

BÀI KIỂM TRA PHÂN TÍCH CHUỖI THỜI GIAN

Họ Và Tên: Nguyễn Đỗ Cường

MSV: 2151264647

Đề 3: Phân tích Sarima, Arimax

Bài 1:

SARIMA:

Mô hình trung bình trượt tích hợp tự hồi quy theo mùa (SARIMA), là một mô hình mở rộng của ARIMA hỗ trợ dữ liệu chuỗi thời gian đơn biến với thành phần theo mùa. SARIMA bổ sung thêm ba tham số mới là thành phần tự hồi quy (AR), tích hợp (I) và trung bình động (MA) cho thành phần theo mùa của chuỗi thời gian đơn biến, cũng như một tham số xác định khoảng thời gian theo mùa. Việc xác định bộ tham số của mô hình SARIMA yêu cầu chọn các tham số cho cả các yếu tố xu hướng và theo mùa của chuỗi.

Ba tham số xu hướng giống như mô hình ARIMA:

- p: Bậc tự hồi quy xu hướng.
- d: Bậc tích hợp xu hướng
- q: Bậc trung bình động xu hướng.

Bốn tham số theo mùa không phải là một phần của ARIMA phải được xác định:

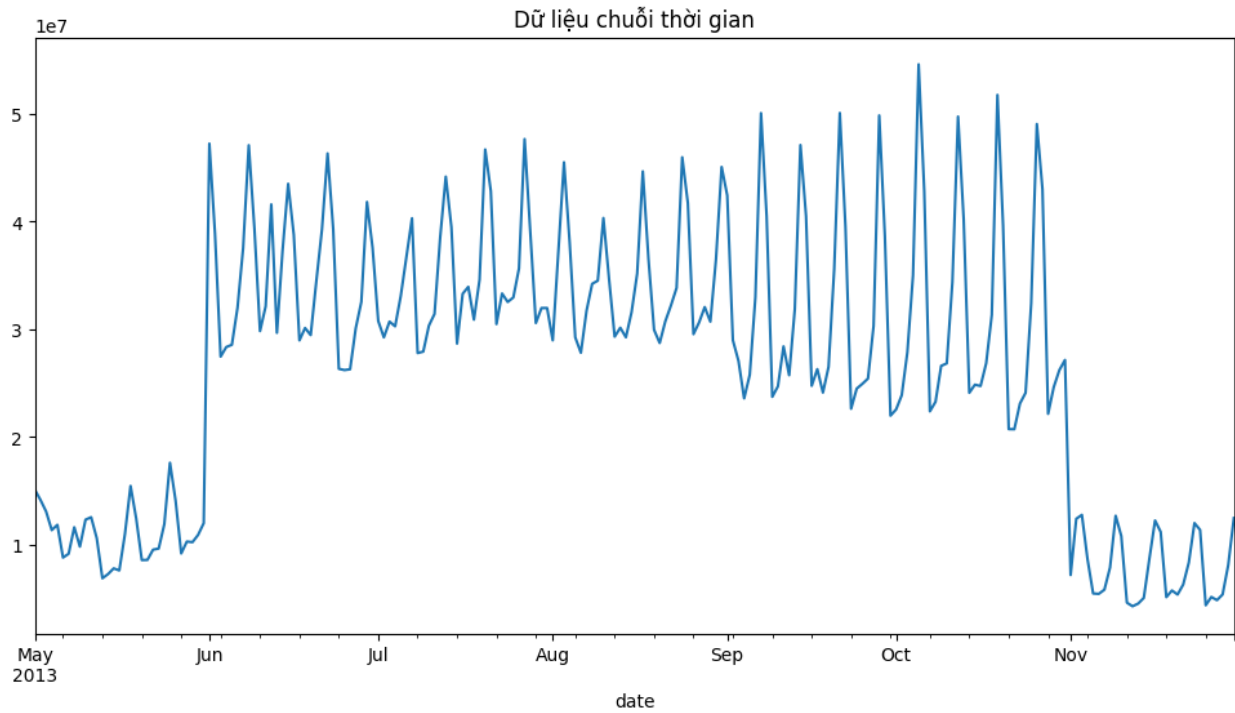
- P: Bậc tự hồi quy theo mùa.
- D: Bậc tích hợp theo mùa.
- Q: Bậc trung bình động theo mùa.
- m: Số bước thời gian (kỳ) trong một mùa.

Như vậy, tùy theo giá trị của các tham số mà mô hình SARIMA có thể là mô hình ARIMA, ARMA, AR hay là MA.

ARIMAX:

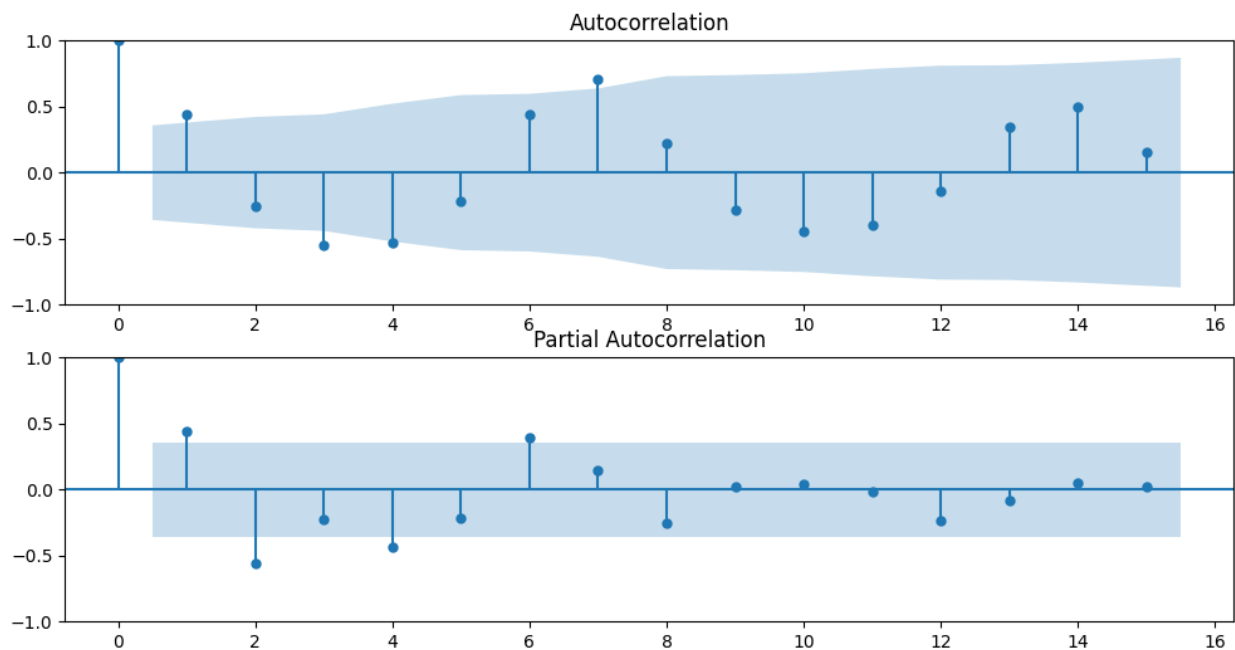
Mô hình ARIMAX cũng là một biến thể mở rộng của mô hình ARIMA, điểm khác biệt của ARIMAX bao gồm các biến ngoại sinh. X(Exogenous variables) là các yếu tố bên ngoài không thuộc chuỗi thời gian nhưng có thể tác động đáng kể đến chuỗi thời gian.

Bài 2:



Dữ liệu của “truong_3” từ tháng 5 đến tháng 11 trong năm 2013 có xu hướng giảm vào tháng 5 và tháng 11 duy trì trong mức từ 0-2, các tháng còn lại duy trì ở cao từ 2-5.

SARIMA:



Biểu đồ tự tương quan (ACF) và biểu đồ tự tương quan từng phần (PACF)

Trong biểu đồ ACF, tự tương quan tại độ trễ 1 là dương và đáng kể. Điều này cho thấy rằng giá trị hiện tại của biến thời gian có mối tương quan tích cực với giá trị trước đó của nó. Nói cách khác, nếu giá trị hiện tại cao, thì giá trị trước đó cũng có khả năng cao. Tự tương quan tại các độ trễ cao hơn không đáng kể. Điều này cho thấy rằng không có mối tương quan đáng kể giữa giá trị hiện tại và các giá trị quá khứ hơn một khoảng thời gian.

Trong biểu đồ PACF, Tự tương quan từng phần tại độ trễ 1 là dương và đáng kể. Điều này cho thấy rằng giá trị hiện tại của biến thời gian có mối tương quan tích cực với giá trị trước đó của nó, ngay cả sau khi điều chỉnh cho mối tương quan với các giá trị quá khứ khác. Tự tương quan từng phần tại các độ trễ cao hơn không đáng kể. Điều này cho thấy rằng không có mối tương quan đáng kể giữa giá trị hiện tại và các giá trị quá khứ hơn một khoảng thời gian, ngay cả sau khi điều chỉnh cho mối tương quan với các giá trị quá khứ khác.

```
# Xây dựng và huấn luyện mô hình SARIMA
# Tham số order và seasonal_order có thể được điều chỉnh dựa trên ACF và PACF
model = SARIMAX(y_scaled, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))
results = model.fit()

# Hiển thị kết quả mô hình
print(results.summary())

# Chẩn đoán mô hình
results.plot_diagnostics(figsize=(15, 12))
plt.show()

# Dự báo
forecast_scaled = results.get_forecast(steps=12)
forecast_mean_scaled = forecast_scaled.predicted_mean
forecast_conf_int_scaled = forecast_scaled.conf_int()
```

Train mô hình

Tham số được sử dụng: $p=1$, $d=1$, $q=1$

```

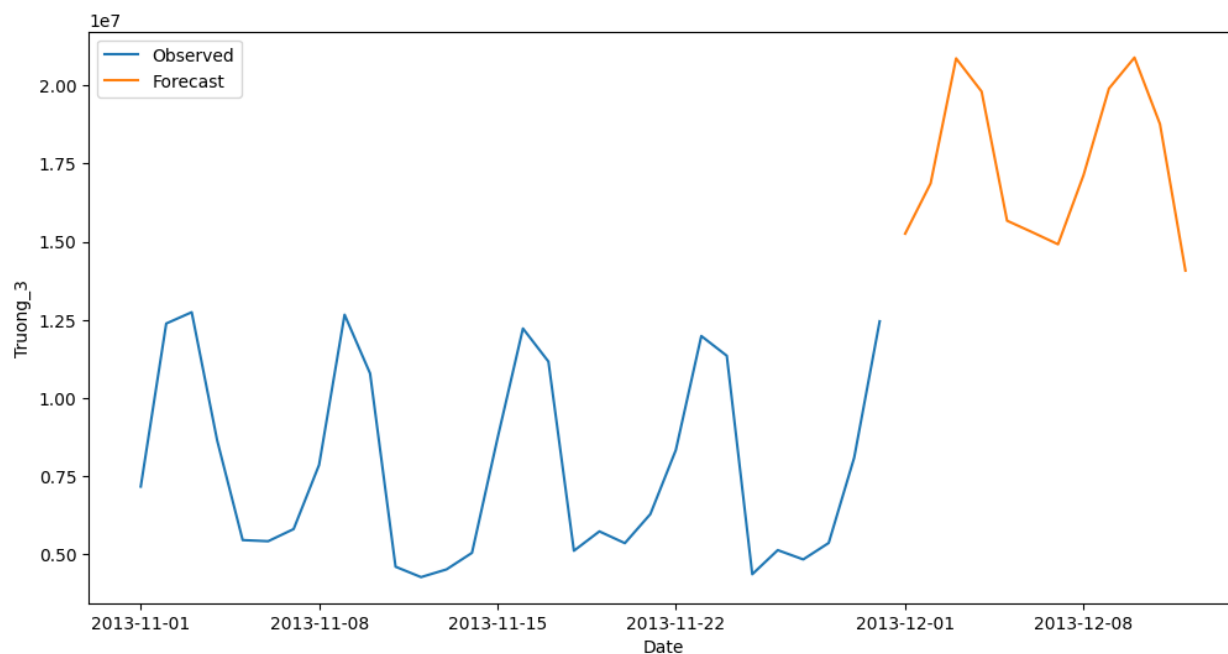
SARIMAX Results
=====
Dep. Variable:          truong_3    No. Observations:          30
Model:                 SARIMAX(1, 1, 1)x(1, 1, 1, 12)    Log Likelihood              -29.672
Date:                  Tue, 04 Jun 2024    AIC                        69.345
Time:                  04:36:54    BIC                       73.511
Sample:                11-01-2013    HQIC                      69.759
                  - 11-30-2013
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1          0.0266      0.518      0.051      0.959     -0.989     1.043
ma.L1          0.6429      0.577      1.114      0.265     -0.488     1.774
ar.S.L12       -0.6205    2.99e+05   -2.07e-06    1.000   -5.86e+05    5.86e+05
ma.S.L12       -0.5560    1.57e+06   -3.53e-07    1.000   -3.09e+06    3.09e+06
sigma2         0.7635    8.27e+05   9.23e-07    1.000   -1.62e+06    1.62e+06
=====
Ljung-Box (L1) (Q):          0.52    Jarque-Bera (JB):          1.56
Prob(Q):                    0.47    Prob(JB):              0.46
Heteroskedasticity (H):      1.47    Skew:                  -0.25
Prob(H) (two-sided):         0.65    Kurtosis:              1.61
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

```

Tóm tắt quá trình train

Kết quả:



Từ biểu đồ ta có thể thấy giá trị dự đoán có phần cao hơn so với giá trị quan sát

ARIMAX:

```
[55] # Chọn biến mục tiêu
      y = df_november['truong_3']

      # Chuẩn hóa dữ liệu nếu cần
      scaler_y = StandardScaler()
      y_scaled = scaler_y.fit_transform(y.values.reshape(-1, 1))

      # Xây dựng DataFrame mới chỉ với biến mục tiêu đã chuẩn hóa
      df_arimax = pd.DataFrame({
          'truong_3': y_scaled.flatten(),
          'truong_1': X_scaled[:, 0],
          'truong_5': X_scaled[:, 1]
      })

      # Xây dựng và huấn luyện mô hình SARIMAX với biến giải thích
      model_arimax = SARIMAX(df_arimax['truong_3'], exog=df_arimax[['truong_1', 'truong_5']],
                             order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))
      results_arimax = model_arimax.fit()

      # Hiển thị kết quả mô hình ARIMAX
      print(results_arimax.summary())
```

Train mô hình

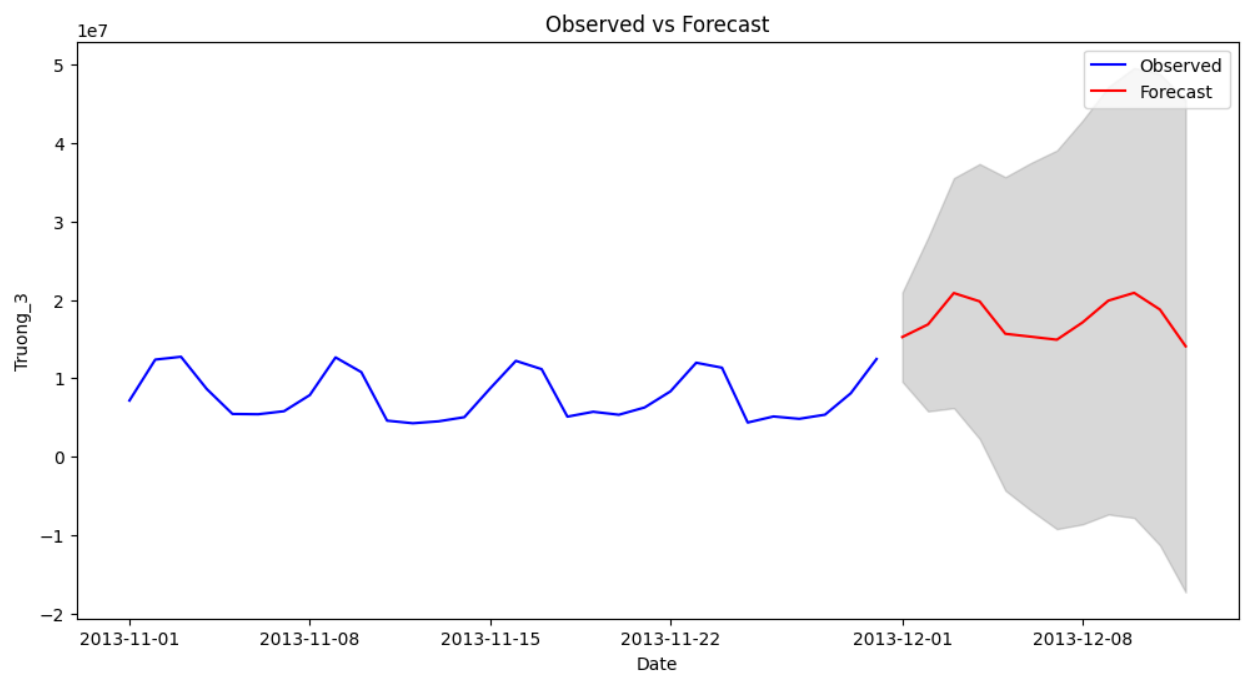
Giá trị ngoại sinh là cột `truong_1` và `truong_5`

SARIMAX Results						
=====						
Dep. Variable:	truong_3			No. Observations:	30	
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)			Log Likelihood	13.046	
Date:	Tue, 04 Jun 2024			AIC	-12.092	
Time:	04:44:49			BIC	-6.260	
Sample:	0			HQIC	-11.512	
	- 30					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

truong_1	1.0212	0.126	8.111	0.000	0.774	1.268
truong_5	-0.0386	0.123	-0.313	0.754	-0.280	0.203
ar.L1	-0.9862	4.195	-0.235	0.814	-9.209	7.237
ma.L1	0.4386	0.752	0.583	0.560	-1.036	1.913
ar.S.L12	-0.6882	2.625	-0.262	0.793	-5.834	4.457
ma.S.L12	-0.7927	49.408	-0.016	0.987	-97.630	96.045
sigma2	0.0035	0.149	0.024	0.981	-0.288	0.295
=====						
Ljung-Box (L1) (Q):	0.02	Jarque-Bera (JB):	5.61			
Prob(Q):	0.90	Prob(JB):	0.06			
Heteroskedasticity (H):	0.28	Skew:	1.13			
Prob(H) (two-sided):	0.15	Kurtosis:	4.67			
=====						
Warnings:						
[1] Covariance matrix calculated using the outer product of gradients (complex-step).						

Tóm tắt quá trình train

Kết quả:



Kết quả dự đoán có phần chính xác hơn so với mô hình SARIMA.

Link Github: https://github.com/CuongNgD203/Time_Series_Test