

Trường Chính sách công và Quản lý Fulbright

Học kỳ Xuân 2022

THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU

BÀI THI CUỐI KỲ

Ngày phát: Thứ bảy 19/02/2022

Ngày nộp: Chủ nhật 13/03/2022

Học viên nộp bản điện tử bao gồm file Word và Eviews trước 8h20 trên ứng dụng Microsoft Teams

Học viên cần trình bày rõ ràng bài làm trên Word

Câu 1: (40 điểm)

Để hoàn thành Câu 1 của bài thi này, Anh/Chị phải tiến hành thu thập thông tin về giá cổ phiếu của một doanh nghiệp đã niêm yết trên thị trường chứng khoán Việt Nam. Anh/Chị có thể chọn bất kỳ cổ phiếu nào mà anh/chị quan tâm, được ký hiệu là Y_t và thu thập dữ liệu trong khoảng thời gian từ ngày 03/01/2021 đến ngày 11/03/2022. Giá cổ phiếu được ghi nhận theo từng ngày giao dịch thực tế. Biến thời gian (t) được tạo tương ứng với mỗi ngày quan sát là một kỳ tương ứng.

Lưu ý: Trong 5 loại giá là (giá mở cửa, giá đóng cửa, giá điều chỉnh, giá cao nhất hoặc giá thấp nhất trong ngày) thì Anh/Chị cần sử dụng giá điều chỉnh để phản ánh đúng nhất giá trị của cổ phiếu.

- Vẽ đồ thị giá cổ phiếu (Y_t) theo thời gian (t). Tính toán một số trị thống kê tổng hợp cho dữ liệu giá cổ phiếu Y_t (số quan sát, trung bình, trung vị, độ lệch chuẩn, giá trị nhỏ nhất, giá trị lớn nhất, khoảng biến thiên). Hãy đưa ra nhận xét ngắn gọn từ đồ thị và kết quả tính toán của Anh/Chị? (2,5 điểm)
- Dựa vào dữ liệu đã thu thập, Anh/Chị hãy ước lượng hai mô hình dưới đây. Từ kết quả ước lượng hãy xác định các giá trị MAD (Mean Absolute Deviation hay Mean Absolute Error – Độ lệch tuyệt đối trung bình) và RMSE (Root Mean Square Deviation hay Root Mean Square Error – Căn bậc hai của Độ lệch bình phương trung bình) của từng mô hình? (2,5 điểm)

Mô hình 1: Mô hình xu hướng tuyến tính: $Y_t = \beta_1 + \beta_2 t + \varepsilon_t$

Mô hình 2: Mô hình tự hồi quy bậc 1 [AR(1)]: $Y_t = \delta_1 + \delta_2 Y_{t-1} + v_t$

- c. Anh/Chị hãy vẽ đồ thị đường giá cổ phiếu thực (Y_t mà anh/chị đã thu thập) cùng với các giá trị đã dự báo từ Mô hình 1 và Mô hình 2 trên cùng một đồ thị. Căn cứ từ kết quả đồ thị cũng như các giá trị MAD và RMSE đã tính, Anh/Chị sẽ chọn mô hình nào trong 2 mô hình ở Mục b? Giải thích ngắn gọn? (2,5 điểm)
- d. Dựa vào các dữ liệu quan sát thu thập, Anh/Chị hãy tính toán hệ số tương quan mẫu (SAC) và hệ số tương quan riêng phần mẫu (SPAC) cho 36 độ trễ của chuỗi Y_t ? Dựa vào kết quả, anh/chị có thể kết luận gì về tính dừng và tính mùa vụ của chuỗi Y_t ? Giải thích? (2,5 điểm)
- e. Anh/Chị hãy thực hiện kiểm định Dickey Fuller đối với chuỗi Y_t ? Kết luận về tính dừng và tính mùa vụ từ kết quả kiểm định Dickey Fuller? Kết quả này có tương thích với kết luận ở các mục trước? (2,5 điểm)
- f. Từ dữ liệu đã thu thập (Y_t), Anh/Chị hãy tạo ra một chuỗi dữ liệu mới (dY_t) bằng cách lấy sai phân bậc 1 của chuỗi Y_t . Sau đó hãy tính toán hệ số tương quan mẫu (SAC) và hệ số tương quan riêng phần mẫu (SPAC) cho 36 độ trễ của chuỗi dY_t . Kết luận về tính dừng và tính mùa vụ của chuỗi dY_t ? (2,5 điểm)
- Ở độ trễ $k = 2$ của chuỗi dY_t : Anh/Chị hãy giải thích ý nghĩa của hệ số tương quan mẫu (SAC)? Thực hiện kiểm định để xem hệ số này bằng 0 hay khác 0 về mặt thống kê ở mức ý nghĩa 5%?
 - Ở độ trễ $k = 3$ của chuỗi dY_t : Anh/Chị hãy giải thích ý nghĩa của hệ số tương quan riêng phần mẫu (SPAC)? Thực hiện kiểm định để xem hệ số này bằng 0 hay khác 0 về mặt thống kê ở mức ý nghĩa 5%?
- g. Từ dữ liệu đã thu thập (Y_t), Anh/Chị hãy tạo ra một chuỗi dữ liệu mới ($d2Y_t$) bằng cách lấy sai phân bậc 2 của chuỗi Y_t . Sau đó hãy tính toán hệ số tương quan mẫu (SAC) và hệ số tương quan riêng phần mẫu (SPAC) cho 36 độ trễ của chuỗi $d2Y_t$. Kết luận về tính dừng và tính mùa vụ của chuỗi $d2Y_t$? (2,5 điểm)
- h. Dựa trên các kết quả ở các mục trước, Anh/Chị hãy xác định bậc p , d , q phù hợp cho mô hình ARIMA (p , d , q). Từ đó, liệt kê các mô hình ARIMA có thể có? (2,5 điểm)
- i. Hãy viết phương trình ARIMA (1, 1, 1) cho chuỗi Y_t ? (2,5 điểm)
- j. Dựa vào kết quả ở Mục h, Anh/Chị hãy chọn ra 2 mô hình ARIMA mà anh/chị cho là hiệu quả nhất, tạm gọi là Mô hình A và Mô hình B. Đối với tổng số quan sát đã thu thập trong chuỗi Y_t , anh/chị hãy tách ra 20% số quan sát cuối cùng (20 quan sát gần nhất) để thực hiện các tính toán và dự báo ngoài mẫu; 80% số quan sát còn lại được sử dụng cho các tính toán trong mẫu. Sau đó, hãy tiến hành thực hiện các tính toán, dự báo trong và ngoài mẫu đối với 2 mô hình mà anh/chị đã lựa chọn. Kết quả sau cùng được tổng hợp như bảng sau (2,5 điểm):

Mô hình	MAD	MAD		RMSE	RMSE
	Trong mẫu (80%)	Ngoài mẫu (20%)		Trong mẫu (80%)	Ngoài mẫu (20%)
Mô hình A					
Mô hình B					

- k. Dựa vào kết quả ở Mục j, Anh/Chị hãy lựa chọn Mô hình phù hợp nhất? Giải thích. (2,5 điểm)
- l. Từ kết quả lựa chọn ở Mục k, Anh/Chị hãy ước lượng lại mô hình đã chọn nhưng thực hiện trên tất cả các quan sát đã thu thập (Y_t)? (2,5 điểm)
- m. Dựa vào kết quả ước lượng, hãy tiến hành dự báo điểm và dự báo khoảng cho 3 quan sát tiếp theo của giá cổ phiếu mà anh chị đang phân tích. Kết quả dự báo được tổng hợp vào bảng sau (10 điểm):

Giá cổ phiếu [Mã cổ phiếu]	Day +1 (14/03/2022)	Day +2 (15/03/2022)	Day +3 (16/03/2022)
Giá trị dự báo điểm			
Giá trị dự báo Min			
Giá trị dự báo Max			
Giá trị thực tế ^(*)			
	(4 điểm)	(3 điểm)	(3 điểm)

^(*) Giá trị thực tế anh/chị sẽ nộp lại cho Ban giảng viên trước 8h20 ngày 17/03/2022 (Anh/Chị chỉ cần nộp bằng kết quả bằng file điện tử qua ứng dụng Microsoft Teams cho Trợ giảng)

Câu 2: (30 điểm)

File Transportation.xlsx là dữ liệu từ một nghiên cứu mẫu ngẫu nhiên bao gồm 172 sinh viên Đại học Quốc gia Tp. Hồ Chí Minh (ĐHQG-HCM) về khả năng chấp nhận chuyển đổi phương tiện giao thông hiện tại sang sử dụng phương tiện công cộng khi Tuyến Metro số 1 (Bến Thành – Suối Tiên) được xây dựng xong và đi vào hoạt động. Từ file dữ liệu, Anh/Chị hãy trả lời các câu hỏi sau:

- a. Anh/Chị kỳ vọng thế nào về dấu và ý nghĩa thống kê của tác động đối với các biến giải thích sau: **sex, part_time, living_location, weekly_public_transport, freq_motorbike, freq_bus, own, income, travel_expenditure, monthly_expenditure, savings** lên khả năng chấp nhận chuyển đổi phương tiện của sinh viên ĐHQG-HCM? Giải thích? (5 điểm)
- b. Ước lượng bằng mô hình logit và viết phương trình hàm hồi quy ước lượng cho mô hình với những biến giải thích: **sex, part_time, living_location, weekly_public_transport, freq_motorbike, freq_bus, own, income, travel_expenditure, monthly_expenditure,**

savings. Từ kết quả có được, theo Anh/chị biến nào có tác động và biến nào không tác động tới khả năng chấp nhận chuyển đổi phương tiện của sinh viên ĐHQG-HCM? Giải thích kết quả? (6 điểm)

- c. Anh/Chị hãy giải thích ý nghĩa hệ số hồi quy đứng trước biến **weekly_public_transport**, **living_location** và **income**? Theo anh chị, vì sao biến **weekly_public_transport** ngược với kỳ vọng? Độ thích hợp của mô hình ước lượng là bao nhiêu? Hãy phát biểu và kiểm định giả thuyết về ý nghĩa chung của mô hình? (8 điểm)
- d. Theo Anh/Chị tác động biên và tác động biên tức thời của thu nhập lên xác suất sinh viên chuyển đổi phương tiện là bao nhiêu xét tại trung bình của các biến giải thích? (6 điểm)
- e. Xét xác suất ban đầu tại giá trị của các biến X như sau: **sex = 1; living_location = 1; weekly_public_transport = 7; freq_motorbike = 1; freq_bus = 4, own = 0; income = 3500; travel_expenditure = 150; monthly_expenditure = 3000; savings = 500.** Theo Anh/chị, trong điều kiện các yếu tố khác không đổi, xác suất sinh viên có đi làm thêm và không đi làm thêm sẽ chấp nhận chuyển đổi phương tiện khác nhau như thế nào? (5 điểm)

Câu 3: (30 điểm)

1. Dựa vào các mô hình ra quyết định trong điều kiện không chắc chắn và trong điều kiện rủi ro đã học trong môn học này; căn cứ vào hoạt động của cơ quan/công ty trong thực tiễn. Anh/Chị hãy:
 - a. Hình thành vấn đề cần ra quyết định (10 điểm):
 - Xác định rõ vấn đề cần giải quyết
 - Liệt kê các phương án có thể có để giải quyết vấn đề
 - Nhận ra các tình huống hay các trạng thái có thể xảy ra
 - Xác định các lợi ích hoặc các chi phí có thể có cho mỗi phương án ứng với mỗi trạng thái
 - b. Lựa chọn mô hình ra quyết định phù hợp với vấn đề cần ra quyết định (2,5 điểm)
 - c. Thu thập dữ liệu cần cho mô hình. Giải thích ngắn gọn việc thu thập dữ liệu (2,5 điểm)
 - d. Áp dụng mô hình để tìm lời giải ra quyết định (2,5 điểm)
 - e. Đề xuất giải pháp giải quyết vấn đề (2,5 điểm)
Giải thích ngắn gọn các nội dung Anh /Chị đã làm.
2. Căn cứ vào hoạt động thực tiễn của cơ quan/công ty Anh/Chị, hãy xác định một số rủi ro có thể xảy ra quá trình hoạt động.
 - a. Đánh giá khả năng xảy ra của các rủi ro này (2,5 điểm)
 - b. Đánh giá tác động của các rủi ro này (2,5 điểm)
 - c. Áp dụng ma trận P.I để đánh giá sắp hạng mức độ quan trọng của các rủi ro này (2,5 điểm)
 - d. Đề xuất giải pháp để quản lý các rủi ro này (2,5 điểm)
Giải thích ngắn gọn các nội dung Anh /Chị đã làm.