**Question 4 - Part I (35 pts): Univariate Time Series Modeling**

You are tasked with developing a **univariate SARIMA model** to forecast future demand.

1. **Data Preparation and Exploratory Analysis**

* Explain how you detect any trend, seasonality, or potential structural breaks in the series?

Trend Detection: By plotting the time series data the trend appears as a long-term increase or decrease in series overtime. This can be confirmed by doing a visual inspection if the overall demand is moving upwards or downwards.

Seasonality: It is detected by monthly, quarterly, etc. by observing the time series plot repeats the patterns at fixed intervals. The pattern cycle is recurring over time. To help separate the seasonal pattern clearer by using the classical decomposition.

Structural Breaks Detection: It can be detected by sudden shift in the plotted area. It can be detected by the changes in the variance of the series.

Code Snippet: Using Visual Studio App

A screenshot of a computer

AI-generated content may be incorrect.

Output:

A graph showing a line

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

* Which tests or approaches would you use to determine if differencing (nonseasonal and/or seasonal) is required?

Nonseasonal Differencing (d): To determine if needed, performing the Augmented Dickery-Fuller (ADF) test will help. If the test returns to a high p-value (above 0.05), it signals that the series is non-stationary and nonseasonal difference (d=1) may required to remove the trend.

Seasonal Differencing (D): To determine if needed, it is evaluated by inspecting the autocorrelation function (ACF) plot. If there is a seasonal lag by showing the significant spikes in the monthly data, it shows that the importance of setting (D=1) might be needed to remove the seasonal effects.

1. **Model Identification and Specification**

* Discuss the roles of the ACF, PACF, and stationarity tests (e.g., Augmented Dickey-Fuller) in guiding your model selection.

Role Stationary Test (d and D): To determine the orders for the AR (autoregressive) and MA (moving average) components, firstly it is important to ensure that the series is stationary. The ACF test will be helpful to assess it. If ADF test results in non-stationary, then set up d>=1. For the seasonal data if the autocorrelation is significant at seasonal lags, then set up D=1.

Roles of ACF and PACF (p and q): The ACF plot is useful to identify the order of the MA component (q). If the sharp cutoff is experienced, it suggests that the number of MA terms required. The PACF plot is helpful to determine the AR order (p). After a certain lag with the sharp cutoff suggesting the number of AR terms.

* Explain how you would choose the orders (p,d,q)(p, d, q) and seasonal orders (P,D,Q)s(P, D, Q)\_s.

Seasonal orders (P and Q): Once the process of addressing the nonseasonal dynamics, it is important to inspect the ACF and PACF plots for seasonal lags.

Seasonal ACF: In the result I can see the significant spikes at seasonal lags 12, 24 it suggests including the seasonal MA term (Q).

Seasonal PACF: The significant spikes in the PACF indicate that need for the seasonal AR terms (P). The seasonal differencing order (D) is chosen based on whether the seasonal effects are strong and need to be removed to achieve stationarity.

Code Snippets:

A screenshot of a computer program

AI-generated content may be incorrect.

Output:

A graph of a graph of a graph

AI-generated content may be incorrect.

As a result of the ADF test to check whether the time series is stationary. The ADF Statistic: 0.626963, the value is not negative as it is positive indicating that the test does not provide evidence against the present of unit root. The p-value: 0.988262, is very high among the thresholds like 0.05 so in the ADF test states the series has a unit root (i.e., non-stationary).

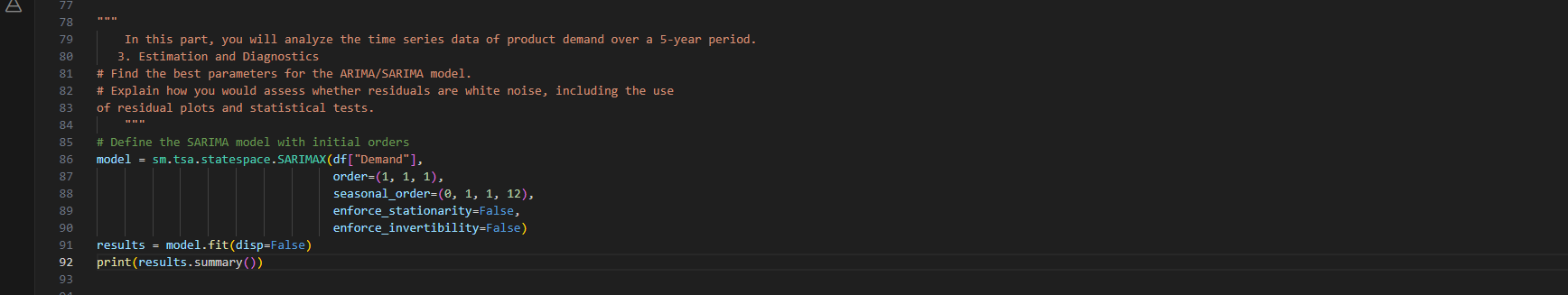
Based on the above two values it is very clear that the series likely exhibits trends or other non-constant statistical properties over time.

1. **Estimation and Diagnostics**

* Find the best parameters for the ARIMA/SARIMA model.

To find the best model parameters the first step is parameter estimation, it is generally done using the maximum likelihood estimation (MLE) or conditional least squares, to best fit the model historical data. The model selection is done based on trying several models and comparing the performance using the AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion), the one with the lowest is preferred to fit with model complexity. Initially, setting up the SARIMA model with orders (1,1,1) for the nonseasonal part and (0,1,1,12) for the seasonal component. It is chosen to refine model selection procedures.

Code Snippet:



Output:

A black screen with white text

AI-generated content may be incorrect.

* Explain how you would assess whether residuals are white noise, including the use of residual plots and statistical tests.

After the model fitting to validate the residuals differences between the observed and model-predicted values act like white noise. In a good model, residuals have no autocorrelation, randomly scattered around zero, plot the residuals over time and if there is no systemic pattern hovering around zero, it is a good sign. ACF plotting helps us to verify the residuals and all the spikes should be within the confidence bounds. The statistical test is done using the Ljung-Box test to test the null hypothesis to check the white noise.

Code Snippet:

A black rectangular object with a black border

AI-generated content may be incorrect.

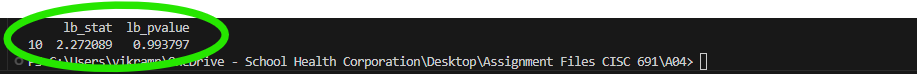
Output:

A graph with a line

AI-generated content may be incorrect.

A graph with blue dots and numbers

AI-generated content may be incorrect.



As a result we can see the p-value is 0.993797, suggesting that there is no significant autocorrelation in the residuals at the 10th lag. It suggests that the SARIMA model fits the data well, the residuals are random, it confirms no further adjustment to the model is necessary and the residuals are essentially white noise.

1. **Testing and Model Validation**

* Generate forecasts for the 5th year, including prediction intervals. Show your results in a plot and give the values also in a table.

Now for the validation part the data is split into the training set (for the first four years) and a test (the 5th year). The training data model is re-estimated. The fitted training set data model generates the forecast for the 5th year. The forecasts should include the 95% confidence levels providing the range of plausible values for the future demand.

Code Snippet:

A screenshot of a computer program

AI-generated content may be incorrect.

* Compare the results with the actuals or test data by plotting the predicted values and the test data (actuals) with the time series.

The forecasted values with the prediction intervals on a plot showing the training and actual test data. It helps understand how well the model predicts the future demand. A table generated to list the forecasted demand values along with the lower and upper bounds of the predicted demand intervals.

Code Snippet:

A screenshot of a computer program

AI-generated content may be incorrect.

Output:

A graph with a line and a red arrow

AI-generated content may be incorrect.

The forecasted demand values show the predicted future values of the product demand over the forecast period. Ex: Jan 2024 the model predicted the demand will be around 153.68 units. The lower bound is 145.99 for Jan 2024, suggesting the 95% confidence level showing the true demand could be as low as 145.99 units. The upper bound is 161.37 for Jan 2024, suggesting the 95% confidence level showing the true demand could go up to 161.37 units. Always the narrow interval suggests that more confidence in prediction.

* Describe which performance metrics (e.g., RMSE, MAE, MAPE) you would use to evaluate the model, and why. Give the results of your analysis.

For this purpose had to install the scikit-learn libraries. By calculating the error metrics such as RMSE (Root Mean Squared Error): It is sensitive to large errors and provides a measure of average error magnitude. MAE (Mean Absolute Error): It helps measure the average absolute difference between predicted and actual values. MAPE (Mean Absolute Percentage Error): It shows error as percentage useful for understanding error in relative values. To note is that for these metrics lower values indicate better performance.

Code Snippet:

A screenshot of a computer

AI-generated content may be incorrect.

Output:

RMSE (3.50):

On average, the predictions deviate from the actual values by 3.50 units. RMSE penalizes larger errors more heavily due to squaring the residuals.

MAE (2.80):

On average, the model predictions are off by 2.80 units. It does not penalize large errors as much as RMSE.

MAPE (1.77%):

The model predictions are on average 1.77% off from the actual values, which indicated good accuracy.

Overall, close to actual values both in absolute terms (MAE, RMSE) and relative terms (MAPE)

**Question 4 - Part II (35 pts): Incorporating Exogenous Variables**

Now assume you wish to incorporate the external factors that potentially influence product demand such as advertising spend and competitor’s price in your model using an ARIMAX or SARIMAX framework.

1. **Model Formulation with Exogenous Regressors**

* How would you determine whether each exogenous variable (and its lags) should be included in the model?

By examining the correlation between product demand and potential exogenous variables (advertising spend, competitors price) plotting their time series. It helps reveal the changes in demand. Using the Granger causality test to statistically access the lagged values of an exogenous variable is useful for the demand forecasting. If the cross- correlation analysis indicated the exogenous variable influence the demand at a lag, including the lags as additional regressors in the model.

Code Snippet:

A computer screen shot of text

AI-generated content may be incorrect.

Output:

A graph with blue dots and lines

AI-generated content may be incorrect.

* Present a general form of the ARIMAX/SARIMAX equation, highlighting how external variables enter the model.

is the product demand at time *t*.

and are the nonseasonal AR and MA polynomials, with orders *p* and *q*, respectively.

and are the seasonal AR and MA polynomials, with seasonal orders P and Q, and a seasonal period *s*.

*d* and D are the orders of nonseasonal and seasonal differencing.

is a constant term (intercept)

is the white noise error term

represents the exogenous variables and their lagged values up to lag L, with corresponding coefficients

This equation reflects how the dependent variable (demand) is influenced by its past value (AR), moving averages (MA), seasonal components (SAR), and exogenous variables ().

1. **Model Parameter Estimation, Diagnostic and stability**

* Explain the challenges posed by multicollinearity among explanatory variables and how you would diagnose or mitigate them.

Multicollinearity can cause the spike in the variance of the coefficient estimates when the multiple exogenous variables are highly correlated. For the diagnosis, calculating the VIFs for the regressors; values above 5 or 10 indicate problematic multicollinearity. Examining the pairwise correlation matrix to identify the highly correlated predictors. Mitigating the challenges by removing or combining the highly correlated variables, dimensions reductions and regularization by using methods like ridge regression when many correlated variables must be included.

Code Snippet:

A screenshot of a computer

AI-generated content may be incorrect.

Output:

The significance of the output values, the **low VIF values** for both AdSpend and CompetitorPrice (around 1.03) suggest that there is no multicollinearity between these two features and other predictors in the model. So these two features are to provide reliable and stable coefficient estimates in the model. The **very high VIF for the constant term** is typical and does not indicate the problem with the model.

* Find the best ARIMAX/SARIMAX model using the test data.

To fit the model extend the SARIMA model by including exogenous variables. By using a similar procedure as before but now pass the exogenous data to the model. By varying the orders and lags for the exogenous variables, evaluate different specifications using information criteria such as AIC or BIC. So the best model will minimize these criteria.

Code Snippet:

A black screen with text

AI-generated content may be incorrect.

Output:

A black screen with white text

AI-generated content may be incorrect.

The ADF Statistics: 0.626963 and p-value: 0.988262, suggests that it fails to reject the null hypothesis that the time series is non-stationary. Therefore, differencing may be required before further modelling.

The Ljung-Box statistics (2.27) are not large enough to indicate significant autocorrelation.

The high p-value (0.993797) suggests that there is no significant autocorrelation in the residuals at lag 10. The model residuals appear to be uncorrelated at lag 10, suggesting that the model has likely captured all the significant patterns in the data.

* Describe how you would verify that adding exogenous variables genuinely improves the model (e.g., comparing information criteria, checking residual plots).

By comparing the AIC/BIC values of the ARIMAX/SARIMAX model with the univariate SARIMA model. The lower value indicates the better balance between the model fit and the complexity. Residual analysis is done for the extended model are still white noise. Now by comparing the forecasting performance using metrics such as RMSE, MAE, and MAPE on the test data. The improved metric suggests that exogenous variables enhance the model predictive power.

Code Snippet:

A green and black background

AI-generated content may be incorrect.

Outputs:

A graph with blue lines

AI-generated content may be incorrect.

A graph with blue dots and numbers

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

The AIC (Akaike Information Criterion): 1275.19, it suggests that ARIMAX model provides a good fit with both accuracy and complexity. The AIC is a measure of the relative quality of statistical model for the given dataset.

Where:

*k* is the number of model parameters (including coefficients and intercepts),

*L* is the likelihood of the model.

The BIC (Bayesian Information Criterion): 1284.16, it suggests that it is slightly higher reflecting a stronger penalty on complexity.

Where:

*n* is the number of data points,

*k* is the number of model parameters,

*L* is the likelihood of the model.

By comparing to other models, a lower AIC and BIC suggest a better fit. But the choice between the two highly depends on the priority either on predictive accuracy (AIC) or simplicity (BIC).

1. **Model Parameter Estimation, Diagnostic and stability**

* Generate the forecast for the 5th year and compare it with the test data. Generate the performance metrics and plots for this model.

As the explanation given earlier, we do the data splitting. To fit the ARIMAX/SARIMAX model the re-estimate the model using the training set and exogenous variables for the training period. For test period, supplying the corresponding exogenous variables data to forecast can incorporate the effects. By using the fitted model forecasting is generated for the test period which includes point predictions and 95% predictions intervals that account for uncertainty.

Code Snippet:

A screenshot of a computer

AI-generated content may be incorrect.

Output: SARIMAX Demand

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

The output values suggest that AdSpend is the only statistically significant predictor with the p-value of 0.001 showing the strong relationship with demand. The CompetitivePrice and the AR and MA terms do not show significant influence on the demand. Diagnostic tests for the residual autocorrelation, normality and heteroskedasticity all indicate that the residuals are well behaved (no significant issues). The model fit indicators AIC, BIC and HQIC suggest a reasonable fit, but still room for improvement in modeling the seasonal components and other predictors.

Plotting and analyzing the performance metrics: By plotting the training data, actual demand from the 5th year using the test data, and the forecasted demand with predicted intervals. By calculating performance metrics like RMSE, MAE and MAPE by using the ARIMAX and the univariate SARIMA model. By comparison between these models the model with lower error metrics indicates that incorporating exogenous variables improves forecasting accuracy. Checking the residuals of the ARIMAX model is white noise using residual plots and Ljung-Box Test. It also provides insight into how external factors AdSpend and CompetitorPrice affect demand helpful for decision making.

Code Snippet:

A screen shot of a computer

AI-generated content may be incorrect.

Outputs:

A graph with a line drawn up

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

The output significance for the forecasted values for the next few months indicates an increase in demand, with narrow confidence intervals. Ex: 2024-01-01, demand is expected to be 157.10 with the 95% confidence interval of (151.41, 162.79) so the price must be between these ranges. The evaluation metrics RMSE (3.31) suggests that model is accurate, with errors around 3.31 units. The MAE (2.21) shows the model’s average error is about 2.98 units. The MAPE (1.90%) indicate model predictions are on average 1.90% off, showing a strong level of accuracy.

So, the ARIMAX model is a good fit to provide reliable forecasts with low uncertainty.

1. **Forecasting and Interpretation**

* If you have (or can predict) the future values of these exogenous variables, describe how you would generate and interpret out-of-sample forecasts.

Having the future values of these exogenous variables can help generate forecasts by using these values into the fitted ARIMAX model. By inputting the AdSpend and CompetitorPrice exogenous variables into the models forecasting method to predict the demand. The prediction intervals will help quantify the uncertainty of the forecasting.

Code Snippet:

A screenshot of a computer program

AI-generated content may be incorrect.

Outputs:

A computer screen shot of a computer screen

AI-generated content may be incorrect.

* Offer guidelines on presenting model results to both technical and non-technical audiences, focusing on how changes in each exogenous variable affect demand.

Technical Audiences: The ARIMAX forecast model for the 5th year point estimates along with 95% prediction intervals. The performance metrics RMSE, MAE and MAPE have lower error values suggest that the model’s forecasts are very close to the observed values suggesting high accuracy. The residual plots and autocorrelation functions confirmed that the residuals behave like white noise, as the model captured the essential patterns in data. The incorporation of the exogenous variables such as AdSpend and CompetitorPrice through the ARIMAX framework led to better performance compared to a univariate SARIMA model. The statistical significance of the exogenous regressors, combined with improvements in AIC/BIC and error metrics support the inclusion of these variables.

Non-Technical Audiences: The forecasting model predicted that in January 2024, the demand for the product will be around 157 units. The range likely between 151 and 163 units. For February the forecast is around 159 units with similar confidence levels. The model’s error is about 3 units on average and the error percentage of just 1.90% means that model can be trusted and forecast to be very close to reality. These forecasts help us be prepared to tackle any challenges for inventory, production, or marketing. If the demand is forecasted to rise more next month the business can adjust the production schedules accordingly. The charts provided and tables clearly show the forecasted demand along with ranges capturing the uncertainty. It should help the decision makers to grasp the trend very quickly. The model also uses external factors like AdSpend and CompetitorPrice. To plan to increase the advertising budget the model can help predict if there might be a positive effect on demand or vice versa with CompetitorPrice.

**Question 4 - Part III (30 pts): Research other forecasting models**

Describe an alternative technique for forecasting product demand and apply it to your dataset. Compare and contrast its results with your previous findings. Which model performs best for your dataset, and why

Alternative methods I considered are the Facebook Prophet and Holt-Winters. In (Abdulla I. Almazrouee, December 2020), they utilized the prophet model with single and multiple regressors to forecast the electricity generation in Kuwait from 2020 to 2030. In addition, (Abdulla I. Almazrouee, December 2020) used the multiple seasonality Holt-Winters models for comparative analysis. In (Vidhya, 2020) the impact of these two models is well explained with results.

Based on the research paper’s findings I have decided to use Holt-Winters as an alternative technique for forecasting product demand. As the model is simpler to implement as it directly decomposes the time series into level, trend and seasonal components. When the exogenous facts are unavailable and seasonal trend patterns dominate, this model can help generate competitive forecasts.

Using Holt-Winters to the Dataset:

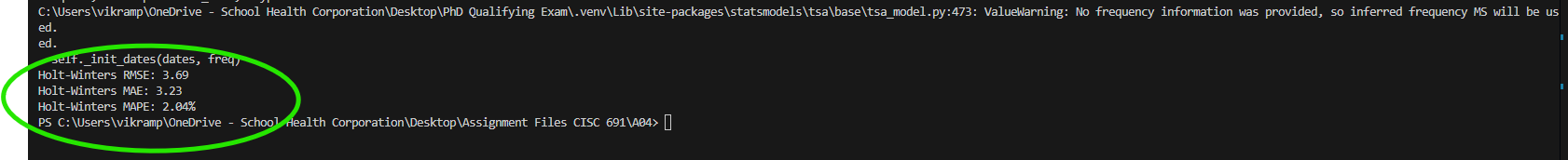
Below code explains the Holt-Winter exponential smoothing model fit on the training data to forecast the 5th year to compute the forecast accuracy metrics.

Code Snippet:

A screenshot of a computer program

AI-generated content may be incorrect.

Output:



A graph with a line drawn on it

AI-generated content may be incorrect.

Comparing Holt-Winters with the ARIMAX Model:

When comparing the values of RMSE, MAE and MAPE of the ARIMAX model to the Holt-Winters RMSE (3.69), MAE (3.23) and MAPE (2.04%) these values indicate that the ARIMAX model forecasts are slightly more accurate. The Accuracy of ARIMAX model with lower evaluation metrics values outperforms the Holt-Winter. The reason is that ARIMAX model ability to incorporate the external factors like AdSpend and CompetitorPrice that causing the significant impact on the demand.

Best Model for the dataset:

For the given dataset the external factors play a major role in driving the demand the ARIMAX model performs best. Its ability to handle the complex exogenous variables leads to improved forecasting accuracy compared ti the pure time series approach like Holt-Winters. If the external data is not provided in the dataset, Holt-Winters would be a good robust alternative due to its simplicity.

**References:**

* Abdulla I. Almazrouee, A. M. (December 2020). Forecasting of Electrical Generation Using Prophet and Multiple Seasonality of Holt-Winters Models: A Case Study of Kuwait. *Applied Sciences*.
* Bruce H. Andrews, M. D. (August 2013). Building ARIMA and ARIMAX Models for Predicting Long-Term Disability Benefit Application Rates in the Public/Private Sectors. *Society of Actuaries Health Section* .
* Hyndman, R. J. (3rd Edition, 2022). *Forecasting: Principle and Practice.* https://otexts.com/fpp3.
* Ramalingareddy Vijaya Kumari, G. R. (November 2020). An application of ARIMAX model for forecasting of castor production in India. *Journal of Oilseeds Research* .
* Vidhya, A. (2020, February 14). *Time Series Forecasting with SARIMA, Holt-Winters’, and Prophet*. Retrieved from Medium: https://medium.com/analytics-vidhya/time-series-forecasting-with-sarima-holt-winters-and-prophet-d6539fbdebe6
* Some parts of the code were developed with the reference of ChatGPT (OpenAI, 2025) to understand the importance of each test method and adapted to suit the project requirements.
* ISEM 700: Smart Enterprises and Strategic Intelligence materials and assignments.