Time Series Analysis for Beginners

Dr. Jeffrey Yau
UC Berkeley and University of San Diego

You will be able to apply one of the most popular classes of statistical time series models to model time series data and use the model for forecasting

Autoregressive Integrated Moving Average Process (ARIMA)

Seasonal ARIMA Process Formulation

The **SARIMA(p,d,q) x (P,D,Q)$_s$** model is specified as

$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^D y_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t$$

```
statsmodels.tsa.statespace.sarimax.SARIMAX(endog, exog=None, order=(1, 0, 0), seasonal_order=(0, 0, 0, 0), trend=None, measurement_error=False, time_varying_regression=False, mle_regression=True, simple_differencing=False, enforce_stationarity=True, enforce_invertibility=True, hamilton_representation=False, **kwargs)
```

https://www.statsmodels.org/0.8.0/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html#statsmodels.tsa.statespace.sarimax.SARIMAX

Why ARIMA Model?

Why ARIMA Model?

- A popular class of statistical time series forecasting models
- Intuitive
- Rich statistical history
- Variants of many successful forecasting methods

Learning Objectives

The attendees will learn the following concepts and techniques in this training course:

1. The key characteristics, which are distinguished from non-time series data, of time series data
2. Statistics for summarizing time series
3. Graphical techniques to describe characteristics of time series
4. Common use cases of the class of ARIMA models
5. Essential concepts required to appropriately apply the class of ARIMA models in practice
6. The advantages and disadvantages of the class of ARIMA models

What to expect in Today's Training?

A (roughly) 5-hour live training

Several sections

~35 minutes	~15 minutes	10 minutes
• Slide lecture • Demos using Jupyter notebooks	An Exercise Q&A	Breaks

- Codes will be explained and applied for forecasting
- The course will not be just coding
- Mathematical notations are used quite extensively

Today’s training is recorded

You will have access to the recording afterwards

Agenda

Section	Topic	Appx. Time
1	Introduction to time series analysis	40 minutes
2	Exploratory Time Series Data Analysis and ARMA Modeling	60 minutes
3	ARIMA Model Formulation	60 minutes
4	ARIMA Modeling	60 minutes
5	Seasonal ARIMA Modeling	60 minutes
6	Concluding Remarks	20 minutes

Segment 1: Introduction to time series analysis

1.1 Introduction and welcome to the course

1.2 Common use cases of time series analysis from different disciplines

1.3 Common characteristics and patterns of time series

1.4 The class of models to be covered today: A demo

Exercise 1

Segment 2: Exploratory Data Analysis and ARMA Modeling

2.1 A brief discussion on basic terminology for time series analysis

2.2 Exploratory Time Series Data Analysis

Exercise

2.3 Mathematical formulation of AR, MA, and ARMA models

2.4 Lag (or backshift) operators

2.5 Properties of the general AR, MA, ARMA models

2.6 ARMA modeling

Exercise

Segment 3: ARIMA Model Formulation

3.1 Notion of non-stationarity

3.2 Mathematical formulation of ARIMA models

3.3 The Box-Jenkins Approach to ARIMA Modeling of non-stationary time series

Exercise 3

Segment 4: ARIMA Modeling

4.1 Model Identification

4.2 Model Diagnostic Checking

4.3 Model performance evaluation (in-sample fit)

4.4 Forecasting and forecast evaluation

4.5 Incorporation of explanatory variables, its use cases, and its practical suggestions

Exercise 4

Segment 5: Seasonal ARIMA Modeling

5.1 Understanding seasonality and examination of seasonal time series

5.2 Mathematical formulation of Seasonal ARIMA (SARIMA) models

5.3 Building a seasonal ARIMA model for forecasting

Exercise 5

Segment 6: Concluding Remarks

6.1 Summary of today’s training

6.2 Course wrap-up and next steps, and where to go from here

Section 1

- Course intro
- Forecasting problem
- Time series patterns

Section 2

- Exploratory Time Series Data Analysis
- Formulating ARMA process

Section 3

- Non-stationarity concept
- Formulating ARIMA process

Section 4

- Developing ARIMA model
- Forecasting using ARIMA model

Section 5

- ARIMA models for time series with seasonality

Section 6

- Recap of today’s training
- A few concluding remarks

Pre-requisite and Setup

Pre-requisites:

- Working knowledge of Python
- Jupyter Notebook or Jupyterlab
- Working knowledge of the classical linear regression model

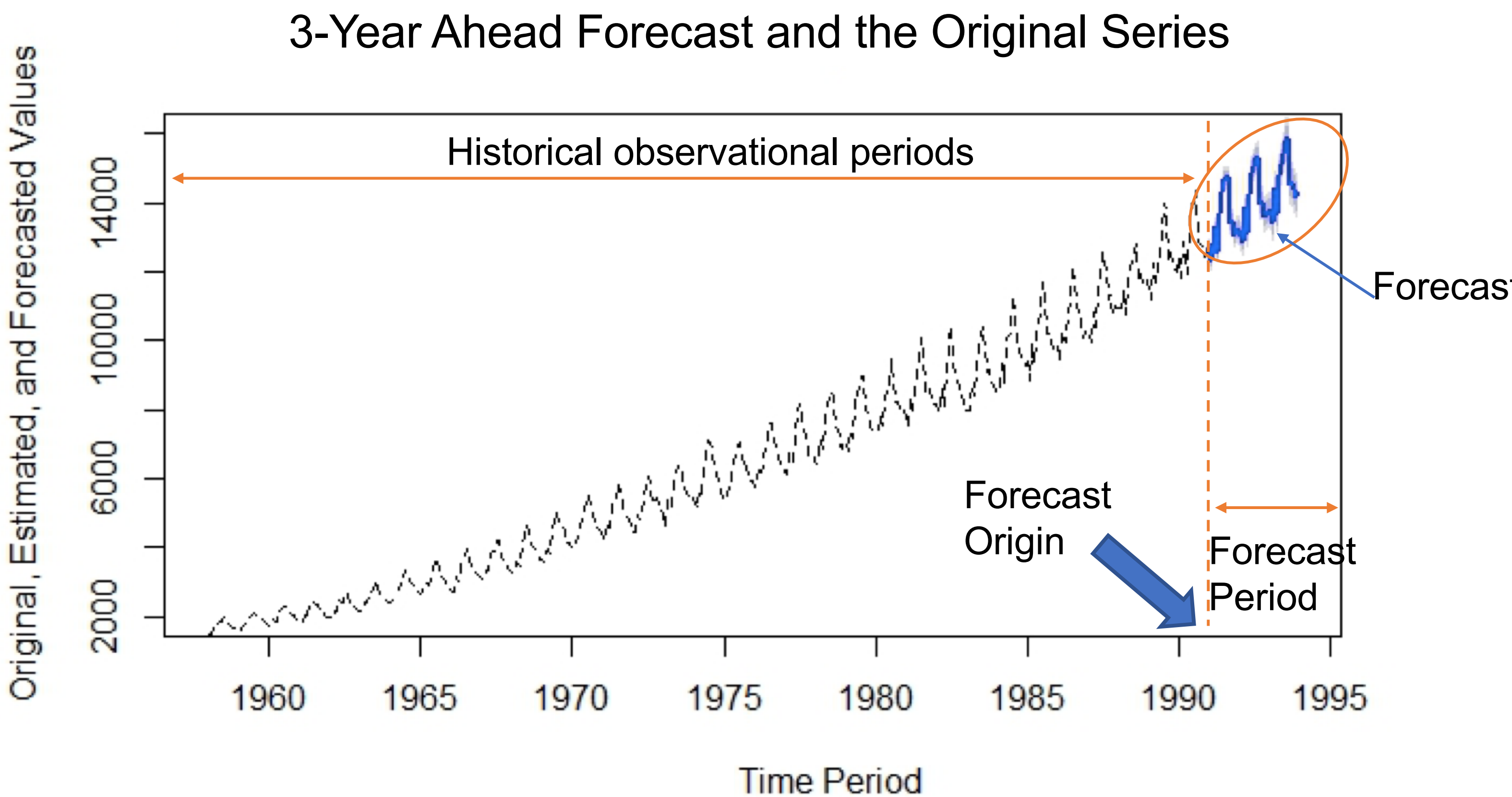
Setup:

- The course slides, datasets, and jupyter notebooks:
  - https://github.com/jeffrey-yau/Pearson-TSA-Training-Beginner.git

- Anaconda's distribution of Python (Jupyter notebook and hundreds of libraries):
  - Anaconda: https://www.anaconda.com/products/individual
  - Package list: https://docs.anaconda.com/anaconda/packages/pkg-docs/

The Forecasting Problem

Introduction to the Forecasting Problem



Forecasting: Problem Formulation

- Forecasting: predicting the **future values** of the series using a model conditional on the **current information set**
- **Current information set** consists of current and past values of the series of interest and perhaps other series

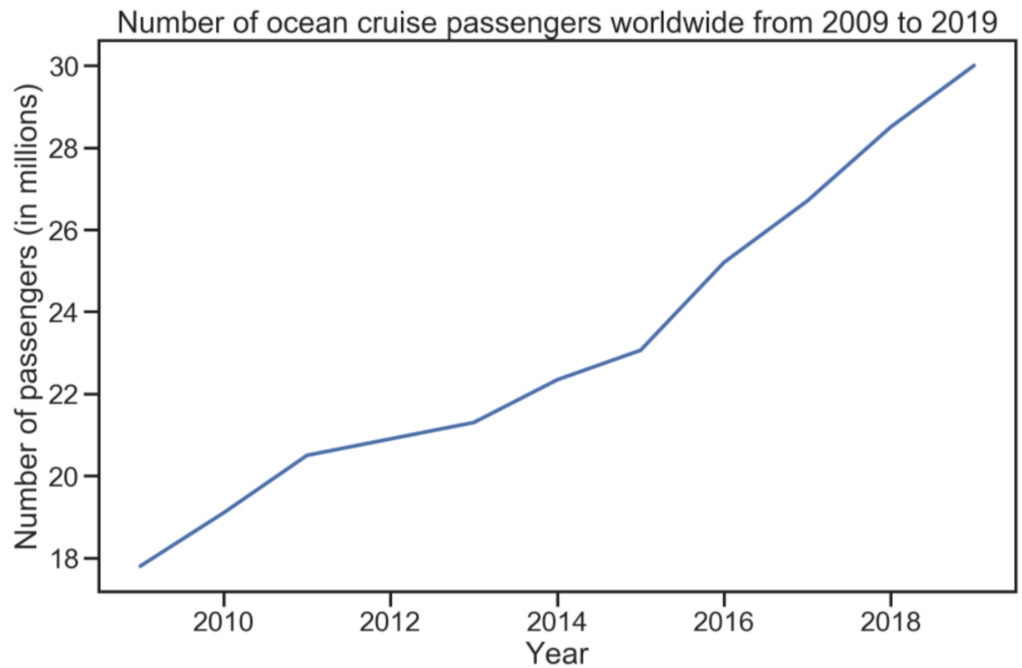
Some Use Cases of TSA

Common Use Cases

- Companies forecast sales
- Government agencies forecast macroeconomic indicators
- Meteorologists forecast various measures of weather
- CMS Projection on National Health Expenditure
- NCES Projections of Education Statistics
- Dynamic resource allocation (e.g., servers)
- Physiological models for health monitoring (e.g., glucose levels in diabetics)

Common Characteristics and Patterns of Time Series

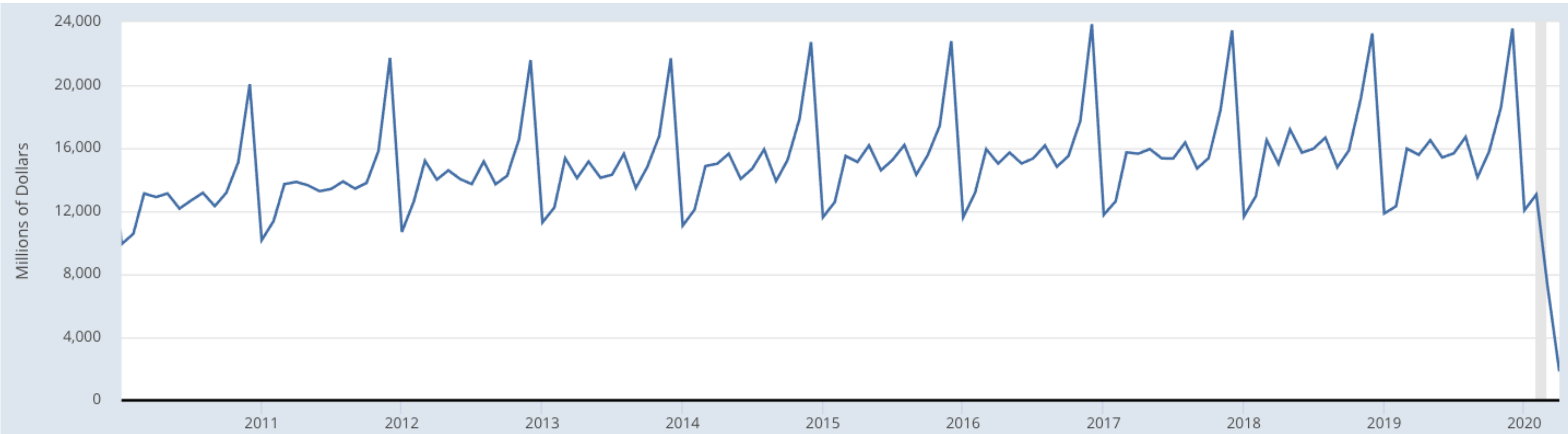
Trend



Number of ocean cruise passengers worldwide from 2009 to 2019
Number of passengers (in millions)
30
28
26
24
22
20
18
2010
2012
2014
2016
2018
Year

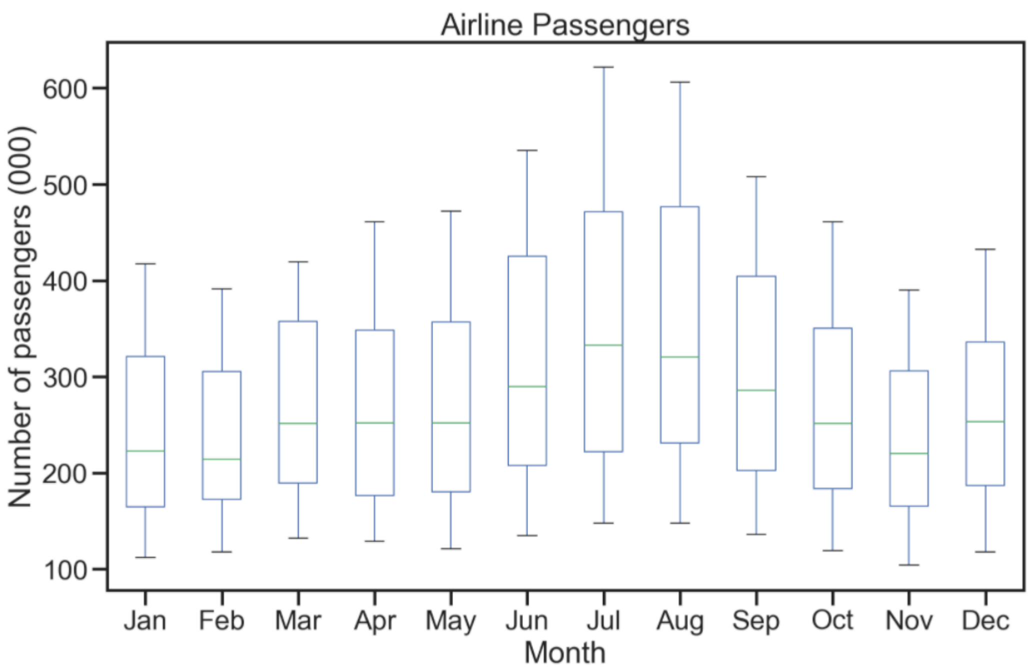
Seasonality

Retail Sales: Clothing Stores



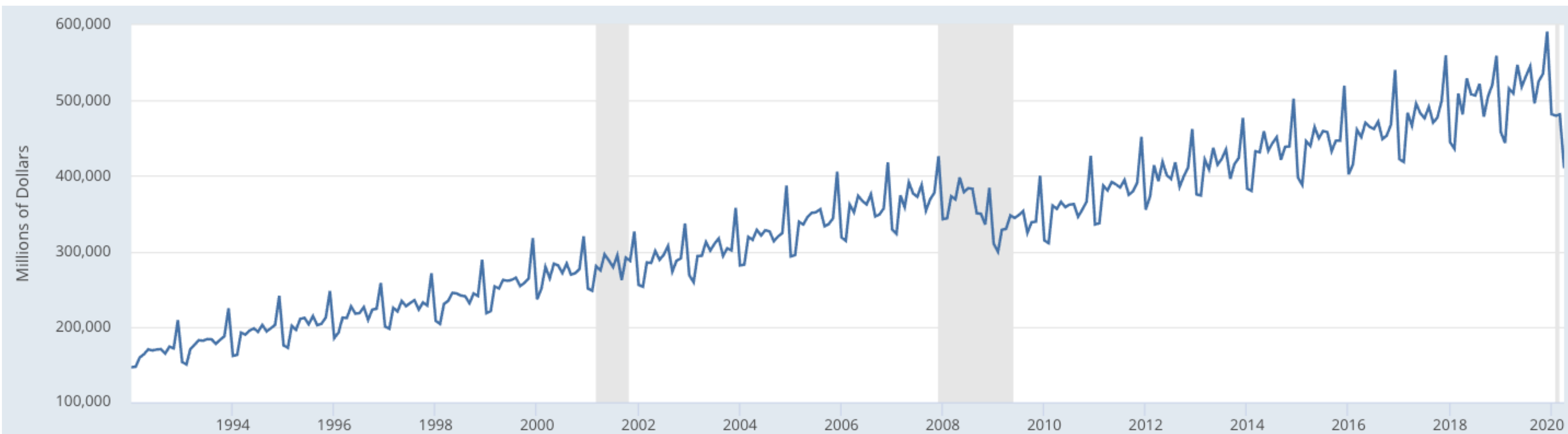
Source: U.S. Census

Seasonality



Trend with Seasonality

Retail Sales: Retail and Food Services



Source: U.S. Census

Demo

Jupyter Notebook

Segment 2: Exploratory Data Analysis and ARMA Modeling

2.1 A brief discussion on basic terminology for time series analysis

2.2 Exploratory Time Series Data Analysis

Exercise

2.3 Mathematical formulation of AR, MA, and ARMA models

2.4 Lag (or backshift) operators

2.5 Properties of the general AR, MA, ARMA models

2.6 ARMA modeling

Exercise

Concepts of Time Series Forecasting

Time Series Forecasting requires Models

Time Series Forecasting Requires Models

$$\hat{y}_{t+H|t} = f(\text{current information set})$$

$$= f(\underbrace{\boxed{y_t, y_{t-1}, \ldots, y_1, \mathbf{X_t}, \mathbf{X_{t-1}}, \ldots, \mathbf{X_1}}})$$

Forecast horizon: *H*

Forecast origin: *t*

A statistical model, a machine learning algorithm, or a rule

Information Set:

$$\Omega_t = \{y_t, y_{t-1}, y_{t-2}, \ldots, y_1, \mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_1\}$$

Takeaways so far …

- Time series forecasting requires a model

Time Series Forecasting requires a Model

$$\begin{aligned}\widehat{y}_{t+H|t} &= f(\text{current information set}) \\ &= f(y_t, y_{t-1}, \ldots, y_1, \mathbf{X_t}, \mathbf{X_{t-1}}, \ldots, \mathbf{X_1})\end{aligned}$$

A Naïve, Rule-based Model:

A model, *f()*, could be as simple as "a rule" - naive model:

The forecast for tomorrow is the observed value today

Persistent forecast

$$\widehat{y}_{t+1|t} = y_t$$

Forecast horizon: *h=1*

Information Set:

$$\begin{aligned}\Omega_t &= \{y_t, y_{t-1}, \ldots, y_1, \mathbf{X_t}, \mathbf{X_{t-1}}, \ldots, \mathbf{X_1}\} \\ &= \{y_t\}\end{aligned}$$

Takeaways so far …

- Time series forecasting requires a model
- The model does not have to be a sophisticated one

"Rolling" Average Model

The forecast for time *t+1* is an average of the observed values from a predefined, past *k* time periods

$$\hat{y}_{t+1|t} = \frac{1}{k} \sum_{s=t-k}^{t} y_s$$

Forecast horizon: *h=1*

Equally weight the last k values

Information Set:

$$\Omega_t = \{y_t, y_{t-1}, \ldots, y_1, \mathbf{X_t}, \mathbf{X_{t-1}}, \ldots, \mathbf{X_1}\}$$

$$\boxed{= \{y_t, \ldots, y_{t-k}\}}$$

Takeaways so far …

- Time series forecasting requires a model
- The model does not have to be a sophisticated one
- Different ”models” use information set differently

Exploratory Data Analysis of Time Series

Refer to Jupyter Notebook

Exercise

EDA

Formulation of Autoregressive Moving Average (ARMA) Process

Autoregressive Moving Average Model (ARMA)

$$\boxed{\hat{y}_{t+H|t} = f(\text{current information set})}$$

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_q \omega_{t-q}$$

lag values from own series

shocks / “error” terms

mean of the series

While this notation of expressing ARMA process is easier to understand, the lag operator notion are being used more often.

Autoregressive Moving Average Model (ARMA)

$$\boxed{\widehat{y}_{t+H|t} = f(\text{current information set})}$$

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + +\theta_q \omega_{t-q}$$

More common notations (that use the lag operators):

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)\widetilde{y}_t = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)\omega_t$$

$$\widetilde{y}_t = y_t - \mu$$

$$\boxed{\phi(L)\widetilde{y}_t = \theta(L)\omega_t}$$

Properties of AR Process

Process	ACF	PACF
AR(p)	Tails off	Cut off after lag p
MA(q)	Cut off after lag q	Tails off
ARMA(p,q)	Tails off	Tails off

Exercise

ARMA model development

Segment 3: ARIMA Model Formulation

3.1 Notion of non-stationarity

3.2 Mathematical formulation of ARIMA models

3.3 The Box-Jenkins Approach to ARIMA Modeling of non-stationary time series

Exercise 3

Stationarity and Invertibility

Formulation of ARIMA Process

ARIMA

A time series y_t follows an **ARIMA(p,d,q)** process if the d^{th} differences of the y_t series is an **ARMA(p,q)** process

$$\underbrace{\phi_p(L)}\underbrace{(1-L)^d}_{d^{th}\text{ differences}}\widetilde{y}_t = \underbrace{\theta_q(L)}\omega_t$$

A polynomial of lag operators associated with the AR(p)

A polynomial of lag operators associated with the MA(q)

ARIMA

A time series y_t follows an **ARIMA(p,d,q)** process if the d^{th} differences of the y_t series is an **ARMA(p,q)** process

$$\phi_p(L)(1-L)^d\tilde{y}_t = \theta_q(L)\omega_t$$

The SARIMA model is specified $(p, d, q) \times (P, D, Q)_s$

$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^D y_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t$$

https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

Exercise

Segment 4: ARIMA Modeling

4.1 Model identification and estimation

4.2 Model diagnostic checking and assumption testing

4.3 Model performance evaluation (in-sample fit)

4.4 Forecasting and forecast evaluation

Exercise 4

General Steps to Build SARIMA Model

1. Ingest the series
2. Split the series for training and validation
3. Conduct exploratory time series data analysis on the training set
4. Determine if the series are stationary
5. Transform the series
6. Build a model on the transformed series
7. Model diagnostic
8. Model selection (based on some pre-defined criterion)
9. Produce forecast using the final, chosen model
10. Inverse-transform the forecast
11. Conduct forecast evaluation

Iterative

Digesting the Model Results

Information about the sample series

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

Series used in model estimation

```
                               SARIMAX Results
==============================================================================
Dep. Variable:               UMCSENT   No. Observations:                  218
Model:             SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:              Thu, 09 Jul 2020   AIC                           1228.430
Time:                      15:14:03   BIC                           1245.329
Sample:                  11-01-2000   HQIC                          1235.256
                       - 12-01-2018
Covariance Type:                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

A specific SARIMAX model being specified

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

Date and time when the estimation was performed

```
                               SARIMAX Results
==============================================================================
Dep. Variable:               UMCSENT   No. Observations:                  218
Model:              SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:               Thu, 09 Jul 2020   AIC                           1228.430
Time:                       15:14:03   BIC                           1245.329
Sample:                   11-01-2000   HQIC                          1235.256
                        - 12-01-2018
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

The sample time period

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Model “fit” measures

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Number of observation contained in the sample

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Log likelihood

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

AIC = -2log(L) + 2(p+q+k+1)

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

$$\mathrm{BIC} = \mathrm{AIC} + [\log(T) - 2](p + q + k + 1)$$

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Hannan-Quinn Information Criterion

$$HQIC = -2 \cdot log(\mathcal{L}_{max}) + 2 \cdot k \cdot log(log(n))$$

Digesting the Model Results

Model estimation results and associated statistics

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

Estimated coefficients

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

Standard errors of estimated coefficients

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

z-statistics (coeff/SE)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

p-value

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

95% C.I.

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Model assumption test statistics

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Ljung-Box test for Serial Correlation

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\rho^2(j)}{n-j}$$

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Test for heteroskedasticity of standardized residuals

$$H(h) = \sum_{t=T-h+1}^{T} \tilde{v}_t^2 \Bigg/ \sum_{t=d+1}^{d+1+h} \tilde{v}_t^2$$

Digesting the Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                UMCSENT   No. Observations:                  218
Model:               SARIMAX(2, 1, 2)   Log Likelihood                -609.215
Date:                Thu, 09 Jul 2020   AIC                           1228.430
Time:                        15:14:03   BIC                           1245.329
Sample:                    11-01-2000   HQIC                          1235.256
                         - 12-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0315      0.316      3.263      0.001       0.412       1.651
ar.L2         -0.5276      0.230     -2.298      0.022      -0.978      -0.078
ma.L1         -1.1181      0.329     -3.402      0.001      -1.762      -0.474
ma.L2          0.4405      0.279      1.579      0.114      -0.106       0.987
sigma2        16.0571      1.533     10.472      0.000      13.052      19.062
===================================================================================
Ljung-Box (Q):                       36.54   Jarque-Bera (JB):                 1.82
Prob(Q):                              0.63   Prob(JB):                         0.40
Heteroskedasticity (H):               0.43   Skew:                            -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.21
===================================================================================
```

Jarque-Bera test for normality

$$JB = n(S^2/6 + (K-3)^2/24)$$

Segment 5: Seasonal ARIMA Modeling

5.1 Understanding seasonality and examination of seasonal time series

5.2 Mathematical formulation of Seasonal ARIMA (SARIMA) models

5.3 Building a seasonal ARIMA model for forecasting

Exercise 5

Seasonal ARIMA Process Formulation

The **SARIMA(p,d,q) x (P,D,Q)$_s$** model is specified as

$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^D y_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t$$

```
statsmodels.tsa.statespace.sarimax.SARIMAX(endog, exog=None, order=(1, 0, 0), seasonal_order=(0, 0, 0, 0), trend=None, measurement_error=False, time_varying_regression=False, mle_regression=True, simple_differencing=False, enforce_stationarity=True, enforce_invertibility=True, hamilton_representation=False, **kwargs)
```

https://www.statsmodels.org/0.8.0/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html#statsmodels.tsa.statespace.sarimax.SARIMAX

Segment 6: Concluding Remarks

6.1 Summary of today's training

6.2 Course wrap-up and next steps, and where to go from here

Recap

- EDA for Time Series (Time Series Plot, ACF, PACF)
- AR
- MA
- ARMA
- ARIMA
- Seasonal ARIMA (SARIMA)

What’s next and where to go from here?

- Review the materials covered today
- Perhaps watching the course a few more times, pausing when needed to ensure you understand the materials
- Remember that today’s course is recorded
- Experiment with your own data, and perhaps simulating some data

References

- Time Series Analysis: Univariate and Multivariate Methods (Classic Version), 2nd Edition, William W.S. Wei.
- Other time Series Analysis and Forecasting training courses - by Jeffrey Yau (work-in-progress)
- Time Series Forecasting for Data Scientists by Jeffrey Yau (work-in-progress)

Review of relevant materials

- **Even You Can Learn Statistics and Analytics**, 3E, Levine and Stephan, 9780133382662
- **Programming Skills for Data Science**, Freeman & Ross, 9780135133101
- ***Machine Learning with Python for Everyone*****,** Mark Fenner. 9780136592259
- **Introduction to Mathematical Statistics and Its Applications**, 6th Edition, Richard J. Larsen, and Morris L. Marx. 9780134114279

Thank You!