

Báo cáo Đồ án Cuối kì

Nhóm 2

Giảng viên: Nguyễn Tiến Đạt

Sinh viên: D.H.N. Trâm - N.T. Mây - N.T.K. Ngân - N.T. Ngân



Mục lục

- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo

Mục lục

- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo

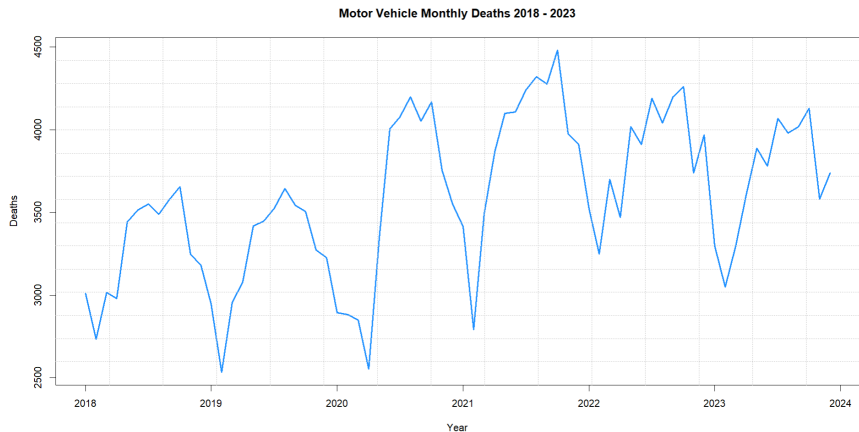
Giới thiệu time series đã chọn

- **Dữ liệu:** Số ca tử vong do tai nạn xe cơ giới ghi nhận theo tháng tại Mỹ, giai đoạn 2018 - 2023.
- **Nguồn dữ liệu:** Thống kê chấn thương - NSC (National Security Council - Hội đồng An ninh Quốc gia Hoa Kỳ).
- **Lí do chọn dữ liệu:** Quan tâm đến lĩnh vực bảo hiểm. Việc thống kê số lượng tai nạn xe cơ giới theo từng tháng trong năm có vai trò quan trọng trong việc tính toán rủi ro, xác định phí bảo hiểm và dự đoán các khoản bồi thường bảo hiểm trong năm.

Mục lục

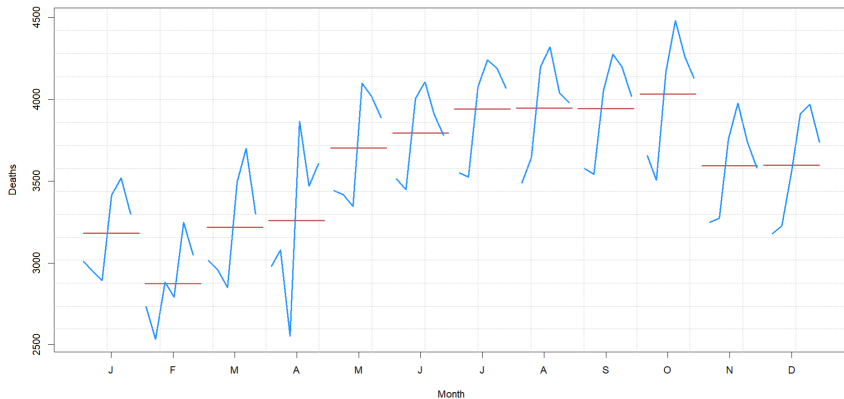
- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu**
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo

Trực quan hoá dữ liệu



Hình 1: Số ca tử vong do tai nạn xe cơ giới theo tháng tại Mỹ (2018 - 2023).

Monthplot



Hình 2: Monthplot cho số ca tử vong do tai nạn xe cơ giới theo tháng tại Mỹ (2018 - 2023).

Mô tả các biểu hiện nổi bật

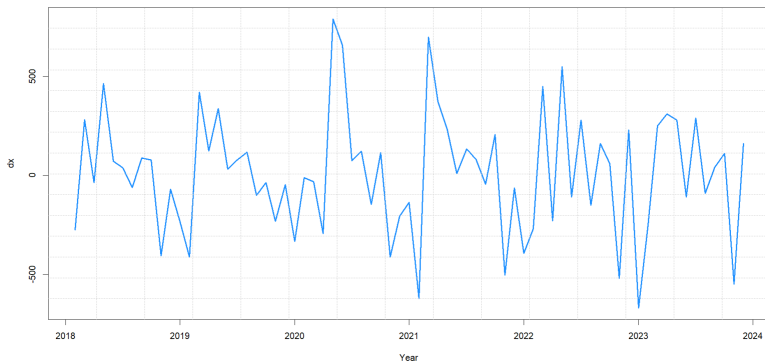
- Số ca tử vong có **xu hướng tăng nhẹ** trong giai đoạn 2018 - 2023.
- Đồ thị về số ca tử vong có xu hướng **lặp lại hình dáng** sau một chu kì nhất định (1 năm).
- Trong mỗi năm, số ca tử vong **cao nhất** vào khoảng tháng 08 đến tháng 10.
- Trong mỗi năm, số ca tử vong **thấp nhất** vào khoảng tháng 02.

Mục lục

- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA**
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo

Loại bỏ yếu tố xu hướng (trend)

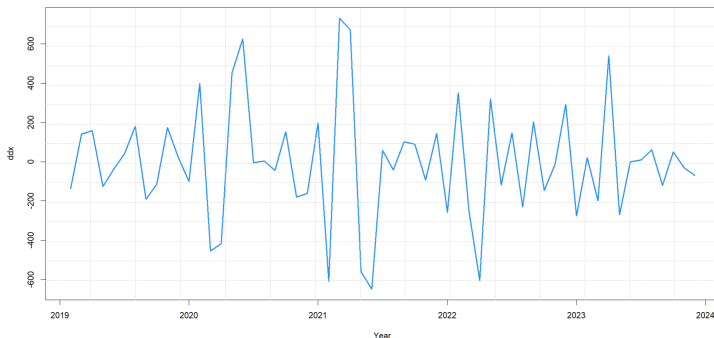
- Dữ liệu ban đầu có xu hướng (trend) tăng nhẹ, áp dụng **difference** cho dữ liệu nhằm loại bỏ yếu tố xu hướng.



Hình 3: Dữ liệu sau khi áp dụng difference lần một.

Loại bỏ yếu tố theo mùa (seasonality)

- Dữ liệu vẫn tồn tại yếu tố theo mùa (seasonality), áp dụng một toán tử **difference bậc $D = 12$** nhằm loại bỏ yếu tố theo mùa.

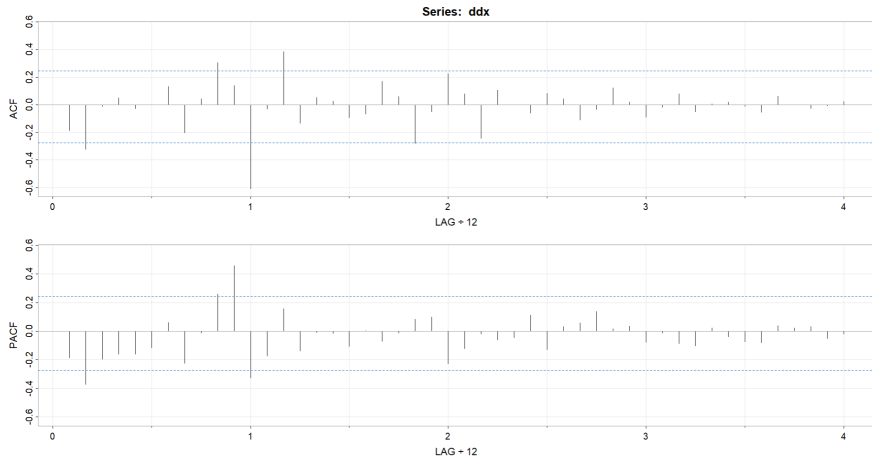


Hình 4: Dữ liệu sau khi áp dụng toán tử difference bậc $D = 12$.

- Dữ liệu có vẻ dừng và loại bỏ được yếu tố theo mùa. Ta tiến hành xây dựng mô hình.

Xác định bậc phụ thuộc của mô hình

■ Hàm ACF và PACF của ddx



Hình 5: Hàm ACF và PACF của ddx

Ước lượng các tham số của mô hình

■ Một số mô hình có thể sử dụng

	AIC	AIC _c	BIC
SARIMA (2, 1, 0) × (1, 1, 1) ₁₂	13.69472	13.70728	13.87079
SARIMA (0, 1, 2) × (1, 1, 1) ₁₂	13.65015	13.66271	13.82621
SARIMA (2, 1, 2) × (1, 1, 1) ₁₂	13.66136	13.68874	13.90784

■ Theo các tiêu chí trên, ta ưu tiên mô hình **SARIMA**(0, 1, 2) × (1, 1, 1)₁₂ để dự báo.

Mục lục

- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình**
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo

Thể hiện kết quả ước lượng tham số

Mô hình **SARIMA**(0, 1, 2) \times (1, 1, 1)₁₂ có dạng

$$(1 - \Phi_1 B^{12})(1 - B^{12})(1 - B)X_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B + \theta_2 B^2)W_t$$

Thay các kết quả ước lượng từ bảng dưới

$$(1 + 0.1123 B^{12})(1 - B^{12})(1 - B)X_t = (1 - 0.9995 B^{12})(1 - 0.4205 B - 0.1509 B^2)W_t$$

Coefficients:

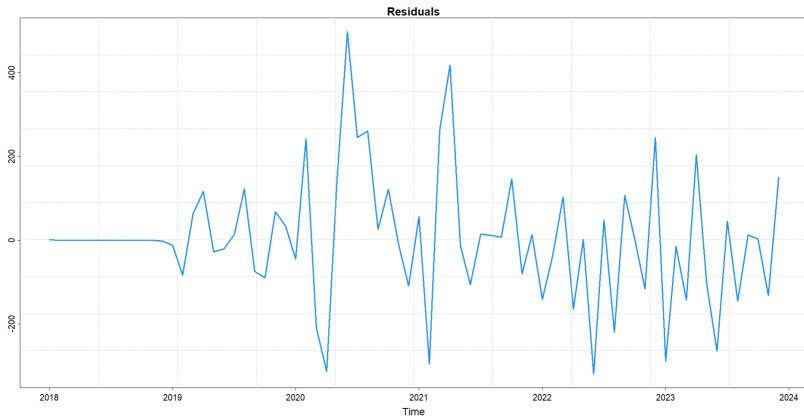
	Estimate	SE	t.value	p.value
ma1	-0.4205	0.1296	-3.2446	0.0020
ma2	-0.1509	0.1288	-1.1714	0.2465
sar1	-0.1123	0.1469	-0.7645	0.4478
sma1	-0.9995	0.4879	-2.0487	0.0453

Hình 6: Ước lượng tham số cho mô hình

Mục lục

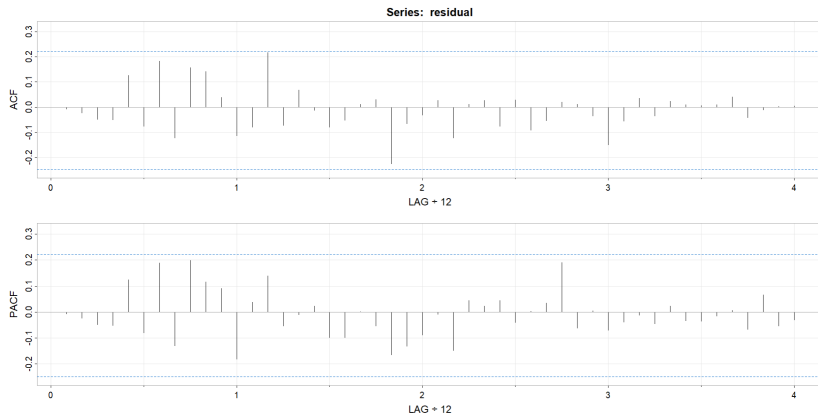
- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals**
- 6 Dự báo cho 12 tháng tiếp theo

Residuals plot



Hình 7: Residual analyst for $\text{SARIMA}(0, 1, 2) \times (1, 1, 1)_{12}$

Hàm ACF và PACF của residuals



Hình 8: Residual analyst for $\text{SARIMA}(0, 1, 2) \times (1, 1, 1)_{12}$

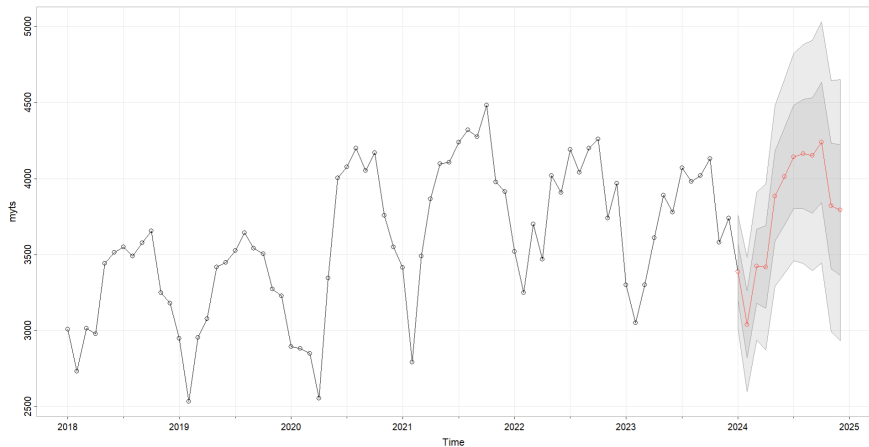
Nhận xét

- Phần dư có trung bình bằng 0 và phương sai là hằng số không phụ thuộc vào t .
- Sự tương quan (dựa vào đồ thị ACF) và sự tương quan riêng (dựa vào đồ thị PACF) giữa các thời điểm khác nhau gần như bằng 0. Nghĩa là X_t và X_s không tương quan (uncorrelated) $\forall t \neq s$.
- **Kết luận:** Phần dư thoả giả định là một White noise.

Mục lục

- 1 Giới thiệu time series data đã chọn
- 2 Trực quan hoá dữ liệu
- 3 Xây dựng mô hình SARIMA
- 4 Thể hiện kết quả ước lượng cho các tham số trong mô hình
- 5 Residuals plot, hàm ACF và PACF của residuals
- 6 Dự báo cho 12 tháng tiếp theo**

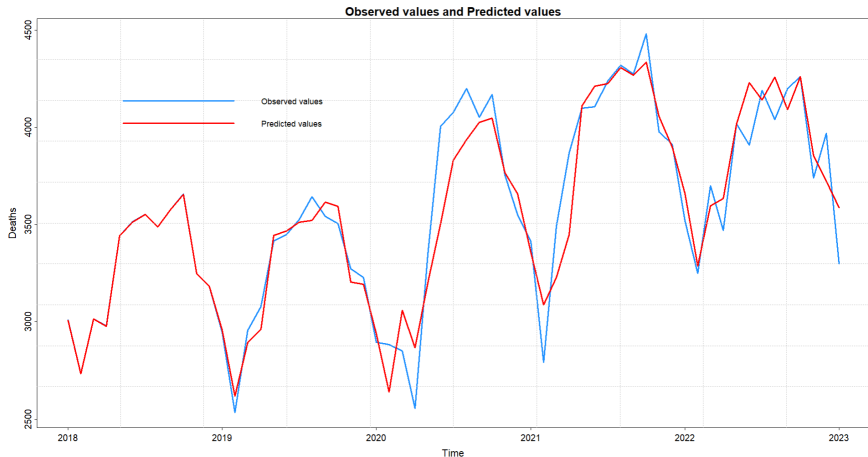
Dự báo cho 12 tháng tiếp theo



Hình 9: Dự báo 12 tháng tiếp theo bằng mô hình $SARIMA(0, 1, 2) \times (1, 1, 1)_{12}$.

Giải thích vì sao Khoảng tin cậy cho dự báo lớn

- Mô hình chưa nắm bắt tốt xu hướng, tính theo mùa của dữ liệu.



Hình 10: Dữ liệu thực tế so với dữ liệu từ mô hình $SARIMA(0, 1, 2) \times (1, 1, 1)_{12}$.

Giải thích vì sao Khoảng tin cậy cho dự báo lớn

- Thời điểm dự báo càng xa so với thời điểm ghi nhận dữ liệu làm cho khoảng tin cậy càng lớn.

- Khoảng tin cậy $(1 - \alpha)$ cho các giá trị dự báo có dạng

$$\hat{X}_{n+m}^n - c_{\frac{\alpha}{2}} \sqrt{P_{n+m}^n} \leq X_{n+m}^n \leq \hat{X}_{n+m}^n + c_{\frac{\alpha}{2}} \sqrt{P_{n+m}^n}$$

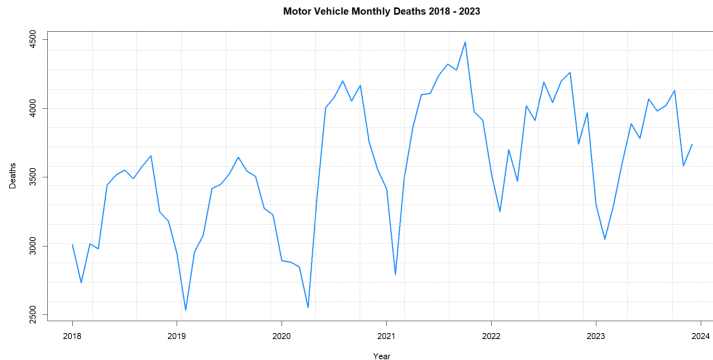
- Khi n đủ lớn, sai số dự báo được xấp xỉ

$$\hat{P}_{n+m}^n = \hat{\sigma}_W^2 \sum_{j=1}^{m-1} \widehat{\psi_j^*}^2$$

- Với m càng lớn, tổng trong \hat{P}_{n+m}^n càng có nhiều số hạng. Do đó, ở thời điểm dự báo càng xa so với thời điểm ghi nhận dữ liệu thì khoảng tin cậy cho dự báo càng lớn.

Giải thích vì sao Khoảng tin cậy cho dự báo lớn

- Số ca tử vong do tai nạn xe cơ giới có thể bị ảnh hưởng bởi các sự kiện nhỏ của từng bang làm cho kết quả dự báo sai lệch.



Hình 11: Số ca tử vong do tai nạn xe cơ giới theo tháng tại Mỹ (2018 - 2023).