Module 13 - Statsmodels Time Series Complete

1. **Time Series Data**: This refers to data that is collected and organized in chronological order, typically over a period of time. This type of data is common in finance (stock prices, economic indicators), environmental monitoring (temperature, pollution levels), and various other fields.
2. **Pandas**: This is a popular Python library specifically designed for working with and analyzing data. It provides data structures like DataFrames that are highly efficient for manipulating and exploring time series data.
3. **Statsmodels**: Statsmodels is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration.1
4. **Autoregression**: The nutshell - Uses observations from previous time steps as input to a regression equation to predict the value at the next time step. Autoregression2 is a time series modeling technique that predicts future values based on past values of the same series. It's like saying, "knowing what happened in the past can help us understand what might happen in the future."
5. **Lagged Value**: A lagged value is simply a previous value in the series.
6. **Order of Autoregression (p)**: The number of lagged values used in the model. For example, an autoregressive model of order 2 uses the two previous values to predict the current value.
7. **Coefficient**: Each lagged value is multiplied by a coefficient that determines its influence on the prediction. These coefficients are estimated from the data.
8. **Error Term**: The model also includes an error term to account for random fluctuations and unpredictable factors.
9. **ETS**: ETS decomposition is a time series analysis method that decomposes a time series into its error, trend, and seasonality components. It is used to understand the underlying patterns and dynamics of a time series and can be helpful for forecasting and anomaly detection.
10. **Additive Models**: Used when trend is linear and/or seasonal variations are constant.
11. **Multiplicative**: Used when the trend is non-linear (exponential) and/or seasonality varies proportionally to the level of the series.
12. **Moving Average**: The moving average is a simple but powerful technique used to smooth out fluctuations in time series data and highlight underlying trends. It works by calculating the average of a specified number of consecutive data points, creating a new series of averages.
13. **EWMA**: EWMA stands for Exponentially Weighted Moving Average. It's a way to smooth out data by giving more weight to recent observations and less weight to older observations. This makes it particularly useful for analyzing time series data where the most recent data points are often the most relevant.
14. **Holt-Winters**: EWMA has just one smoothing factor and doesn't account for trend, seasonality, etc. HW offers three smoothing factors for level, trend, and season and can adjust for divisions per cycle (L). Offers Single, Double (Holt and Trend), and Triple (Holt-Winters and Seasonality) Exponential Smoothing.
15. **Stationary Time Series**: A stationary time series is one whose statistical properties, such as mean, variance, and autocorrelation, remain constant3 over time. This means that the data looks roughly the same no matter what period you examine.
16. **ACF**: ACF stands for Autocorrelation Function. It's a tool used in time series analysis to measure the correlation between a time series and lagged versions of itself. In simpler terms, it helps you understand how a data point at a particular time is related to data points that came before it.
17. **PACF**: PACF stands for Partial Autocorrelation Function. It's another important tool in time series analysis, and it's closely related to the ACF (Autocorrelation Function). However, there's a key difference: it measures the correlation between a time series and its lagged versions after removing the effect of intermediate lags.
18. **ARIMA**: ARIMA stands for Autoregressive Integrated Moving Average. It's a powerful statistical model used for analyzing and forecasting time series data. The AR part of the model uses past values to predict future values. The I part deals with making the time series stationary. The MA part of the model uses past forecast errors to predict future values.
19. **SARIMAX**: SARIMA (Seasonal ARIMA) is an extension of ARIMA that explicitly includes seasonal components. SARIMAX (Seasonal ARIMA with exogenous regressors) is an extension of SARIMA that allows you to include external variables (exogenous variables) in your model.
20. **Forecasting**: Predicting future values based on historical trends.