

ỨNG DỤNG NGÔN NGỮ R TRONG PHÂN TÍCH DỮ LIỆU CHỨNG KHOÁN

Đông Thị Hồng Ngọc¹, Trần Thị Ngọc Linh²,
Hoàng Thanh Hải³

Tóm tắt

Dữ liệu về thị trường chứng khoán là một trong những dạng dữ liệu chuỗi thời gian điển hình bởi chúng luôn biến động và thường không có tính quy luật. Bài viết trình bày cách sử dụng ngôn ngữ R trong việc thu thập và phân tích dữ liệu chứng khoán, từ đó đưa ra một số chỉ số kỹ thuật thường được những người hoạt động trong thị trường này sử dụng. Phương pháp nghiên cứu chính trong bài viết là tổng hợp cơ sở lý thuyết dựa trên các tài liệu thứ cấp, sử dụng bộ dữ liệu từ các nguồn có sẵn và thực hiện phân tích số liệu thống kê bằng R. Kết quả thu được của bài viết giúp các nhà đầu tư chứng khoán có thêm một phương pháp phân tích dữ liệu một cách trực quan, các chỉ số kỹ thuật được tính toán hoàn toàn tự động, từ đó giúp cho việc phân tích dữ liệu cô phiếu theo thời gian thực đạt được hiệu quả tốt nhất.

Từ khoá: Thị trường chứng khoán; giá trị cô phiếu; ngôn ngữ R; quantmod; phân tích dữ liệu.

APPLICATION OF R LANGUAGE IN STOCK DATA ANALYSIS

Abstract

Stock market data is one of the typical types of time series data because it is volatile and often irregular. The article presents how to use R language in collecting and analyzing stock data, thereby providing a number of technical indicators that are commonly used by people operating in this market. The main research methods in the article include synthesizing theories based on secondary documents, using data sets from available sources and performing statistical analysis using R. The results of the article help stock investors have an additional method of data analysis in an intuitive way; technical indicators are calculated automatically, thereby helping to analyze stock data in the real time with the best effect.

Keywords: Stock market; stock value; R language; quantmod; data analysis.

JEL classification: C81; C88.

1. Đặt vấn đề

Đối với các nhà quản lý kinh tế, thông tin về thị trường chứng khoán đóng một vai trò rất quan trọng. Thông tin này cho phép đánh giá “*tình trạng ổn định hay bất ổn*” của một nền kinh tế để từ đó các nhà kinh tế có thể đưa ra biện pháp quản lý kinh tế vĩ mô một cách hiệu quả và phù hợp (Philip Arestis, Panicos O. Demetriades, Kul B. Luintel, 2001; Nguyễn Thị Minh Huệ, Lê Thị Hương Lan, 2019). Đối với những nhà đầu tư, những cá nhân tham gia thị trường chứng khoán, đây là một thước đo quan trọng trong việc thu thập thông tin về giá trị doanh nghiệp, xu hướng của thị trường để từ đó đưa ra sự lựa chọn hoặc quyết định đầu tư một cách đúng đắn. Việc phân tích một cách nhanh chóng và chính xác các thông tin này đóng vai trò then chốt trong việc hoạch định chính sách kinh tế vĩ mô của các nhà quản lý và quyết định góp vốn của các nhà đầu tư (Nguyễn Thị Minh Huệ, Trần Đăng Khâm, 2016).

Alexandru & cs (2014) đã trình bày những khó khăn trong việc phân tích dữ liệu định lượng dành cho thị trường chứng khoán sử dụng các phần mềm thống kê thông thường, người phân tích sử dụng các phần mềm này gặp phải nhiều khó khăn trong việc thu thập dữ liệu theo thời gian thực, đa số là các phần mềm phải trả phí và bị phụ thuộc rất nhiều vào các hàm tính toán thống kê được lập trình sẵn, các công cụ trực quan dữ liệu cũng rất hạn chế. Nhóm nghiên cứu đưa ra phương pháp tính toán hiệu quả hơn bằng ngôn ngữ R

hoàn toàn miễn phí và gói quantmod khi phân tích dữ liệu chứng khoán.

Trong phân tích dữ liệu định lượng, tính ưu việt khi sử dụng gói quantmod được thể hiện rõ rệt, đặc biệt là với các dạng dữ liệu mang tính chuỗi thời gian và liên quan đến thị trường kinh tế, tài chính (Nicolae-Marius Jula, Nicoleta Jula, 2016; Mahantesh C. Angadi, Amogh P. Kulkarni, 2015; (F. R. Johnston, J. E. Boyland, M. Meadows and E. Shale, 1999). Angadi & cs đã chỉ ra sự chính xác của việc áp dụng phần mềm R và gói quantmod kết hợp với học sâu trong việc phân tích dữ liệu và xây dựng mô hình dự đoán ARIMA dành cho dữ liệu thị trường tài chính (Mahantesh C. Angadi, Amogh P. Kulkarni, 2015).

Một số nghiên cứu gần đây đều cho thấy các nhà nghiên cứu đã tiến tới sử dụng phương pháp phân tích định lượng bằng ngôn ngữ R và gói lệnh quantmod như một công cụ phân tích phổ biến và hiệu quả về thị trường chứng khoán: Akuegbu đánh giá ảnh hưởng của đại dịch COVID-19 tới thị trường chứng khoán của 4 quốc gia bao gồm: Vương quốc Anh, Tây Ban Nha, Pháp và Đức (Akuegbu, 2020). Trong nghiên cứu đưa ra, He đã lựa chọn sử dụng ứng dụng R và các kỹ thuật để phân tích tính thời vụ của thị trường chứng khoán Hoa Kỳ từ tháng 01 năm 1941 tới tháng 12 năm 2020 thông qua gói quantmod (He, 2022).

Nghiên cứu của Schuenemann & cs đã sử dụng phần mềm R đánh giá tác động của toàn cầu hóa lên thị trường chứng khoán thông qua việc sử

dụng phương pháp phân tích định lượng về lợi nhuận cổ phiếu hàng từ dữ liệu các doanh nghiệp của Nhật Bản và Trung Quốc trong khoảng thời gian từ năm 2009 đến năm 2015. Kết quả phân tích bằng R cho thấy sự tiện lợi và khả năng linh hoạt trong thu thập dữ liệu từ các nguồn khác nhau, khả năng trực quan và đưa ra các tính toán thống kê một cách nhanh chóng, đáng tin cậy (Jan-Hendrik Schuenemanna, Natalia Ribberinka, Natallia Katenka, 2019).

Ngôn ngữ R ban đầu được xây dựng như một phần mềm mã nguồn mở dùng cho phân tích thống kê và vẽ biểu đồ. Phần mềm này được khởi xướng bởi R. Ihaka và R. Gentleman vào năm 1966 (Ross Ihaka, Robert Gentleman, 1996). Ý tưởng này của R được giới thống kê học ủng hộ và tập trung phát triển thành một ngôn ngữ máy tính đa năng và hoàn toàn miễn phí.

Nghiên cứu của nhóm tác giả Lê Văn Tuấn và Phùng Duy Quang (2014) đã sử dụng ngôn ngữ R trong việc phân tích dữ liệu cổ phiếu Việt Nam thông qua trang *cophieu68.com*. Trong nghiên cứu, các tác giả đã đưa ra các tham số thống kê phù hợp với dạng dữ liệu loại này, đồng thời, giới thiệu một số hàm tính toán xây dựng sẵn trong ngôn ngữ R để đưa ra kết quả cho các chỉ số kỹ thuật. Tuy nhiên, nghiên cứu chưa đưa ra được cách lấy dữ liệu trực tiếp và khai thác dữ liệu theo thời gian thực mà chỉ dừng lại ở việc phân tích và xử lý dữ liệu với một file dữ liệu dạng .csv có sẵn.

Nhiên cứu của tác giả Hoàng Nhật Quy (2019) đã đưa ra cái nhìn tổng quan hơn và giới thiệu gói *quantmod* để có thể lấy dữ liệu từ Internet theo thời gian thực.



Hình 1. Dữ liệu cổ phiếu của công ty Apple (AAPL) trong thời gian 03/2007 - 02/2022 thu thập bằng gói *quantmod*.

Trong R có một số thư viện có thể dùng để khai thác dữ liệu tài chính như: *tidyquant*, *tseries*, *quantmod*. Các thư viện này đều có một lượng đa dạng các hàm khai thác dữ liệu và mỗi một thư viện đều hỗ trợ những kiểu dữ liệu khác nhau (J. Ryan, J. Ulrich, W. Thielen, P. Teetor, S. Bronder, 2020), (Ross Ihaka, Robert Gentleman, 1996). Một chức năng rất mạnh của R là cho phép lấy trực tiếp các dữ liệu trên Internet, nhất là dữ liệu tài chính theo thời gian (FTS - Financial Time

Series) bằng thư viện *quantmod*. Với thư viện *quantmod* trong R cho phép tải trực tiếp dữ liệu khoảng 40000 chuỗi dữ liệu thời gian từ các nguồn như Yahoo Finance, Google Finance, ...) (Mahantesh C. Angadi, Amogh P. Kulkarni, 2015: J. Ryan, J. Ulrich, W. Thielen, P. Teetor, S. Bronder, 2020).

Vấn đề đặt ra đối với các nhà nghiên cứu mới tiếp cận phân tích thị trường chứng khoán đó là phần mềm R và gói *quantmod* sử dụng như thế nào?

Tại Việt Nam, những năm gần đây ngôn ngữ R bắt đầu được sử dụng trong giảng dạy tại một số trường Đại học lớn như Trường Đại học Kinh tế Quốc dân, Trường Đại học Thương mại, Trường Đại học Đà Nẵng,... Mục tiêu của bài viết này là giới thiệu cho người phân tích dữ liệu các chức năng của gói *quantmod*, phục vụ cho phân tích kỹ thuật về bất kỳ chuỗi thời gian tài chính như cổ phiếu, cặp ngoại hối, trái phiếu, hàng hóa, v.v.). *Quantmod* là gói phân tích dữ liệu chuỗi thời gian chuyên biệt được tạo ra để phân tích định lượng bằng ngôn ngữ R.

2. Phương pháp nghiên cứu

Bài viết sử dụng phương pháp nghiên cứu tài liệu thứ cấp, phương pháp nghiên cứu phân tích định lượng để đưa ra được các tính chất cơ bản của dạng dữ liệu cổ phiếu chứng khoán và các chỉ số kỹ thuật liên quan tới kiểu dữ liệu này. Trong bài viết có sử dụng dữ liệu cổ phiếu theo chuỗi thời gian của công ty Apple với mã cổ phiếu được niêm yết trên thị trường chứng khoán là AAPL. Dữ liệu được thu thập trực tiếp và theo thời gian thực thông qua nguồn của trang Yahoo Finance (<https://finance.yahoo.com/>). Đây là nguồn lấy dữ liệu chứng khoán và các thông tin khác về kinh tế khá uy tín, được cập nhật thường xuyên theo thời gian thực đối với tất cả các mã cổ phiếu đã được niêm yết. Sau khi dữ liệu đã được thu thập, nhóm tác giả sử dụng gói *quantmod* và các hàm tính toán, trực quan dữ liệu được tích hợp trong R để tiến hành xử lý và phân tích dữ liệu.

Gói *quantmod* trong R không những có thể trực quan dữ liệu cổ phiếu của các công ty nước ngoài mà còn có thể lấy trực quan cho dữ liệu của bất kỳ công ty nào muốn phân tích, với điều kiện công ty đó có mã niêm yết được đăng trên các nguồn trang thông tin dữ liệu cổ phiếu. Chẳng hạn, mã cổ phiếu của Tập đoàn Vingroup được niêm yết là VIC và được đăng thông tin trên trang *cophieu68.vn* (phụ lục).

3. Kết quả nghiên cứu

3.1. Trực quan dữ liệu

3.1.1. Thu thập dữ liệu

Mã cổ phiếu của công ty Apple có thể nhận được thông qua Internet hoặc qua nguồn Yahoo Finance. Để thu thập dữ liệu về biến động cổ phiếu

của công ty Apple chúng ta tiến hành thực hiện các lệnh trong R như sau:

```
install.packages("quantmod")- Cài đặt thư viện quantmod;
library(quantmod) - Tải thư viện quantmod vào môi trường R;
Sys.Date() - Hiển thị thời gian hiện tại;
getSymbols(Symbols = "AAPL", src = "yahoo") - Lấy dữ liệu cổ phiếu của APPLE thông qua mã AAPL và nguồn dữ liệu từ Yahoo Finance;
```

```
AAPL <- data.frame(AAPL) - Chuyển đổi thành dữ liệu cấu trúc dataframe;
fix(AAPL) - Hiển thị dữ liệu thu thập được;
str(AAPL) - Xem cấu trúc dữ liệu của các biến trong dữ liệu AAPL;
head(AAPL) - Hiển thị 6 dữ liệu đầu tiên của tập dữ liệu;
tail(AAPL) - Hiển thị 6 dữ liệu gần nhất của tập dữ liệu.
```

	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2007-01-03	3.081786	3.092143	2.925000	2.992857	1238319600	2.562706
2007-01-04	3.001786	3.069643	2.993571	3.059286	847260400	2.619588
2007-01-05	3.063214	3.078571	3.014286	3.037500	834741600	2.600933
2007-01-08	3.070000	3.090357	3.045714	3.052500	797106800	2.613777
2007-01-09	3.087500	3.320714	3.041071	3.306071	3349298400	2.830903
2007-01-10	3.383929	3.492857	3.337500	3.464286	2952880000	2.966379

Hình 2. Hiển thị 6 dữ liệu đầu tiên của cổ phiếu APPLE trong thời gian từ 03/2007 - 02/2022.

Nguồn: Nhóm tác giả phân tích

2022-02-03	174.48	176.24	172.12	172.90	89418100	172.6800
2022-02-04	171.68	174.10	170.68	172.39	82391400	172.3900
2022-02-07	172.86	173.95	170.95	171.66	77251200	171.6600
2022-02-08	171.73	175.35	171.43	174.83	74829200	174.8300
2022-02-09	176.05	176.65	174.90	176.28	71285000	176.2800
2022-02-10	174.14	175.48	171.55	172.12	90558600	172.1200

Hình 3. Hiển thị 6 dữ liệu gần đây nhất của cổ phiếu APPLE tại thời điểm truy cập ngày 10/02/2022

Nguồn: Nhóm tác giả phân tích

Quan sát các biến số của tập dữ liệu cổ phiếu AAPL, dữ liệu này là một tập dữ liệu chuỗi thời gian, gồm 6 biến số mô tả:

- AAPL.Open - Giá mở cửa của phiên giao dịch.
- AAPL.High - Là giá cao nhất trong một phiên giao dịch hoặc trong một chu kỳ theo dõi biến động giá.
- AAPL.Low - Là giá thấp nhất trong một phiên giao dịch hoặc trong một chu kỳ theo dõi biến động giá.
- AAPL.Close - Giá đóng cửa là giá thực hiện tại lần khớp lệnh cuối cùng trong ngày giao dịch.
- AAPL.Volume - Khối lượng giao dịch.
- AAPL.Adjusted - Giá đóng cửa hiệu chỉnh.

Ngoài ra chúng ta có thể chi tiết hóa dữ liệu thu thập được theo chu kỳ, theo khoảng thời gian, hoặc phân tích một biến số như chi phân tích giá đóng cửa, giá mở cửa, ... hoặc lấy nhiều giá trị cổ phiếu của nhiều công ty cùng một lúc bằng các lệnh như sau:

```
Open <- Op(AAPL); Close <- Cl(AAPL);
High <- Hi(AAPL); Low <- Lo(AAPL); AdjClose <- Ad(AAPL); Volume <- Vo(AAPL).
```

Những nội dung này cho phép người sử dụng lấy giá trị của cổ phiếu theo từng biến riêng biệt như giá mở cửa; giá đóng cửa; giá trị cao nhất cổ phiếu; giá thấp nhất; giá đóng cửa hiệu chỉnh; khối lượng cổ phiếu giao dịch và gán cho từng đối

tượng có tên lần lượt là Open; Close; High; Low; AdjClose; Volume.

```
getSymbols("AAPL", from='year-month-day',to='year-month-day') - Nội dung này cho phép phân tích giá trị cổ phiếu trong một thời gian nhất định;
```

```
AAPL <- last(AAPL,'1 year') - Lấy giá trị cổ phiếu trong năm gần đây nhất;
```

```
AAPL <- first(AAPL,'3 years') - Lấy giá trị cổ phiếu trong ba năm đầu tiên;
```

```
getSymbols(c("AAPL","GOOG")) - Lấy đồng thời giá trị cổ phiếu của công ty APPLE và công ty GOOGLE;
```

```
stocklist <- c("AAPL","GOOG") - Lấy đồng thời giá trị cổ phiếu của công ty APPLE và công ty GOOGLE và gán cho một đối tượng có tên là stocklist;
getSymbols(stocklist) - Lấy dữ liệu cổ phiếu stocklist;
```

```
WeekVoYa <- apply.weekly(Vo(AAPL),sum)
- Phân tích dữ liệu cổ phiếu biến động theo tuần;
MonthVoYa <- apply.monthly(Vo(AAPL),sum)
- Phân tích dữ liệu cổ phiếu biến động theo tháng;
QuarterVoYa <- apply.quarterly(Vo(AAPL),sum)
```

```
YearVoYa <- apply.yearly(Vo(AAPL),sum)
- Phân tích dữ liệu cổ phiếu biến động theo năm;
```

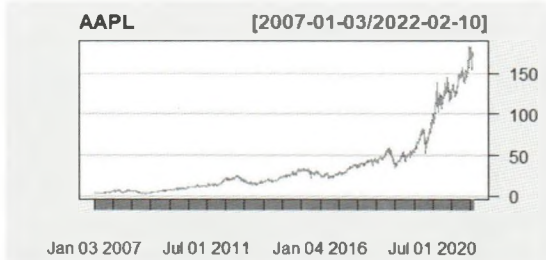
```
WeekAveVoCIYa <- apply.weekly(Vo(AAPL),mean)
- Phân tích dữ liệu cổ phiếu biến động theo giá trị trung bình theo tuần.
```

3.1.2. Phân tích trực quan dữ liệu

Thư viện `quantmod` hỗ trợ trực quan dữ liệu rất hiệu quả với dữ liệu ở dạng chuỗi thời gian. Biểu đồ đường là loại biểu đồ phổ biến trong các ngành khoa học vì tính trực quan và dễ hiểu. Đây là loại biểu đồ cơ bản nhất được sử dụng trong tài chính. Đối với chứng khoán, biểu đồ đường mô tả giá đóng cửa của chứng khoán theo thời gian, có thể được sử dụng cho bất kỳ chu kỳ thời gian nào (theo ngày, theo tuần, theo tháng,...) nhưng chúng thường được sử dụng để mô tả sự thay đổi về giá theo ngày. Biểu đồ đường phổ biến với các nhà đầu tư vì giá đóng cửa là một phần dữ liệu phổ biến liên quan đến chứng khoán.

Giá sử chúng ta muốn trực quan dữ liệu của cổ phiếu công ty APPLE để xem biến động của giá cổ phiếu theo thời gian với giá mở cửa của giá cổ phiếu AAPL từ ngày 03/01/2007 đến ngày 20/10/2021 và sử dụng biểu đồ đường thông qua câu lệnh:

```
chartSeries(AAPL, theme = "white")
```



Hình 4. Giá cổ phiếu APPLE mô tả bằng biểu đồ đường trong thời gian từ 03/01/2007 - 20/10/2021.

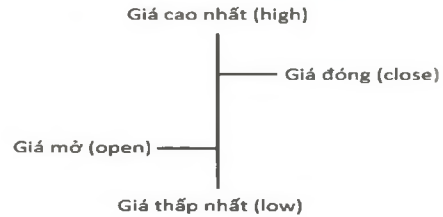
Nguồn: Nhóm tác giả phân tích

Trong biểu đồ được mô tả trong Hình 4, trục x biểu diễn chiều thời gian, trục y biểu diễn khoảng giá của cổ phiếu Apple, tham số `theme` để sử dụng cho nền biểu đồ là màu trắng. Ưu điểm của dạng biểu đồ này chính là sự đơn giản, các nhà phân tích có thể dễ dàng nhận biết được xu hướng của giá ngay khi nhìn vào biểu đồ. Biểu đồ giá cổ phiếu APPLE trên cho thấy được xu hướng tăng mạnh của cổ phiếu này bắt đầu từ nửa cuối tháng 2/2019 và trước đó là giai đoạn giá đi ngang.

Tuy nhiên, loại biểu đồ này không thể hiện được mức độ biến động của giá trong một phiên giao dịch, bởi vì biểu đồ đường thường chỉ hiển thị giá đóng cửa, chúng giảm nhiều từ những thời điểm ít quan trọng hơn trong ngày giao dịch, chẳng hạn như giá mở cửa, giá cao và giá thấp. Do đó trong điều kiện thị trường phức tạp, dao động giá trong phiên giao dịch cao điểm thì sử dụng biểu đồ đường không mang lại hiệu quả phân tích, dẫn đến không có nhiều cơ sở quan trọng cho việc ra quyết định đầu tư.

Biểu đồ dạng thanh (Bar chart): Dạng biểu đồ này cung cấp thêm thông tin về biến động của giá chứng khoán trong một phiên giao dịch. Mỗi

thanh trên biểu đồ thể hiện thông tin của chứng khoán trong phiên giao dịch (hoặc trong một khoảng thời gian) như sau: thanh dọc thể hiện giá mức giá thấp nhất và giá cao nhất tại thời gian giao dịch tương ứng, thanh ngang trái cho biết giá mở cửa, thanh ngang phải cho biết giá đóng cửa (Omer Berat Sezer, Murat Ozbayoglu, 2019).



Hình 5. Cấu trúc mô tả dữ liệu về giá cổ phiếu trên biểu đồ thanh.

Nguồn: <https://knowledge.sapp.edu.vn/>

Để vẽ được biểu đồ thanh trong R, sử dụng câu lệnh sau:

```
chartSeries(AAPL, type = "bar", theme = chartTheme('white'))
```



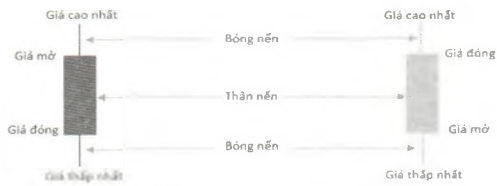
Hình 6. Giá cổ phiếu của công ty APPLE mô tả bằng biểu đồ thanh trong thời gian từ 05/2021 - 08/2021.

Nguồn: Nhóm tác giả phân tích

Biểu đồ thanh tại Hình 6 cung cấp các thông tin liên quan đến giá trị cổ phiếu APPLE trong giai đoạn từ tháng 05/2021 đến tháng 08/2021, phần dưới cùng của biểu đồ thể hiện khối lượng giao dịch chứng khoán Volume. Các thanh được thể hiện bằng màu sắc khác nhau thể hiện giá tăng hay giảm vào thời gian giao dịch nào đó. Chẳng hạn, màu vàng biểu thị thanh giảm so với thanh trước, trong khi màu xanh cho biết sự tăng so với thanh trước đó.

Tuy nhiên, biểu đồ thanh không đơn giản để có thể đọc hiểu, đặc biệt là đối với các nhà đầu tư mới tiếp cận dạng biểu đồ này.

Biểu đồ nến (Candlestick Chart) được người Nhật Bản sáng tạo và áp dụng đầu tiên trên thị trường chứng khoán của họ. Với nhiều ưu điểm, loại biểu đồ này đang dần trở thành loại được sử dụng phổ biến nhất trên hầu hết các thị trường chứng khoán hiện đại (Tsung-Hsun Lu, Yung-Ming Shiu, Tsung-Chi Liu, 2012). Biểu đồ nến tạo ra bởi các chuyển động lên và xuống của giá cổ phiếu. Các nhà giao dịch có thể dựa vào biểu đồ nến để đưa ra quyết định giao dịch dựa trên tần suất xuất hiện của các hình, từ đó giúp dự báo hướng ngắn hạn của giá cổ phiếu đó.



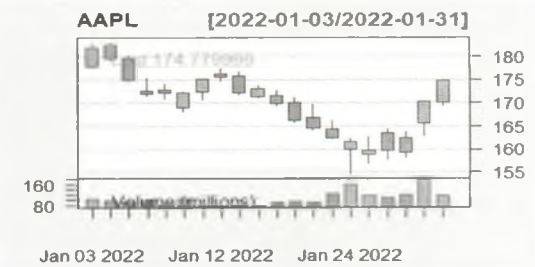
Hình 7. Cấu trúc mô tả dữ liệu về giá cổ phiếu trên biểu đồ nến.

Nguồn: <https://knowledge.sapp.edu.vn/>

Trên biểu đồ nến cũng chứa đựng bốn thông tin như biểu đồ thanh trong suốt khoảng thời gian mà nhà giao dịch lựa chọn khảo sát. Thân nến là một phần rộng, đại diện cho phạm vi giá giữa mở và đóng của giao dịch ngày hôm đó. Khi phần thân được tô màu hoặc có màu đen, nghĩa là giá đóng cửa thấp hơn giá mở cửa. Nếu phần thân trống, điều đó có nghĩa là giá đóng cửa cao hơn giá mở cửa. Bóng nến cho thấy giá cao và thấp của giao dịch ngày hôm đó. Nếu bóng trên của nến giảm ngắn, điều đó cho thấy mức mở cửa của ngày hôm đó gần với mức cao nhất trong ngày.

Để vẽ được biểu đồ nến trong R sử dụng thư viện quantmod, chúng ta thực hiện câu lệnh sau:

```
chartSeries(AAPL,type="candlesticks",
subset='2022-01',theme=chartTheme('white'))
```



Nguồn: Nhóm tác giả phân tích

Hình 8. Giá cổ phiếu APPLE trong giai đoạn tháng 01/2022 mô tả bằng biểu đồ nến.

So với biểu đồ thanh, khi nhìn vào biểu đồ nến, sự biến động của giá được thể hiện rõ ràng và dễ nhận biết hơn. Do đó, các nhà đầu tư dễ dàng nhận thấy sự khác biệt và mối quan hệ giữa các mức giá đóng, giá mở, giá cao nhất và giá thấp nhất bởi chúng được minh họa rõ ràng và trực quan hơn bằng cách mã hóa màu sắc của nến, độ rộng của thân, giúp làm nổi bật sự khác biệt của giá giữa thời điểm mở cửa và đóng cửa.

3.2. Một số chỉ số kỹ thuật thống kê

3.2.1. SMA – Simple Moving Average

Đường SMA thường được gọi là đường trung bình động đơn giản – công cụ chỉ báo phản ánh biến động giá cổ phiếu. Đây được xem là công cụ cơ bản nhất dành cho các nhà đầu tư nếu muốn sử dụng để phân tích kỹ thuật. Bản chất của đường SMA là số trung bình cộng của giá, thường là giá đóng cửa thị trường, trong khoảng thời gian nhất định sau khi đã loại bỏ các yếu tố bất thường (F. R. Johnston, J. E. Boyland, M. Meadows and E.

Shale, 1999). Do đó, đường SMA được đánh giá mang lại hiệu quả cao cho các nhà đầu tư khi có nhu cầu xác định sự biến động về giá. Đường trung bình động thường được đặt tên cho khoảng thời gian sử dụng cho tính toán, chẳng hạn như, SMA 10 chính là đường trung bình động tính theo giá đóng cửa của 10 ngày trước đó và các đường trung bình động khác cũng có ý nghĩa tương tự.

Các nhà đầu tư sẽ phân loại các đường SMA theo thời gian như sau:

- Đường trung bình động phù hợp các nhà đầu tư yêu thích phân tích biến động ngắn hạn: SMA 10, SMA 14, SMA 20.
- Đường trung bình động phù hợp các nhà đầu tư yêu thích phân tích biến động trung hạn là SMA 50.
- Đường trung bình động phù hợp các nhà đầu tư yêu thích phân tích biến động dài hạn là SMA 100, SMA 200.

Cách tính đường SMA

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n} \quad (1)$$

trong đó: $A_i (i = 1, 2, \dots, n)$: là mức giá đóng cửa tại giai đoạn i ;

n : tổng số giai đoạn xét sự biến động.

Để tính toán giá trị SMA trong R, sử dụng cú pháp sau:

- sma <- SMA(Close,n=20) - Tính toán giá trị SMA 20 với giá trị cổ phiếu đóng cửa;
- tail(sma,n=5) - Thống kê ra 5 giá trị gần đây nhất của SMA 20.

```
> tail(sma,n=5)
SMA
2021-10-14 143.1860
2021-10-15 143.1250
2021-10-18 143.3055
2021-10-19 143.5720
2021-10-20 143.7425
```

Hình 9. Liệt kê 5 giá trị gần đây nhất ứng với SMA 20 tính tại thời điểm truy cập là ngày 20/10/2021.

Nguồn: Nhóm tác giả phân tích

Thư viện quantmod trong R cho phép trực quan dữ liệu ứng với các đường SMA khác nhau. Hình 10 biểu diễn giá trị cổ phiếu APPLE trong giai đoạn từ tháng 05/2021 đến tháng 08/2021 có bổ sung thêm 2 đường SMA 30 màu xanh da nước biển và SMA 200 màu đỏ tương ứng. Vẽ các đường SMA trong biểu đồ ta thực hiện câu lệnh:

```
chartSeries(AAPL,subset='2021-05::2021-08',
theme=chartTheme('white'))
addSMA(n=30,on=1,col="blue")
addSMA(n=200,on=1,col="red")
```

Theo biểu đồ trong Hình 10, ta có thể thấy, SMA 200 khá “mịn”, có khoảng cách với đường giá và đường như không ảnh hưởng đến biến động

của giá. Trong khi đó, SMA 30 gần giá hơn SMA 200, có chuyển động theo từng nhịp biến động của giá. Cả 2 đường SMA đều đi lên trong các thời kì cuối biểu thị xu hướng đi lên. Theo lý thuyết, giá giao với các đường SMA được sử dụng để kích hoạt các tín hiệu giao dịch. Khi giá vượt qua đường SMA, có thể muốn mua hoặc bán; khi chúng cắt xuống dưới đường SMA, có thể muốn mua ngắn hạn hoặc thoát lệnh mua.



Hình 10. Đường SMA 30 và SMA 200.

Nguồn: Nhóm tác giả phân tích

Ngoài ra, SMA Giao nhau SMA là một tín hiệu giao dịch phổ biến khác. Khi SMA trong thời gian ngắn vượt lên trên SMA dài hạn, có thể muốn mua dài hạn. Còn nếu các nhà đầu tư muốn mua ngắn hạn thì quan sát thời điểm khi đường SMA ngắn hạn cắt xuống dưới đường SMA dài hạn. Trong trường hợp hình 10, không có sự giao nhau giữa hai đường SMA.

3.2.2. Bollinger Band

Bollinger band là một công cụ phân tích kỹ thuật xác định bởi đường trung bình đơn giản (Simple Moving Average - SMA) ở giữa, dải trên và dải dưới. Dải Bollinger bands sẽ tự điều chỉnh mở rộng trong các giai đoạn thị trường biến động và thu hẹp trong các giai đoạn thị trường ít biến động hơn. Bollinger bands được phát triển và sở hữu bản quyền bởi một nhà phân tích kỹ thuật nổi tiếng tên là John Bollinger (Camillo Lento, Nikola Gradojevic, Chris Wright, 2007).

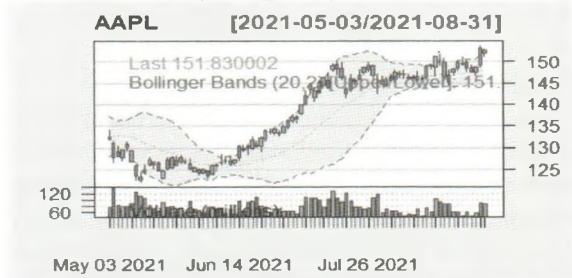
Bollinger Bands là một chỉ báo phổ biến trong thị trường chứng khoán. Giá cổ phiếu càng di chuyển đến dải trên của dải Bollinger Bands, thị trường càng mua quá nhiều và giá càng di chuyển đến dải thấp hơn, thị trường càng bán quá mức.

Dải Bollinger Bands siết chặt (thu hẹp): Siết chặt là khái niệm quan trọng của Bollinger Bands. Dải Bollinger siết chặt khi khoảng cách giữa dải trên và dải dưới với đường SMA được thu hẹp. Bollinger Bands siết chặt cho biết cổ phiếu đang trong giai đoạn biến động thấp. Các nhà đầu tư cho rằng đây là một dấu hiệu cho biết giá sẽ di biến động mạnh trong tương lai và có thể xuất hiện thêm nhiều cơ hội giao dịch. Ngược lại, các dải di chuyển rộng ra, có khả năng biến động sẽ giảm. Tuy nhiên, những diễn biến này không phải là tín hiệu giao dịch, vì nó không cho biết giá sẽ biến động theo hướng tăng hay giảm.

Bứt phá: Khoảng 90% biến động giá xảy ra giữa dải trên và dải dưới. Bất kỳ khi nào giá vượt qua dải trên hoặc dải dưới đều là sự kiện lớn. Tương tự như khi siết chặt, bứt phá không phải là một tín hiệu giao dịch. Sai lầm mà hầu hết mọi người mắc phải là tin rằng giá chạm hoặc vượt một trong các dải là tín hiệu để mua hoặc để bán. Sự bứt phá không cung cấp các manh mối về hướng và mức độ của sự di chuyển giá cổ phiếu trong tương lai.

Tính toán và trực quan dải Bollinger Bands với quantmod có thể thực hiện như sau:

```
chartSeries(AAPL,subset='2021-05::2021-08', theme=chartTheme('white'))
addBBands(n=20.sd=2)
```



Hình 11. Giá cổ phiếu trong thời gian từ 05/2021 - 08/2021 được mô tả qua dải Bollinger Bands.

Nguồn: Nhóm tác giả phân tích

4. Kết luận

Mục đích của bài báo nhằm giới thiệu về ngôn ngữ R dùng trong phân tích dữ liệu chứng khoán. Các dữ liệu chứng khoán luôn biến đổi và có nhiều thông tin cần khai thác, dự đoán. Ngôn ngữ R và thư viện quantmod được sử dụng như một công cụ trợ giúp phân tích thống kê và trực quan dữ liệu hiệu quả. Từ đó đưa ra các nhận xét, dự đoán về giá trị tương lai của mỗi loại cổ phiếu khác nhau.

Phần 3 của bài viết đã trình bày về cách trực quan dữ liệu cổ phiếu và phân tích một số chỉ số kỹ thuật chứng khoán cơ bản, áp dụng vào cổ phiếu của công ty APPLE. Tuy nhiên, việc đánh giá thông qua các chỉ số và biểu đồ thống kê như vậy chưa đủ để mang lại cái nhìn toàn diện và phân tích một cách hiệu quả giá trị của các cổ phiếu trên thị trường chứng khoán hiện nay, do đó nhóm tác giả đề xuất một số các phương pháp và hướng nghiên cứu mới như sau:

- Phân tích thêm một số các chỉ số kỹ thuật chuyên sâu về cổ phiếu như Momentum, chỉ báo ROC, chỉ báo MACD nhằm góp phần đánh giá hiệu quả hơn nữa các phép phân tích thống kê giá trị cổ phiếu.

- Trực quan dữ liệu động theo chuỗi thời gian kết hợp với phần mềm trực quan Power BI.

- Tìm hiểu thêm các thuật toán về tính toán hệ số rủi ro của giá cổ phiếu trên thị trường chứng khoán.

TÀI LIỆU THAM KHẢO

- [1]. Akuegbu, C. (2020). *An investigative analysis on the effect of a pandemic on countries' stock market indexes within Europe: A case study of (covid 19) coronavirus*. National College of Ireland.
- [2]. Camillo Lento, Nikola Gradojevic, Chris Wright. (2007). Investment information content in Bollinger Bands? *Applied Financial Economics Letters*, 3(4), 263-267.
- [3]. Ciprian Alexandru, Nicoleta Caragea, Ana - Maria Dobre. (2014). R-evolution in Time Series Analysis Software Applied on R-omanian Capital Market. *Computational Methods in Social Sciences*, 2(1), 28-34.
- [4]. F. R. Johnston, J. E. Boyland, M. Meadows and E. Shale. (1999). Some Properties of a Simple Moving Average when Applied to Forecasting a Time Series. *The Journal of the Operational Research Society*, 50(12), 1267-1271.
- [5]. He, X. J. (2022). Seasonality Analysis of US Stock Market. In *Advances in Information and Communication* (pp. 263-270). Springer, Cham.
- [6]. Hoàng Nhật Quy, Huỳnh Thị Oanh Triều, Trương Thị Minh Hiếu. (2019). Ứng dụng phần mềm R phân tích số liệu trực tuyến thị trường chứng khoán trong nước và quốc tế. *Hội thảo khoa học Quốc gia CITA lần thứ 8*, (pp. 304 - 310).
- [7]. J. Ryan, J. Ulrich, W. Thielen, P. Teetor, S. Bronder. (2020). *cran.r-project.org*. Retrieved 10 25, 2021, from <https://cran.r-project.org/web/packages/quantmod/quantmod.pdf>
- [8]. Jan-Hendrik Schuenemanna, Natalia Ribberinka, Natallia Katenka. (2019). Japanese and Chinese Stock Market Behaviour in Comparison can analysis of dynamic networks. *Asia Pacific Management Review*, 25(2), 99-110.
- [9]. Lê Văn Tuấn, Phùng Duy Quang. (2012, 4 16). Retrieved from <https://tuanvanle.wordpress.com/>: <https://tuanvanle.files.wordpress.com/2012/04/ung-dung-phan-mem-r-trong-phan-tich-co-phieu.pdf>
- [10]. Mahantesh C. Angadi, Amogh P. Kulkarni. (2015). Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R. *International Journal of Advanced Research in Computer Science*, 6(6), 104-108.
- [11]. Nicolae-Marius Jula, Nicoleta Jula. (2016). Using R for Analyzing Financial Markets. *Challenges of the Knowledge Society*, 990-995.
- [12]. Nguyễn Thị Minh Huệ, Lê Thị Hương Lan. (2019). *Giáo trình Thị trường chứng khoán*. NXB Đại học Kinh tế quốc dân.
- [13]. Nguyễn Thị Minh Huệ, Trần Đăng Khâm. (2016). *Giáo trình Phân tích đầu tư chứng khoán*. NXB Đại học Kinh tế quốc dân.
- [14]. Omer Berat Sezer, Murat Ozbayoglu. (2019). Financial Trading Model with Stock Bar Chart Image Time Series with Deep Convolutional Neural Networks. *Intelligent Automation and Soft Computing*, 26(2), 323-334.
- [15]. Philip Arestis, Panicos O. Demetriades, Kul B. Luintel. (2001). Financial Development and Economic Growth: The Role of Stock Markets. *Journal of Money, Credit and Banking*, 33(1), 16-41.
- [16]. Ross Ihaka, Robert Gentleman. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.
- [17]. Tsung-Hsun Lu, Yung-Ming Shiu, Tsung-Chi Liu. (2012). Profitable candlestick trading strategies-The evidence from a new perspective. *Review of Financial Economics*, 21(2), 63-68.

Thông tin tác giả:

1. Đồng Thị Hồng Ngọc

- Đơn vị công tác: Trường Đại học Kinh tế & QTKD
- Địa chỉ email: dongthihongngoc@tueba.edu.vn

2. Trần Thị Ngọc Linh

- Đơn vị công tác: Trường Đại học Kinh tế & QTKD

3. Hoàng Thanh Hải

- Đơn vị công tác: Trường ĐH Kinh tế và QTKD

Ngày nhận bài: 5/4/2022

Ngày nhận bản sửa: 9/5/2022

Ngày duyệt đăng: 27/5/2022