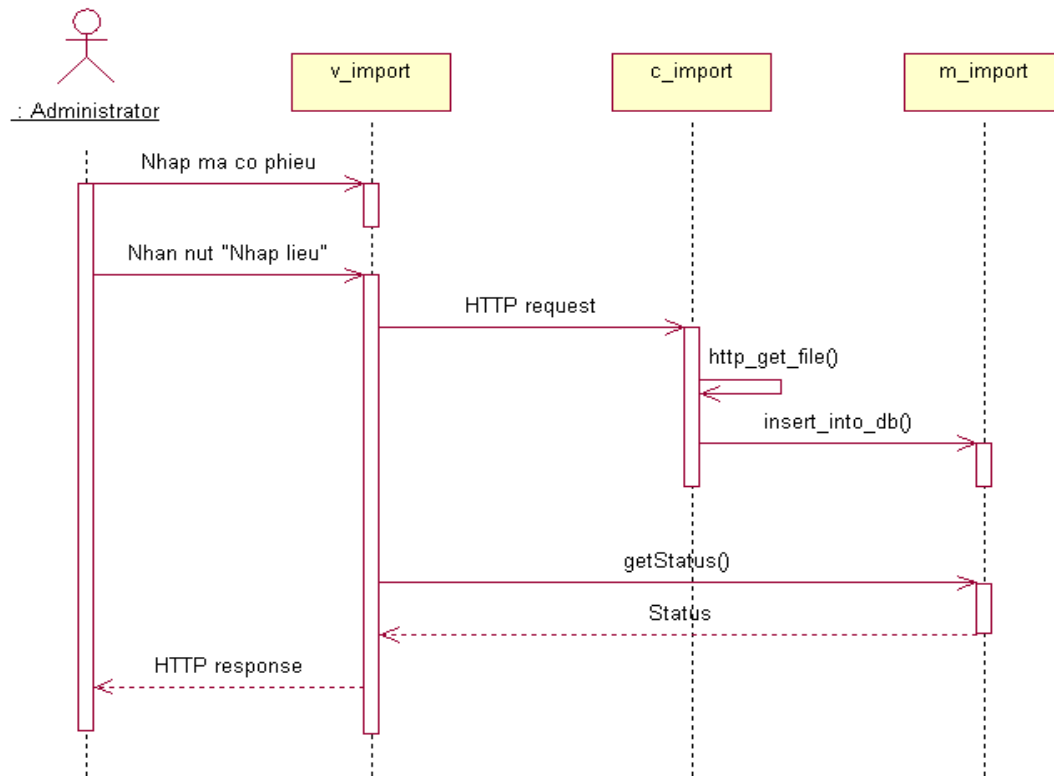
**Điều kiện bắt đầu:** đã đăng nhập vào trang quản lý của admin.

**Luồng sự kiện chính:**

Actor	System
1. Nhập mã cổ phiếu	
2. Nhấn nút “Nhập liệu”	3. Tải tập tin từ trang cophieu68.com và đưa vào cơ sở dữ liệu
	4. Hiện thị thông báo kết quả

**Điều kiện kết thúc:** dữ liệu giả được cập nhật vào cơ sở dữ liệu.

**Sequence Diagram:**



Hình 14. SD – Nhập liệu

### 1.2.8. Lấy tin

**Mô tả:** lấy RSS từ các trang tin chứng khoán để lưu trữ vào cơ sở dữ liệu.

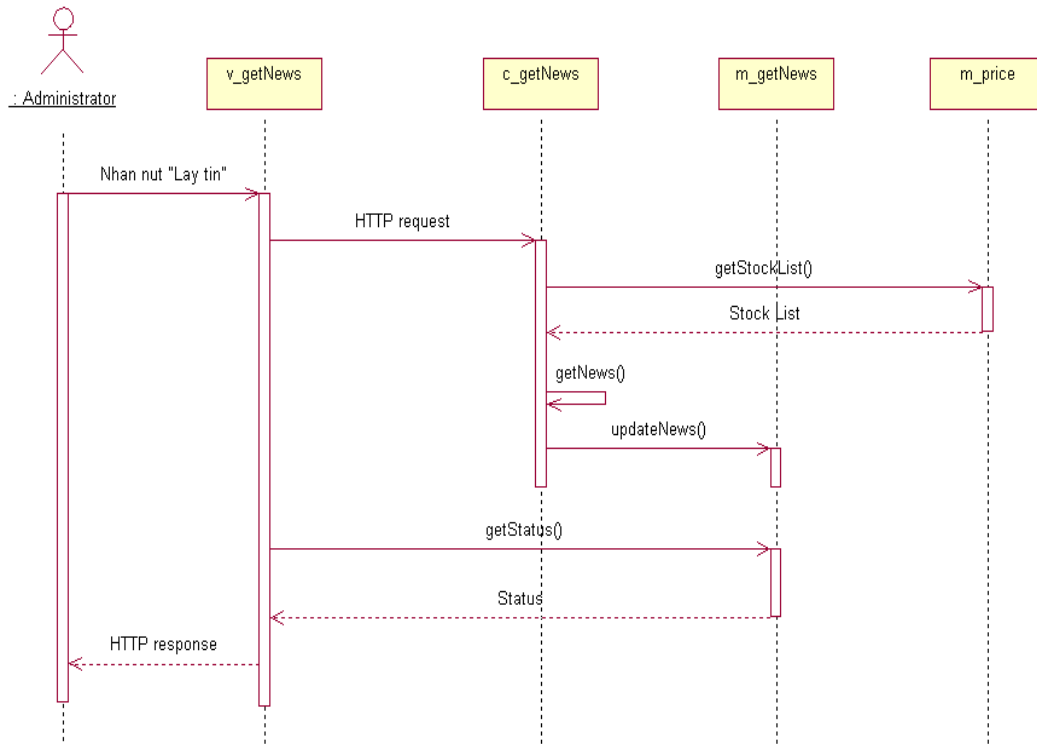
**Điều kiện bắt đầu:** đã đăng nhập vào trang quản lý của admin.

**Luồng sự kiện chính:**

Actor	System
1. Nhấn vào nút “Lấy tin”	2. Lấy thông tin từ các trang tổng hợp theo danh sách các mã cổ phiếu trong hệ thống
	3. Hiển thị thông báo số tin lấy về

**Điều kiện kết thúc:** tin tức mới nhất đã được cập nhật vào cơ sở dữ liệu.

### Sequence Diagram:



Hình 15. SD – Lấy tin

### 1.2.9. Lập mô hình

**Mô tả:** đưa ra mô hình ARIMA dựa trên dữ liệu giá cổ phiếu.

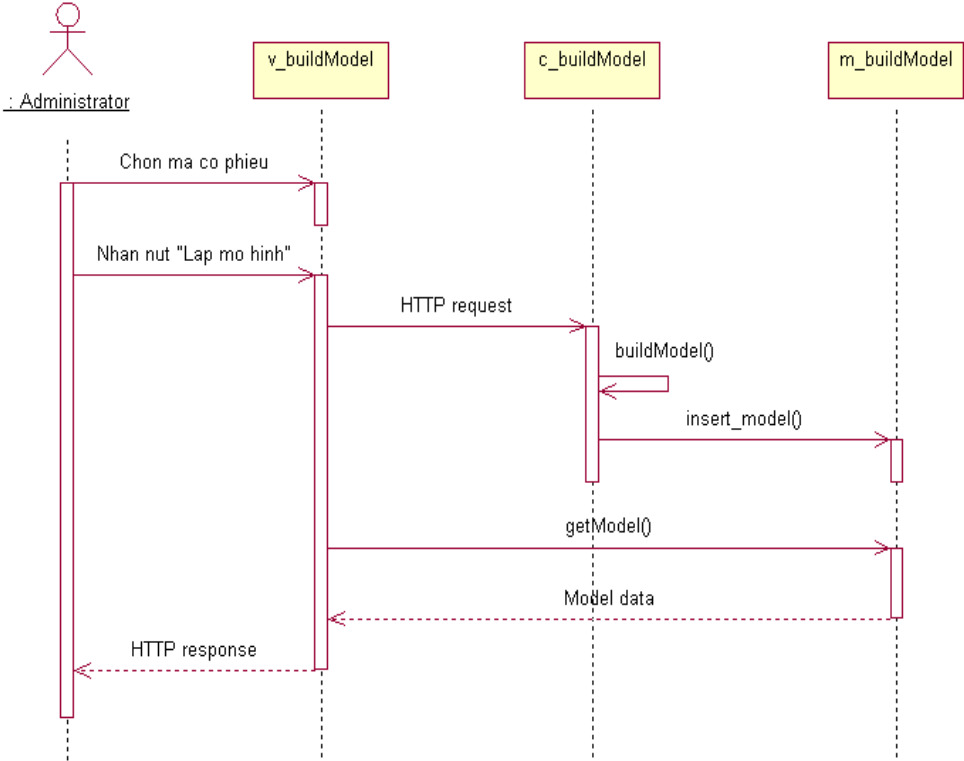
**Điều kiện bắt đầu:** đã đăng nhập vào trang quản lý của admin.

**Luồng sự kiện chính:**

Actor	System
1. Chọn mã cổ phiếu	
2. Nhấn nút “Lập mô hình”	3. Lập mô hình và lưu vào cơ sở dữ liệu
	4. Hiện thị mô hình vừa được lập

**Điều kiện kết thúc:** mô hình dự báo được lưu trữ vào cơ sở dữ liệu.

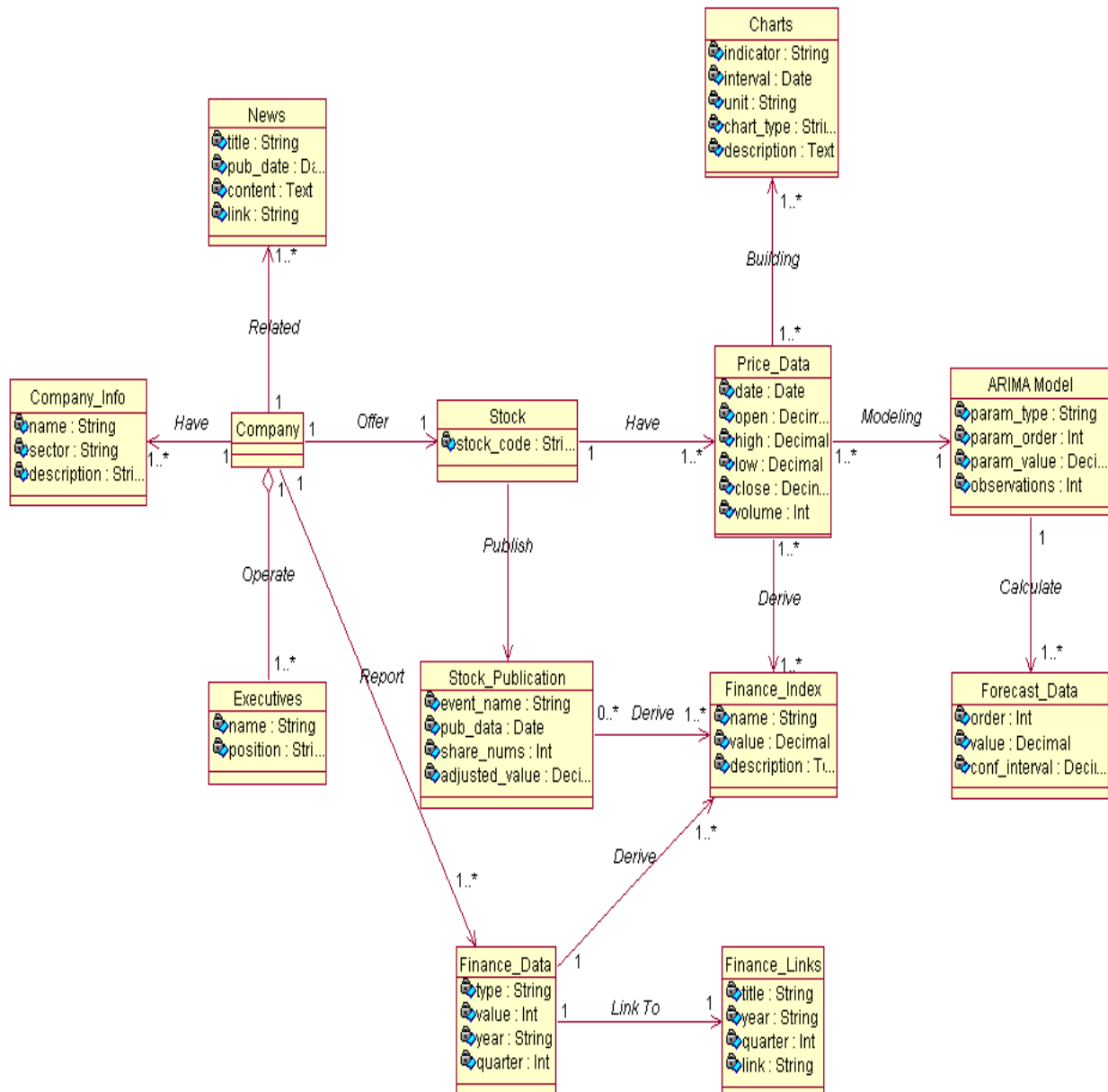
**Sequence Diagram:**



*Hình 16. SD – Lập mô hình*

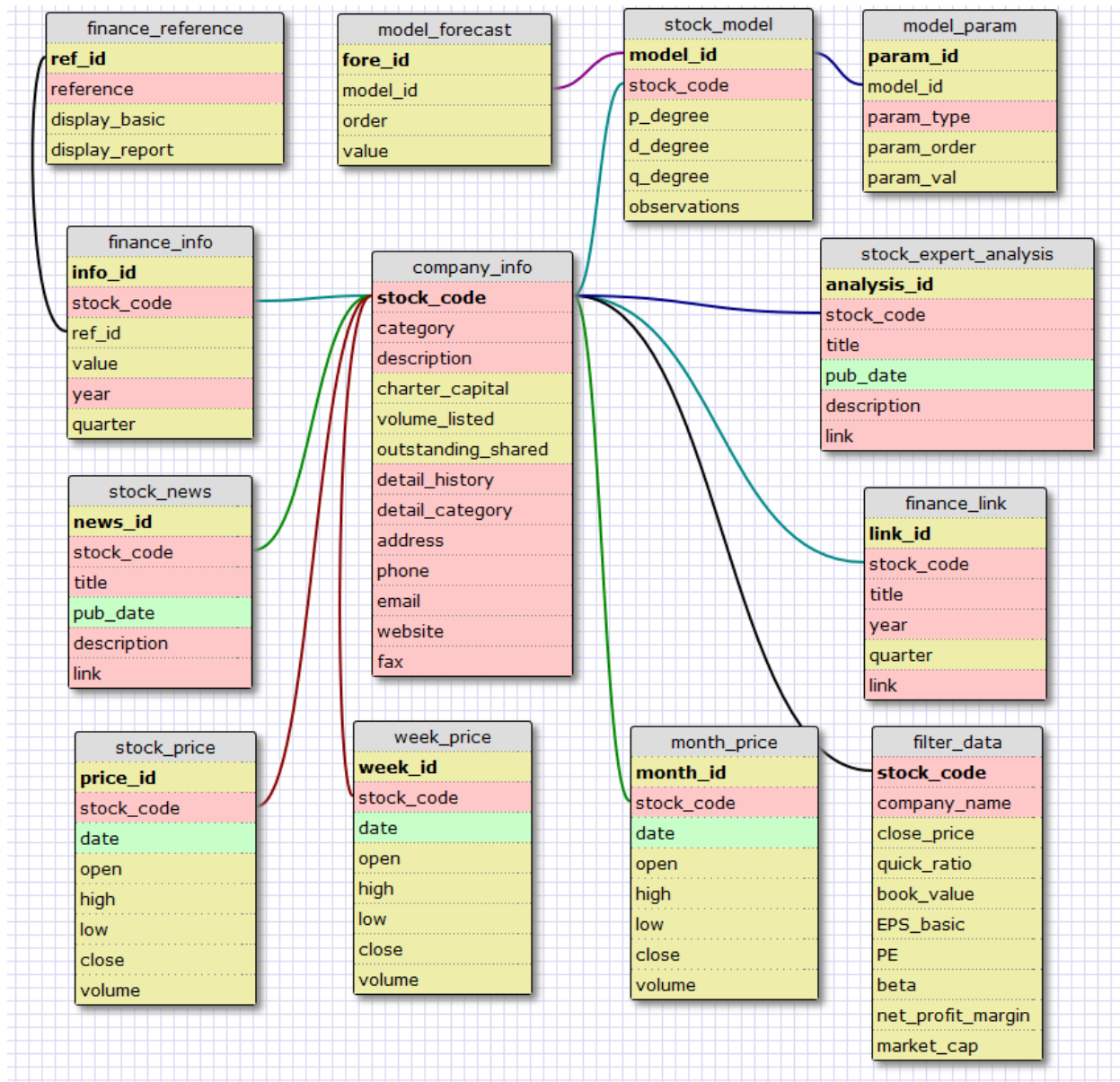


### 1.3. Domain Model



Hình 17. Domain Model

## 2. Cấu trúc cơ sở dữ liệu



Hình 18. Sơ đồ database

Tất cả các bảng đều có khóa chính kiểu int tăng dần và với các bảng chứa field stock\_code đều có khóa ngoại tham chiếu đến field stock\_code của bảng company\_info.

## company\_info

Lưu trữ các thông tin cơ bản của công ty, bao gồm:

Tên	Kiểu dữ liệu	Giải thích
stock_code	varchar (10)	Mã chứng khoán của công ty (khóa chính)
category	varchar (100)	Nhóm ngành
description	varchar (200)	Tên đầy đủ của công ty
charter_capital	varchar (25)	Vốn điều lệ
volume_listed	varchar (25)	Khối lượng cổ phiếu niêm yết
oustanding_shared	varchar (25)	Khối lượng cổ phiếu đang lưu hành
detail_category	text	Lĩnh vực kinh doanh
detail_history	text	Lịch sử công ty
address	text	Địa chỉ
phone	text	Số điện thoại
email	text	Địa chỉ email
website	text	Trang web của công ty
fax	text	Số fax

## finance\_reference

Là bảng lưu trữ tên gọi các thông tin tài chính.

Tên	Kiểu dữ liệu	Giải thích
ref_id	char (3)	Mã tham chiếu thông tin tài chính (khóa chính)
reference	varchar (100)	Tên tham chiếu của thông tin tài chính
display_basic	tinyint (1)	Phân biệt thông tin có phải là chỉ số cơ bản hay không
display_report	tinyint (1)	Phân biệt thông tin có phải thuộc báo cáo tài chính hay không

### finance\_info

Lưu trữ các giá trị tài chính cần thiết, có khóa ngoại *ref\_id* liên kết đến bảng *finance\_reference*.

Tên	Kiểu dữ liệu	Giải thích
info_id	integer	Khóa chính của mỗi record dữ liệu
stock_code	varchar (10)	Mã chứng khoán của công ty
ref_id	char (3)	Mã tham chiếu thông tin tài chính
value	varchar (25)	Giá trị
year	varchar (4)	Năm tài chính
quarter	integer	Quý trong năm

**Lưu ý:** bảng 1 và 2 có thể kết hợp lại thành một bảng duy nhất với mỗi cột đại diện cho một giá trị thông tin tài chính. Tuy nhiên cấu trúc như vậy sẽ gặp khó khăn trong việc chỉnh sửa các thông tin tài chính trong cơ sở dữ liệu. Do đó, nhóm thống nhất rằng sẽ lưu dưới dạng cặp name/value để dễ tùy chỉnh và nâng cấp. Thêm vào đó, bảng *finance\_reference* được thêm vào để giảm dữ liệu lặp.

### finance\_link

Lưu trữ các liên kết để tải báo cáo tài chính gốc.

Tên	Kiểu dữ liệu	Giải thích
link_id	integer	Khóa chính của mỗi record
stock_code	varchar (10)	Mã chứng khoán của công ty
link_name	varchar (100)	Tên đầy đủ của báo cáo gốc
link	varchar (300)	Liên kết tải về
year	varchar (4)	Năm tài chính
quarter	integer	Quý trong năm

### **stock\_news**

Các tin tức có liên quan tới chứng khoán và công ty tương ứng sẽ được lưu trữ bằng bảng này, bao gồm khóa chính và

<b>Tên</b>	<b>Kiểu dữ liệu</b>	<b>Giải thích</b>
news_id	integer	Khóa chính mỗi record tin tức
stock_code	varchar (10)	Mã chứng khoán của công ty
pub_date	datetime	Ngày đăng tin
title	varchar (200)	Tiêu đề bản tin
description	varchar (500)	Tóm tắt mẫu tin
link	varchar (300)	Liên kết tới trang gốc

### **stock\_expert\_analysis:**

Bảng stock\_expert\_analysis có cùng cấu trúc với bảng stock\_news, dùng để lưu trữ nhận định thị trường của các chuyên gia.

### **stock\_price**

Lưu trữ thông tin giá cổ phiếu trong các phiên giao dịch.

<b>Tên</b>	<b>Kiểu dữ liệu</b>	<b>Giải thích</b>
price_id	integer	Khóa chính mỗi record thông tin giá
stock_code	varchar(10)	Mã chứng khoán của công ty
date	date	Ngày giao dịch
open	decimal	Giá mở cửa
high	decimal	Giá trần
low	decimal	Giá sàn
close	decimal	Giá đóng cửa
volume	integer	Khối lượng giao dịch

### **week\_price, month\_price**

Bảng week\_price và month\_price có cùng cấu trúc và chức năng với bảng stock\_price, chỉ khác biệt về mặt ý niệm. Bảng stock\_price lưu giá theo ngày còn week\_price lưu theo tuần, month\_price lưu theo tháng. Lý do có sự tồn tại của hai bảng này là để tăng hiệu suất khi người dùng yêu cầu truy vấn dữ liệu giá của tuần và tháng hoặc những chỉ số tính toán có liên quan.

### **filter\_data**

Lưu trữ các thông tin dùng để lọc cổ phiếu, bao gồm

<b>Tên</b>	<b>Kiểu dữ liệu</b>	<b>Giải thích</b>
stock_code	varchar(10)	Mã chứng khoán của công ty
company_name	text	Tên công ty
close_price	decimal	Giá đóng cửa
quick_ratio	decimal	Chỉ số quick ratio
book_value	decimal	Giá trị sổ sách
EPS_basic	decimal	Lợi nhuận cơ bản trên một cổ phiếu
PE	decimal	Tỷ suất giá trên EPS
beta	decimal	Chỉ số thể hiện mức biến động của giá so với thị trường
net_profit_margin	decimal	Hệ số biên lợi nhuận ròng
market_cap	decimal	Lợi nhuận cơ bản trên một cổ phiếu

## stock\_model

Lưu trữ thông tin về mô hình dự báo của mã cổ phiếu.

Tên	Kiểu dữ liệu	Giải thích
model_id	integer	Khóa chính mỗi record mô hình dự báo
stock_code	varchar(10)	Mã chứng khoán của công ty
p_degree	integer	Bậc p
p_degree	integer	Bậc d
p_degree	integer	Bậc q
observations	integer	Số quan sát được dùng để đưa ra mô hình

**Lưu ý:** Việc tính toán đưa ra mô hình dự báo cho một mã cổ phiếu cần nhiều phép tính và thời gian, thêm vào đó, đối với những chuỗi thời gian đủ dài thì ta có thể lập mô hình tương đối hiệu quả mà không cần cập nhật liên tục các giá trị mới. Với những lý do trên, việc lập mô hình dự báo sẽ được tính toán từ trước và lưu trữ vào cơ sở dữ liệu để tăng hiệu suất và tính tiện dụng cho hệ thống.

## model\_param

Lưu trữ thông tin về tham số mô hình dự báo.

Tên	Kiểu dữ liệu	Giải thích
param_id	integer	Khóa chính mỗi record tham số mô hình dự báo
model_id	integer	Tham chiếu đến mô hình tương ứng
param_type	varchar(50)	Kiểu tham số
param_order	integer	Bậc tham số
param_val	decimal	Giá trị của tham số

## model\_forecast

Lưu trữ thông tin về giá trị dự báo dẫn xuất từ mô hình

Tên	Kiểu dữ liệu	Giải thích
fore_id	integer	Khóa chính mỗi record giá trị dự báo
model_id	integer	Tham chiếu đến mô hình tương ứng
order	integer	Thứ tự dự báo
value	decimal	Giá trị dự báo
confi_50	decimal	Khoảng tin cậy 50%
confi_95	decimal	Khoảng tin cậy 95%

**Lưu ý:** để tính toán sai số dùng trong mô hình dự báo, cần phải lưu lại kết quả dự báo từ những phiên trước. Khi mô hình được lập nên, giá trị dự báo cho 10 phiên sẽ được đưa vào bảng model\_forecast. Với 5 là mốc thời gian, thứ tự từ 1 – 5 là giá trị từ quá khứ cho đến hiện tại. Và từ 6 – 10 là các giá trị dự báo cho tương lai. Mỗi cuối ngày, giá trị dự báo sẽ tự động cập nhật theo phương pháp trượt (moving), cập nhật giá trị từ 2 – 6 cho 1 – 5, sau đó cập nhật dự báo cho giá trị từ 6 – 10. Cơ sở dữ liệu sẽ được giữ sao cho bất cứ lúc nào cũng chỉ có 10 giá trị dự báo tương ứng với một model.

### 3. Thu thập và xử lý dữ liệu

#### 3.1. Dữ liệu tài chính

Các dữ liệu tài chính liên quan phục vụ cho việc phân tích cơ bản tài chính doanh nghiệp chủ yếu được trích từ báo cáo tài chính theo quý của từng doanh nghiệp. Báo cáo tài chính chứa nhiều chỉ số tài chính cụ thể khác nhau, do đó, trong bước xử lý dữ liệu, cần phải trích lọc chỉ số nào cần thiết trong việc tính toán các chỉ số khác. Bảng dưới đây tổng hợp các chỉ số liên quan.

Bên cạnh đó, sẽ cần sử dụng tới bảng giá thường nhật của các mã chứng khoán (xem thêm mục 3.2)



Mã số	Chỉ số tài chính
100	Tài sản ngắn hạn
200	Tài sản dài hạn
300	Nợ phải trả
310	Nợ ngắn hạn
330	Nợ dài hạn
400	Vốn chủ sở hữu
1	Tổng doanh thu hoạt động kinh doanh
30	Lợi nhuận thuần từ hoạt động kinh doanh
50	Tổng lợi nhuận kế toán trước thuế
60	Lợi nhuận sau thuế thu nhập DN
61	Lợi ích của cổ đông thiểu số

### 3.2. Dữ liệu giá

Dữ liệu về giá chứng khoán có thể được xem như là dữ liệu quan trọng nhất đối với các nhà đầu tư, khi họ đầu tư vào một mã chứng khoán nào đó gần như đồng nghĩa với việc họ kỳ vọng rằng giá của mã cổ phiếu đó có thể tăng lên trong tương lai. Do đó, dữ liệu giá cổ phiếu là một chỉ số phân tích và dự báo chủ chốt trong hệ thống của chúng tôi. Hiện nay, dữ liệu giá cổ phiếu được cung cấp rộng rãi bởi nhiều trang thông tin chứng khoán, nổi bật trong đó là trang <http://cophieu68.com>, với kho dữ liệu chi tiết và chính xác được sắp xếp theo hệ thống rõ ràng. Đó cũng là những lý do vì sao nhóm chúng tôi chọn trang này làm nguồn cung cấp dữ liệu.

Để thuận tiện cho việc phát triển và nghiên cứu, việc nhập dữ liệu giá được tự động hóa. Từ một mã cổ phiếu bất kỳ do admin nhập, hệ thống sẽ download file dữ liệu giá định dạng csv của mã cổ phiếu tương ứng từ trang <http://cophieu68.com>. Sau đó, duyệt file và insert dữ liệu vào database.

Theo như chuẩn thì dữ liệu giá chứng khoán sẽ được ghi nhận theo ngày. Nhưng để thuận tiện cho việc phân tích và thể hiện thông tin, hệ thống sẽ ghi nhận lại dữ liệu giá theo tuần và theo tháng.

### **III. Giải pháp công nghệ**

#### **1. HTML 5**

HTML (Hyper text markup language – Ngôn ngữ đánh dấu siêu văn bản) là một ngôn ngữ đánh dấu văn bản được dùng để định nghĩa cấu trúc của một trang web.

Trong hơn 7 năm tồn tại của mình, HTML từ một ngôn ngữ đánh dấu văn bản đơn giản chỉ với một số ít các thẻ đã phát triển thành một ngôn ngữ tiêu chuẩn để định nghĩa các trang web. Phiên bản hoàn chỉnh mới nhất là HTML 4.01 được phát hành vào cuối năm 1999.

HTML 5 được giới thiệu lần đầu vào đầu năm 2008 và hiện tại vẫn đang còn trong giai đoạn phát triển, là phiên bản tiếp theo HTML 4.01. HTML 5 đem lại rất nhiều cải tiến so với HTML 4.01:

- Nhất quán trong xử lý lỗi. HTML có một đặc tính là cho dù được viết đúng hay sai thì trình duyệt vẫn luôn cố xử lý và thể hiện thành một trang web. Ở các phiên bản HTML trước, việc xử lý những đoạn mã sai cú pháp không được tổ chức W3C định nghĩa trước mà mỗi trình duyệt có một cách xử lý khác nhau tạo nên sự không nhất quán trong cách thể hiện trang web. HTML 5 đã xử lý được vấn đề này.
- Biến trình duyệt web thành một nền tảng phát triển ứng dụng mới. Với sự bổ sung thêm của các thẻ <video>, <audio> và <canvas>, một trang web có thể hiển thị các nội dung phim ảnh, âm thanh hay các hoạt cảnh mà không cần phụ thuộc các công cụ của hãng thứ 3 như Adobe Flash hay Microsoft Silverlight.
- Bổ sung thêm những thẻ mới như thẻ <section> để định nghĩa những thành phần chính của trang web, thẻ <article> để định nghĩa các bài báo, ... giúp cho cấu trúc của trang web trở nên trực quan, dễ đọc, dễ hiểu hơn.

#### **2. CSS 3**

CSS (Cascading style sheet - tập tin định kiểu theo tầng) là ngôn ngữ dùng để định nghĩa cách thể hiện một trang web trên trình duyệt.

CSS 3 là phiên bản mới nhất của CSS, được giới thiệu lần đầu vào năm 1999. Khác với các phiên bản trước, CSS 3 được chia thành nhiều module và từng module sẽ trải qua nhiều giai đoạn kiểm tra bởi W3C trước khi được triển khai rộng rãi. Điều này tạo điều kiện để CSS 3

được hỗ trợ bởi nhiều trình duyệt hơn. Ngoài ra CSS 3 cũng có nhiều ưu điểm so với các phiên bản trước:

- Bổ sung, cải tiến khả năng xác định các đối tượng thông qua thuộc tính.
- Bổ sung thêm nhiều pseudo class hữu dụng (pseudo class được dùng để xác định các đối tượng đặc biệt, một số pseudo class: first-child, last-child, hover, active, ... Ví dụ: p:last-child sẽ chọn ra thẻ <p> cuối cùng trong trang web).
- Bổ sung thêm những khả năng mới như cho phép sử dụng kết hợp nhiều hình nền trong trang web, sử dụng hình ảnh để tạo đường viền cho các đối tượng, ...
- Bổ sung nhiều thuộc tính mới như border-radius cho phép các đối tượng có các góc được bo tròn, box-shadow để tạo hiệu ứng đổ bóng, ... Những thuộc tính mới này cho phép các trang web ít phụ thuộc vào hình ảnh để tạo các hiệu ứng đặc biệt, từ đó tăng được tốc độ tải trang.

### 3. AJAX

AJAX (Asynchronous javascript and XML – Javascript và XML không đồng bộ) là một phương pháp kết hợp các công nghệ HTML, CSS, DOM, XML, Javascript, XMLHttpRequest để xây dựng các ứng dụng mang tính tương tác cao trên nền web.

Khi sử dụng AJAX, những thành phần của một trang web có thể được thay đổi tức thì theo thao tác của người dùng mà không cần phải tải lại toàn bộ trang web. Đây là một lợi thế lớn của AJAX do nó giảm được đáng kể thời gian chờ của người dùng, tăng tính tương tác cho trang web. Nhờ vậy AJAX đã mở ra nhiều khả năng mới cho việc phát triển, xây dựng các ứng dụng web, cho phép người dùng tương tác và sử dụng trang web như sử dụng phần mềm máy tính.

### 4. JQuery

jQuery là một trong những thư viện javascript phổ biến nhất hiện nay. jQuery được thiết kế với mục tiêu đơn giản hóa công việc lập trình javascript.

jQuery là một thư viện miễn phí, mã nguồn mở. Cú pháp của jQuery được thiết kế để tạo sự đơn giản trong việc tìm kiếm, thay đổi các đối tượng HTML trong trang web, tạo các hoạt cảnh, xử lý các sự kiện và tạo các ứng dụng AJAX.

Một điểm mạnh khác của jQuery là nó cho phép các lập trình viên viết thêm những plug-in và gắn vào jQuery, bổ sung thêm chức năng cho nó mà không cần phải thay đổi mã nguồn. Điều này khiến cho khả năng mở rộng của jQuery trở nên vô tận và thu hút được rất nhiều người dùng.

#### IV. Giới thiệu chương trình

##### 1. Cấu trúc trang web

Trang chủ	Phân tích và dự báo	Trợ giúp	Tìm kiếm <input type="text"/>
Thông tin công ty			
Phân tích kỹ thuật		Tin tức	
Phân tích cơ bản		Nhận định thị trường	
Dự báo		Diễn biến thị trường	

Hình 19. Cấu trúc trang web

Trang web gồm 3 phần chính:

**Navigation:** được đặt trên đầu trang web, chứa đường dẫn đến những trang chính của website là trang chủ, trang phân tích dự báo và trang trợ giúp.

**Nội dung chính:** được chia thành 3 phần:

- Thông tin công ty: chứa các thông tin cơ bản của công ty đang xem, gồm:
  - Tên công ty, mã cổ phiếu.
  - Giá đóng cửa phiên mới nhất.
  - Giá dự báo của phiên tiếp theo.
  - Nhóm ngành của công ty.
  - Vốn điều lệ.
  - EPS cơ bản.
  - Khối lượng giao dịch của phiên mới nhất.

- Giá cao nhất trong phiên giao dịch mới nhất.
- Giá thấp nhất trong phiên giao dịch mới nhất.
- Giá cao nhất trong 52 tuần kể từ hiện tại.
- Giá thấp nhất trong 52 tuần kể từ hiện tại.
- Phân tích kỹ thuật: chứa biểu đồ về giá, khối lượng giao dịch và các chỉ số phân tích kỹ thuật.
- Phân tích cơ bản: chứa các chỉ số phân tích cơ bản, báo cáo tài chính 4 quý mới nhất và các biểu đồ về tài chính đáng quan tâm.
- Dự báo: chứa dữ liệu dự báo cho 5 phiên giao dịch tiếp theo gồm giá dự báo, khoảng tin cậy 50% và 95%. Ngoài ra những dữ liệu này còn được thể hiện thành biểu đồ để tăng tính trực quan.

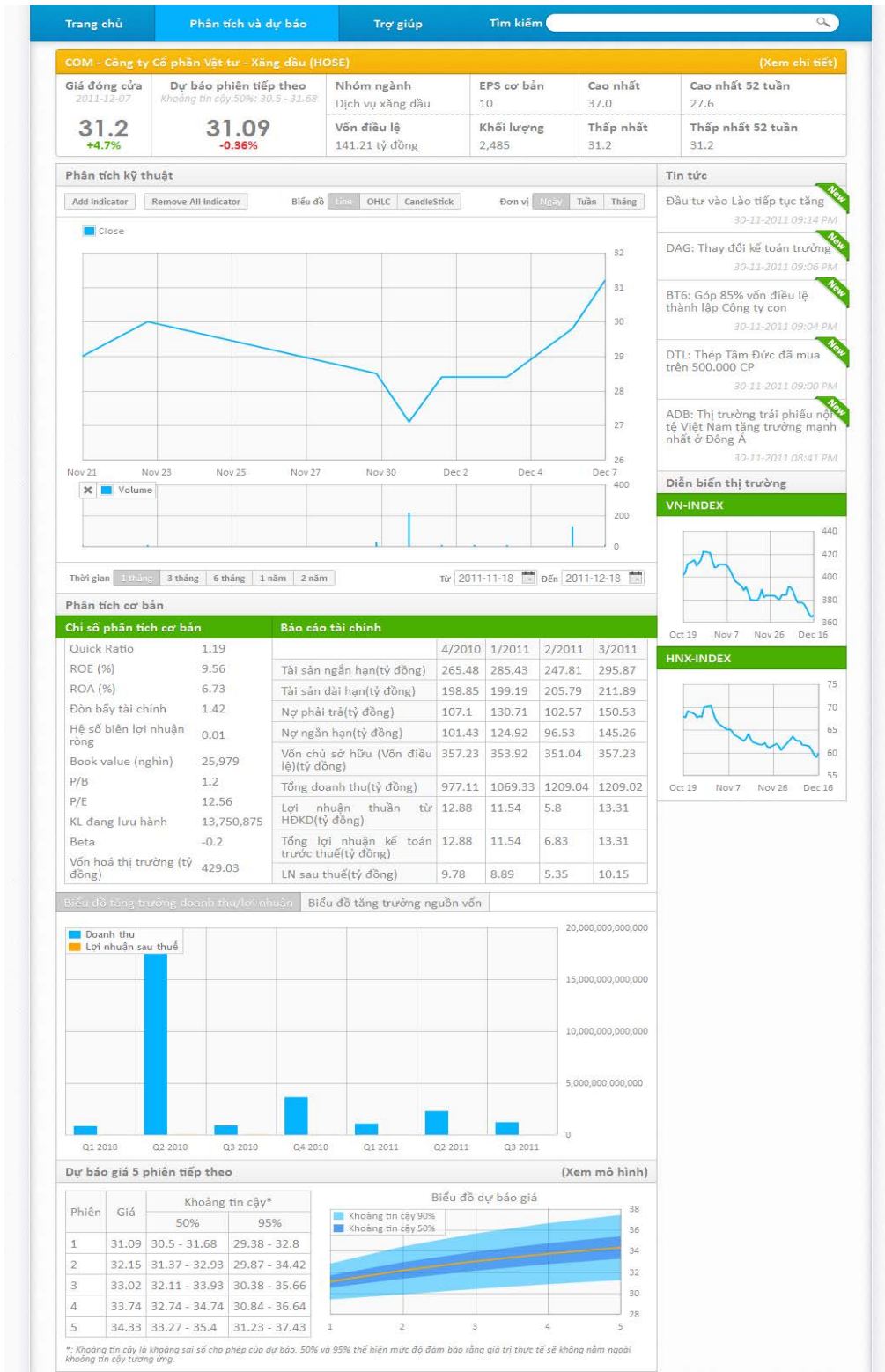
**Nội dung phụ:** được chia thành 2 phần:

- Tin tức: thể hiện các tin kinh tế mới nhất.
- Nhận định thị trường: hiển thị phân tích của các chuyên gia.
- Diễn biến thị trường: gồm biểu đồ thể hiện 2 chỉ số thị trường là VN-INDEX và HNX-INDEX.

## 2. Chức năng chính

- Cung cấp thông tin cơ bản và chi tiết của các công ty.
- Cung cấp thông tin về phân tích kỹ thuật qua các biểu đồ và các chỉ số kỹ thuật.
- Cung cấp thông tin về phân tích cơ bản qua các chỉ số cơ bản, các báo cáo và biểu đồ tài chính.
- Cung cấp dự báo giá cho các phiên giao dịch trong tương lai.
- Cung cấp các tin tức mới nhất về thị trường.
- Cung cấp thông tin về diễn biến thị trường chứng khoán qua 2 chỉ số VN-INDEX và HNX-INDEX.
- Cho phép người dùng tìm kiếm các công ty theo tên và theo mã cổ phiếu.
- Cho phép người dùng tìm các công ty bằng cách lọc theo các chỉ số tài chính và giá đóng cửa.

### 3. Giao diện hoàn chỉnh



Hình 20. Giao diện trang phân tích dự báo

## **4. Hướng dẫn sử dụng**

### **4.1. Phân tích kỹ thuật**

#### **4.1.1. Thêm chỉ số kỹ thuật**

Bước 1: đưa trỏ chuột vào nút Add indicator, danh sách các chỉ số sẽ xuất hiện ngay bên dưới.

Bước 2: chọn một chỉ số để thêm vào biểu đồ, một khung nhập dữ liệu sẽ xuất hiện.

Bước 3: nhập thông số cho chỉ số đã chọn.

Bước 4: nhấn OK.

#### **4.1.2. Bỏ một chỉ số kỹ thuật**

Nhấn vào tên của chỉ số cần bỏ (tên của các chỉ số nằm ở trên biểu đồ giá) hoặc nhấn vào biểu tượng mang dấu “X” ở góc trên cùng bên trái của các biểu đồ chỉ số được vẽ riêng.

### **4.2. Tìm kiếm công ty**

Nhập từ khóa cần tìm vào khung tìm kiếm ở góc phải trên cùng của trang web. Kết quả sẽ được hiển thị ngay dưới khung tìm kiếm. Người dùng có thể sử dụng các phím lên, xuống để chọn lựa giữa các kết quả sau đó nhấn Enter hoặc nhấp chuột để đến trang phân tích của công ty cần tìm.

### **4.3. Lọc mã cổ phiếu**

Điền những thông số phù hợp vào các trường dữ liệu tương ứng để bắt đầu lọc. Người dùng có thể nhấn biểu trưng refresh ở các trường dữ liệu để khôi phục dữ liệu mặc định hoặc nhấn nút Refresh All để lấy lại bộ dữ liệu ban đầu.

### **4.4. Xem thông tin chi tiết công ty**

Nhấn vào link “Xem chi tiết” nằm ở bên phải tên công ty trong trang phân tích và dự báo.

# TỔNG KẾT

## I. Đánh giá tổng quan

Nhìn chung, hệ thống đã thể hiện được những thông tin cơ bản trong việc phân tích thị trường chứng khoán và đáp ứng nhu cầu xem dự báo giá cho các phiên giao dịch trong tương lai.

Có nhiều phương pháp và cách thức phân tích thị trường chứng khoán. Trong phạm vi khóa luận, chúng tôi đã cố gắng hiện thực hầu hết các hình thức phân tích thị trường chứng khoán. Có thể nói hệ thống mà nhóm đã xây dựng trải rộng hầu hết các phương pháp, bao gồm: phân tích chỉ số tài chính cơ bản, phân tích biểu đồ giá kỹ thuật và dự báo dựa trên quy luật vận động của chuỗi biến động giá. Mỗi phương pháp phân tích đều có những điểm mạnh và điểm yếu riêng, chúng không đối nghịch nhau, mà ngược lại mang mối quan hệ tương hỗ, bù trừ lẫn nhau. Ở đây, chúng tôi cố gắng đưa ra những phương pháp cơ bản nhất dưới một góc độ trung lập, sự tổng quát này sẽ cung cấp những thông tin có giá trị tùy theo nhiều góc nhìn khác nhau, từ đó hỗ trợ người đầu tư đưa ra quyết định đúng đắn.

Nếu như nói phân tích cơ bản là xác định giá trị tiềm năng của một công ty để đầu tư thì phân tích kỹ thuật sẽ đóng vai trò chủ chốt trong việc xác định thời điểm (*timing*) xuất nhập thị trường thông qua việc xác định tín hiệu và xu hướng giá. Giá trị dự báo được dùng làm giá trị tham khảo và xác nhận quy luật vận động của giá thị trường, khi giá trị thực tế vượt ngoài khoảng tin cậy có nghĩa là đang có những biến động bất thường báo hiệu đang có một sự kiện lớn sắp xảy ra hoặc một quy luật vận động mới sắp bắt đầu.



## II. Hướng phát triển

- Phát triển thuật toán tự tìm kiếm và truy hồi báo cáo tài chính trên mạng.
- Xây dựng một hệ thống gồm các chỉ số cụ thể hơn về phân tích cơ bản lẫn kỹ thuật, từ đó nâng cao độ tin cậy trong việc đánh giá – nhận định cổ phiếu trên các chỉ tiêu do hệ thống hỗ trợ.
- Hỗ trợ tính năng tự động xác định các tín hiệu đặc biệt để kịp thời cảnh báo các biến động thị trường, đồng thời ghi nhận các sự kiện có liên quan như: chia cổ phiếu thưởng, phát hành thêm cổ phiếu, mua lại cổ phiếu quỹ... và thể hiện chúng trên biểu đồ.
- Hỗ trợ các chức năng vẽ và đánh dấu để người dùng có thể tương tác trên biểu đồ.
- Tìm cách parse những tập tin RSS phức tạp để phân tin tức thêm phong phú và không bị phụ thuộc vào một nguồn nhất định.
- Sử dụng các giải thuật AI (Neuron Network, Decision Tree, Genetic Algorithm...) để đánh giá tổng quan các yếu tố biến động thị trường, từ đó đưa ra dạng mô hình dự báo thuyết phục hơn.
- Cải tiến việc lập mô hình bằng phương pháp *Phân tích có can thiệp* (intervention analysis)
- Tiếp tục phát triển thuật toán lập mô hình đối với chuỗi thời gian đa biến (multivariate) từ mô hình đơn biến (univariate) hiện tại. Do trong thực tế có nhiều chỉ số khác có thể ảnh hưởng tới giá trong tương lai chẳng hạn như khối lượng giao dịch, khối lượng niêm yết.

## TỪ ĐIỂN THUẬT NGỮ

Thuật ngữ	Miêu tả
Mua quá mức	Hiện tượng giá tăng lên rất cao, cao hơn giá trị thực của cổ phiếu. Do đó trong những phiên giao dịch tiếp theo, giá sẽ giảm về mức bình thường.
Bán quá mức	Hiện tượng giá giảm xuống rất thấp, thấp hơn giá trị thực của cổ phiếu. Do đó trong những phiên giao dịch tiếp theo, giá sẽ tăng về mức bình thường.
Phân kỳ	Hiện tượng xảy ra khi đường giá và các chỉ số phân tích kỹ thuật di chuyển ngược chiều nhau. Phân kỳ là dấu hiệu cho sự đảo chiều của xu hướng giá hiện tại.
Hội tụ	Hiện tượng xảy ra khi đường giá và các chỉ số phân tích kỹ thuật di chuyển tập trung về cùng một điểm
Tài sản ngắn hạn ( <i>Current Assets</i> )	Bao gồm tiền mặt, tài khoản nhận được, đầu tư, chứng khoán có thể bán được, chi phí trả trước hoặc những tài sản có tính thanh khoản khác.
Tài sản cố định vô hình	Là những tài sản không có hình thái vật chất, thể hiện một lượng giá trị đã được đầu tư thoả mãn các tiêu chuẩn của tài sản cố định vô hình, tham gia vào nhiều chu kỳ kinh doanh như: một số chi phí liên quan trực tiếp tới đất sử dụng, chi phí về quyền phát hành, bằng phát minh, bằng sáng chế, bản quyền tác giả...
Vốn chủ sở hữu ( <i>Owners' equity</i> )	Là các nguồn vốn thuộc sở hữu của chủ doanh nghiệp và các thành viên trong công ty liên doanh hoặc các cổ đông trong các công ty cổ phần. Bao gồm số tiền góp vốn của các nhà đầu tư, tổng số tiền tạo ra từ kết quả hoạt động sản xuất kinh doanh (lợi nhuận chưa phân phối) và chênh lệch đánh giá lại tài sản.
Nợ ngắn hạn ( <i>Current liabilities</i> )	Một khoản nợ của công ty hay một nghĩa vụ nợ được xác định trong khoảng thời gian 1 năm (năm tài chính). Tài sản nợ ngắn hạn trên bảng cân đối kế toán của công ty bao gồm những khoản nợ ngắn hạn, tài khoản có thể phải trả, các khoản nợ tích lũy và các loại khoản nợ khác.

<b>Thuật ngữ</b>	<b>Miêu tả</b>
Hàng tồn kho ( <i>Inventory</i> )	Là những tài sản được giữ để bán trong kỳ sản xuất - kinh doanh bình thường; đang trong quá trình sản xuất - kinh doanh dở dang; nguyên liệu, vật liệu, công cụ, dụng cụ để sử dụng trong quá trình sản xuất - kinh doanh hoặc cung cấp dịch vụ.
Cổ phiếu lưu hành hiện tại ( <i>Outstanding shares</i> )	Số cổ phần hiện đang được các nhà đầu tư nắm giữ, bao gồm cả các cổ phiếu giới hạn được sở hữu bởi nhân viên của công ty hay cá nhân khác trong nội bộ công ty.
Bảng cân đối kế toán ( <i>Balance sheet</i> )	Cho biết tình trạng tài sản của công ty, nợ và vốn cổ đông vào một thời điểm ấn định.
Lợi nhuận ròng ( <i>Net income</i> )	Là tổng thu nhập của một doanh nghiệp, được tính toán từ tổng thu nhập có điều chỉnh thêm các khoản chi phí hoạt động, khấu hao, lãi suất, thuế và các chi phí khác liên quan đến hoạt động kinh doanh của doanh nghiệp.
Độ trễ ( <i>Lag</i> )	Khoảng cách giữa một quan sát và quan sát kế tiếp. Vd: với độ trễ là 4 thì nếu quan sát hiện tại là 1 thì quan sát kế tiếp sẽ là 5.
Hàm tự tương quan ( <i>Autocorrelation Function</i> )	Là hiệp phương sai của một quan sát với quan sát trước đó (cùng một đối tượng, với trục hoành là thời gian) thể hiện sự tương đồng của hàm thời gian giữa các quan sát với nhau
Hàm tự tương quan riêng phần ( <i>Partial Autocorrelation Function</i> )	Tương tự như hàm tự tương quan, là hiệp phương sai của một quan sát với quan sát trước đó. Tuy nhiên, chữ riêng phần (partial) ở đây thể hiện sự khác biệt giữa chúng, đó là hiệp phương sai này sẽ loại bỏ đi sự ảnh hưởng của các quan sát trung gian nếu có. Vd: độ trễ k là 4, quan sát hiện tại là 1 và quan sát kế tiếp là 5, chúng ta sẽ tính hiệp phương sai giữa 1 và 5, trừ đi hiệp phương sai giữa 2 và 5; 3 và 5; 4 và 5.

Thuật ngữ	Miêu tả
Giả thuyết null ( <i>Null Hypothesis</i> )	Một khái niệm về thống kê, thường dùng trong các phép kiểm định. Giả thuyết null được xem như là mặc định của một trường hợp. Người ta sẽ thu thập dữ liệu để chứng minh giả thuyết này. Có hai trường hợp xảy ra, thứ nhất giả thuyết null đúng và được chấp nhận; thứ hai, không đủ dữ kiện để chứng minh giả thuyết null đúng, khi đó chúng ta tạm thời lựa chọn giả thuyết thay thế ( <i>alternative hypothesis</i> ).
AIC( <i>Akaike Information Criteria</i> ), BIC( <i>Bayesian Information Criteria</i> )	Chỉ số đánh giá mức độ thông tin sai lệch khi dùng một mô hình toán học để suy diễn thực tế
Mẫu ( <i>Sample</i> )	Một tập giá trị con được lựa chọn tượng trưng cho toàn bộ tập giá trị ban đầu.
Phân phối Chi bình phương ( <i>Chi Squared Distribution</i> )	Phân phối với tổng bình phương của các độ tự do $k$ ( <i>k degrees of freedom</i> ), trong đó $k$ là các biến ngẫu nhiên chuẩn độc lập. Đây là phân phối được sử dụng rất rộng rãi trong các phép kiểm định xác suất thống kê.
Khoảng tin cậy ( <i>Confidence</i> )	Khoảng sai số cho phép của dự báo. Thể hiện mức độ đảm bảo rằng giá trị thực tế sẽ không nằm ngoài khoảng tin cậy tương ứng.

# PHỤ LỤC

Bảng 1.

Ngày 1	Ngày 2	Ngày 3	Ngày 4	Ngày 5
10	5	7	14	3

Bảng 2. Kết quả ví dụ MACD

Ngày	Giá	EMA(8)	EMA(5)	MACD	Đường tín hiệu
1	89.5	89.12	89.70	0.57	0.45
2	89.5	89.20	89.63	0.42	0.37
3	90	89.38	89.75	0.37	0.30
4	90	89.51	89.83	0.31	0.22
5	89.5	89.51	89.72	0.21	0.14
6	89.5	89.51	89.65	0.13	-0.04
7	89.5	89.50	89.60	0.09	-0.44
8	85.5	88.62	88.24	-0.37	-0.88
9	81.5	87.05	86.02	-1.03	-1.22
10	82	85.94	84.69	-1.25	
11	81	84.85	83.47	-1.38	

Bảng 3.

Ngày	Giá high	Giá low	Giá đóng cửa
1	128.64	127.4	128.27
2	129.14	127.49	128.69
3	128.65	126.9	127.6
4	128.09	127	127.6
5	128.47	127.6	128.11
6	129.29	127.61	127.93
7	129.12	128.06	128.6

Bảng 4 – Công thức tính các chỉ số.

Tên trường dữ liệu	Công thức
<i>Nhóm chỉ số định giá</i>	
EPS cơ bản	$EPS = \frac{\text{Tổng lợi nhuận sau thuế 4 quý gần nhất}}{\text{Tổng số lượng cổ phiếu lưu hành}}$
P/E cơ bản	$P/E = \frac{\text{Giá trị thị trường}}{\text{Lợi nhuận ròng trên một cổ phần (EPS)}}$
Book value	$\text{Book value} = \frac{\text{Vốn chủ sở hữu - Tài sản cố định vô hình}}{\text{KLCP lưu hành hiện tại}}$
P/B	$P/B = \frac{\text{Giá đóng cửa}}{\text{Book value}}$
Beta	$\beta = \frac{\text{covar}(R_i, R_m)}{\text{var}(R_m)}$

Vốn hóa thị trường (Market capitalization)	$\text{Market capitalization} = \text{Giá đóng cửa} \times \text{Tổng số lượng cổ phiếu lưu hành}$
<i>Nhóm hiệu quả quản lý</i>	
ROE (Return On Equity)	$\text{ROE} = \frac{\text{Lợi nhuận sau thuế thu nhập doanh nghiệp}}{\text{Vốn chủ sở hữu}} \times 100$
ROA (Return On Assets)	$\text{ROA} = \frac{\text{Lợi nhuận ròng}}{\text{Tổng tài sản}} \times 100$
<i>Nhóm sức khỏe tài chính</i>	
Đòn bẩy tài chính (Financial Leverage)	$\text{Financial Leverage} = 1 + \frac{\text{Tổng nợ}}{\text{Vốn chủ sở hữu}}$
Quick Ratio	$\text{Quick ratio} = \frac{\text{Tài sản ngắn hạn} - \text{Hàng tồn kho}}{\text{Tổng nợ ngắn hạn}}$
<i>Nhóm chỉ số lợi nhuận</i>	
Hệ số biên lợi nhuận ròng (Net Profit Margin)	$\text{Net Profit Margin} = \frac{\text{Lợi nhuận ròng}}{\text{Doanh thu}}$

# TÀI LIỆU THAM KHẢO

- [1] “*Time Series Analysis: forecasting and control*”, Revised Edition 1976, George E. P. Box and Gwilym M. Jenkins.
- [2] “*Tính dừng (tĩnh tại), các nghiệm đơn vị và tính đồng kết hợp*”, Chương trình giảng dạy kinh tế FullBright của Gujarati D.N. chương 21 Cao Hào Thi và các cộng sự dịch.
- [3] “*Dự báo với mô hình ARIMA và VAR*”, Chương trình giảng dạy kinh tế FullBright của Gujarati D.N. chương 22 Cao Hào Thi và các cộng sự dịch.
- [4] “*Basic Econometrics*”, Time series econometrics: some basic concepts, Gujarati Damodar.
- [5] “*Akaike information criterion*”, truy cập tại địa chỉ [http://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](http://en.wikipedia.org/wiki/Akaike_information_criterion)
- [6] “*Bayesian information criterion*”, truy cập tại địa chỉ [http://en.wikipedia.org/wiki/Bayesian\\_information\\_criterion](http://en.wikipedia.org/wiki/Bayesian_information_criterion)
- [7] “*Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship*”, Henry de-Graft Acquah, truy cập tại địa chỉ <http://www.academicjournals.org/jdae/PDF/Pdf2010/Jan/Acquah.pdf>
- [8] “*AIC vs BIC*”, truy cập tại địa chỉ <http://emdbolker.wikidot.com/blog:aic-vs-bic>
- [9] “*Identifying the order of differencing*”, truy cập tại địa chỉ <http://www.duke.edu/~rnau/411arim2.htm>
- [10] “*Identifying the orders of AR or MA terms*”, truy cập tại địa chỉ <http://www.duke.edu/~rnau/411arim3.htm>
- [11] “*Box-Jenkins Model Identification*”, truy cập tại địa chỉ <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc446.htm>
- [12] “*Sample PACF; Durbin – Levinson algorithm*”, truy cập tại địa chỉ <http://www.pstat.ucsb.edu/faculty/feldman/174-03/lectures/113.pdf>



- [13] “*Maximum Likelihood*”, truy cập tại địa chỉ [http://en.wikipedia.org/wiki/Maximum\\_likelihood](http://en.wikipedia.org/wiki/Maximum_likelihood)
- [14] “*Maximum Likelihood Estimation*”, truy cập tại địa chỉ <http://mercury.bio.uaf.edu/courses/wlf625/readings/MLEstimation.PDF>
- [15] “*Tutorial on maximum likelihood estimation*”, truy cập tại địa chỉ <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.74.671&rep=rep1&type=pdf>
- [16] “*Maximum Likelihood Estimation*”, truy cập tại địa chỉ <http://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm>
- [17] “*Maximum Likelihood & Bayesian Parameter Estimation*”, truy cập tại địa chỉ <http://www.cedar.buffalo.edu/~srihari/CSE555/Chap3.Part1.pdf>
- [18] “*MLE (Maximum Likelihood) Parameter Estimation*” truy cập tại địa chỉ [http://www.weibull.com/AccelTestWeb/mle\\_maximum\\_likelihood\\_parameter\\_estimation.htm](http://www.weibull.com/AccelTestWeb/mle_maximum_likelihood_parameter_estimation.htm)
- [19] “*Empirical Time Series Analysis and Maximum Likelihood Estimation*”, truy cập tại địa chỉ [http://www.dcsc.tudelft.nl/Research/PubSSC/dsis/publications/deWaele\\_S/MLtimser.pdf](http://www.dcsc.tudelft.nl/Research/PubSSC/dsis/publications/deWaele_S/MLtimser.pdf)
- [20] “*What is the differences between least squares estimation and maximum likelihood estimation*”, truy cập tại địa chỉ [http://www.linkedin.com/answers/management/business-analytics/MGM\\_ANA/388159-32651588](http://www.linkedin.com/answers/management/business-analytics/MGM_ANA/388159-32651588)
- [21] “*Customizable tables of p values for teaching the t, F, and chi-square distributions*” truy cập tại địa chỉ [http://www.sociology.ohio-state.edu/people/ptv/publications/p%20values/p\\_value\\_tables.html](http://www.sociology.ohio-state.edu/people/ptv/publications/p%20values/p_value_tables.html)
- [22] “*Technical Analysis of the Financial Market*”, John J.Murphy, New York Institute of Finance, 1999
- [23] “*Technical Indicators and Overlays*”, truy cập tại địa chỉ [http://stockcharts.com/school/doku.php?id=chart\\_school:technical\\_indicators](http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators)

- [24] “*Technical Analysis*”. Truy cập tại địa chỉ <http://www.investopedia.com/university/technical/#axzz1gNU16iPn>
- [25] “*HTML 5*”, truy cập tại địa chỉ <http://en.wikipedia.org/wiki/HTML5>
- [26] “*CSS 3*”, truy cập tại địa chỉ <http://en.wikipedia.org/wiki/Css>
- [27] “*AJAX*”, truy cập tại địa chỉ [http://en.wikipedia.org/wiki/Ajax\\_\(programming\)](http://en.wikipedia.org/wiki/Ajax_(programming))
- [28] “*jQuery*”, truy cập tại địa chỉ <http://en.wikipedia.org/wiki/Jquery>
- [29] “*Kiến thức chứng khoán*”, truy cập tại địa chỉ <http://www.ssc.gov.vn/portal/page/portal/ubck/ktck>
- [30] “*Bộ môn Phân tích báo cáo tài chính*”, Phan Thị Nhi Hiếu, trường Đại học Hoa Sen.
- [31] “*Phân tích cơ bản*”, truy cập tại địa chỉ <http://www.saga.vn/Taichinh/Kithuattaichinh/Phantichcoban.saga>
- [32] “*Variance và Covariance*”, truy cập tại địa chỉ [http://en.wikipedia.org/wiki/Algorithms\\_for\\_calculating\\_variance](http://en.wikipedia.org/wiki/Algorithms_for_calculating_variance)
- [33] “*Danh mục thuật ngữ*”, truy cập tại địa chỉ <http://saga.vn/dictlist.aspx>
- [34] “*Phân tích tài chính*”, Nguyễn Minh Kiều, Chương trình giảng dạy Kinh tế Fulbright, 2007-2008