

Phân tích, dự đoán giá chứng khoán bằng mô hình thống kê

Trần Thị Mỹ Xoan - 21522815

Lê Anh Tuấn Dũng - 21521974

Lê Thị Ánh Hồng - 21520245

Nguyễn Thị Mai Liên - 21522283

Đỗ Sĩ Đạt - 21521932

Ngày 9 tháng 6 năm 2024

Tóm tắt nội dung: Khi thực hiện đầu tư vào các cổ phiếu chứng khoán, việc sử dụng các mô hình dự đoán giá cổ phiếu trước khi thực hiện một giao dịch đầu tư có thể giúp cho ta có các quyết định đúng đắn hơn khi ra quyết định thực hiện 1 giao dịch. Bài báo này trình bày phương pháp dự đoán giá cổ phiếu của ba công ty lớn Samsung, LG và Sony. Chúng tôi sẽ thực hiện dự đoán bằng cách sử dụng các thuật toán thống kê: ARIMA, Linear Regression, VARMA. Bộ dữ liệu được sử dụng lấy từ ngày 1/3/2019 đến ngày 1/6/2024. Sau khi thực hiện dự đoán thì ta sẽ thực hiện sử dụng các độ đo MAPE, MSE, RMSE để đánh giá các mô hình. Cuối cùng, sử dụng các mô hình để thực hiện dự đoán giá chứng khoán cho 30, 60 và 90 ngày tiếp theo.

Từ khoá: ARIMA, Linear Regression, VARMA. Phân tích, dự đoán giá chứng khoán.

1 GIỚI THIỆU

Dự đoán giá cổ phiếu là một vấn đề quan trọng trong lĩnh vực đầu tư tài chính. Việc dự đoán chính xác giá cổ phiếu có thể giúp các nhà đầu tư đưa ra quyết định đầu tư sáng suốt và giảm thiểu rủi ro. Có nhiều phương pháp khác nhau để dự đoán giá cổ phiếu, bao gồm phân tích kỹ thuật, phân tích cơ bản và học máy. Samsung, LG, SONY là những công ty hàng đầu trong ngành công nghệ, giá cổ phiếu của 3 công ty đóng vai trò là chỉ số quan trọng về động lực thị trường, khiến việc phân tích của họ trở nên cần thiết đối với các nhà đầu tư.

Báo cáo này nhằm mục đích phân tích, dự đoán giá cổ phiếu của các công ty có ảnh hưởng bằng cách sử dụng 3 thuật toán ARIMA, Linear Regression, VARMA để đưa ra dự đoán cho giá cổ phiếu. Nhờ đó các nhà đầu tư có thể lựa chọn chính xác hạng mục đầu tư vào giá cổ phiếu.

Bằng cách tận dụng những phương pháp này, chúng tôi mong muốn cung cấp hướng dẫn có giá trị cho các nhà đầu tư trong việc định hướng bối cảnh năng động của lĩnh vực công nghệ.

2 NGHIÊN CỨU LIÊN QUAN

[1] Mô hình VARMA thường được sử dụng để dự báo dữ liệu chuỗi thời gian đa biến. Mô hình VARMA là sự mở rộng của mô hình ARMA trong chuỗi thời gian đơn biến (Lütkepohl, 2005; Wei, 1990) và được sử dụng với điều kiện rằng dữ liệu phải ổn định theo thời gian (Lütkepohl, 2005). Mô hình VARMA (p,q) là sự kết hợp của mô hình VAR (p) và mô hình vector trung bình động (q) (VMA (q)). Bài báo này xác định 4 mô hình VARMA (p,q) lần lượt là (1,1), (2, 1), (3, 1), (4, 1), Việc lựa chọn mô hình tốt nhất được thực hiện bằng cách sử dụng một số tiêu chí thông tin (AICC, HQC, AIC, và SBC). Các giá trị nhỏ nhất

của những tiêu chí này cho biết mô hình tốt nhất. Mô hình phù hợp nhất được chọn là VARMA(2, 1). Kết quả dự báo cho thấy sai số chuẩn tăng lên theo thời gian; sai số chuẩn trong tháng đầu tiên tương đối nhỏ so với dự đoán trung bình, nhưng tăng lên theo thời gian đến dự báo cho 12 tháng tiếp theo. Điều này cho thấy mô hình là hợp lý khi dự báo cho các khoảng thời gian ngắn, nhưng kết quả không ổn định (do sai số chuẩn cao hơn) khi dự báo cho các khoảng thời gian dài.

3 PHƯƠNG PHÁP

3.1 THUẬT TOÁN VARMA

Mô hình VARMA thường được sử dụng để dự báo dữ liệu chuỗi thời gian đa biến. Mô hình VARMA là một mở rộng của mô hình ARMA trong chuỗi thời gian đơn biến (Lutkepohl, 2005; Wei, 1990) và được sử dụng với điều kiện dữ liệu phải là dừng theo thời gian (Lutkepohl, 2005). Mô hình VARMA (p,q) là sự kết hợp của mô hình VAR (p) và mô hình trung bình trượt vector (q) (VMA (q)).

- Công thức:

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^p \Phi_i \mathbf{y}_{t-i} + \sum_{j=1}^q \Theta_j \epsilon_{t-j} + \epsilon_t \quad (1)$$

Trong đó:

- \mathbf{y}_t là vector của các biến tại thời điểm t .
- \mathbf{c} là một vector hằng số.
- Φ_i là ma trận hệ số của chuỗi thời gian (độ trễ, lag) i của mô hình AR.
- Θ_j là ma trận hệ số của chuỗi thời gian (độ trễ, lag) j của mô hình MA.
- ϵ_t là vector của các nhiễu ngẫu nhiên tại thời điểm t .

4 THỰC NGHIỆM

4.1 MÔ HÌNH DỰ ĐOÁN

4.1.1 ARIMA

4.1.2 LINEAR REGRESSION

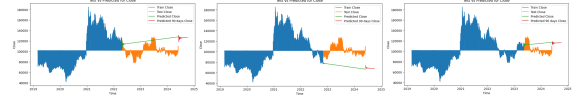
4.1.3 VARMA



Hình 1: LG 30 DAYS 6:4, 7:3, 8:2



Hình 2: LG 60 DAYS 6:4, 7:3, 8:2



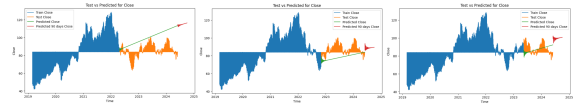
Hình 3: LG 90 DAYS 6:4, 7:3, 8:2



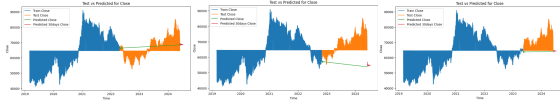
Hình 4: SONY 30 DAYS 6:4, 7:3, 8:2



Hình 5: SONY 60 DAYS 6:4, 7:3, 8:2



Hình 6: SONY 90 DAYS 6:4, 7:3, 8:2



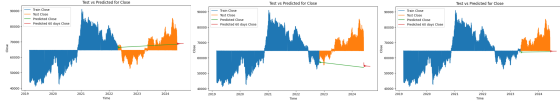
Hình 7: SAMSUNG 30 DAYS 6:4, 7:3, 8:2



Hình 9: SAMSUNG 90 DAYS 6:4, 7:3, 8:2

5 TÀI LIỆU THAM KHẢO

[1]: Warsono, Edwin Russel, Wamiliana*, Widiarti, Mustofa Usman, "Modeling and Forecasting by the Vector Autoregressive Moving Average Model for Export of Coal and Oil Data (Case Study from Indonesia over the Years 2002-2017)". International Journal of Energy Economics and Policy, 2019



Hình 8: SAMSUNG 60 DAYS 6:4, 7:3, 8:2