MISO Energy Demand

Cody Rorick and Dennis Peter
MA 5781
February 2025

Abstract

To maintain the grid's dependability, precise energy consumption forecasts within the MISO power grid footprint is essential. This study examines the MISO monthly aggregated energy data to determine which model best fits the data observed. The important questions we are interested in addressing include: does the data exhibit a seasonal trend, is there growth in load over time after controlling for seasonality, and what can we anticipate for future load? This study answers these questions by diving into the consumption patterns through the use of statistical tests, model diagnostics, and forecasting techniques.

Table of Contents

BACKGROUND

Forecasting energy usage is essential for utilities to efficiently plan and distribute resources. Demand for the MISO power grid, which covers a large area, varies seasonally, with peaks usually occurring during months of severe weather. Seasonal trends can be more clearly identified because to the dataset's hourly energy usage statistics that are combined into monthly totals. The goal of the study is to evaluate the long-term trends in energy use and choose the best seasonal model for describing consumption patterns.

DATASET DESCRIPTION

The time series data that we will analyze is energy consumption within the MISO power grid footprint. The data we have selected is hourly, but we have aggregated it into monthly for ease of analysis.

Figure 1: Monthly Aggregated Data

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2021		37140635	50739821	47462594	50868175	60108108	63731708	65247577	53430095	52048311	50938840	52666334
2022	60933204	53334078	52814996	49073624	54090926	60244719	63338999	62739803	54269112	49583464	51353652	56141686
2023	55980478	49771136	52804770	47556816	52058058	58248889	44575332	40740431	51122027	48369032	52157868	56104631
2024	13326215	50363418	47910045	53060146	59166158	63640439	64004503	54807750	51616679	50319468	54412905	

Figure 1 shows the MISO megawatt load that has been aggregated from hourly data into a monthly sum. It is apparent that the load varies from month to month. Extreme weather months, particularly winter and summer, are showing up as local minimums and maximums for energy usage.

METHOD

For the analysis, the dataset hourly energy consumption totals were organized as a time series. The data was then combined from hourly into monthly totals in order to have it ready for analysis. Once the data was aggregated, we created time series plots for visual inspection. We then fit and compared different models through further visual inspections of their forecasts against the actual values, along with comparing their AIC and BIC values. Once a model was chosen, we analyzed residuals to determine if certain model assumptions were met. After model assumptions were determined to be met, the stationarity of the residuals were analyzed through the Augmented Dickey-Fuller (ADF) test. To evaluate time dependency, the runs test was performed and analyzed. To further evaluate time and lag dependencies, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were analyzed.

DETERMINISTIC MODEL FITTING

To start the analysis, we created a time series plot that shows the MISO power grid footprint's monthly total energy usage from 2021 to 2025. Seasonal patterns and anomalies can be examined by labeling each data point with the relevant month.

Figure 2: Monthly Aggregated MISO Data plot

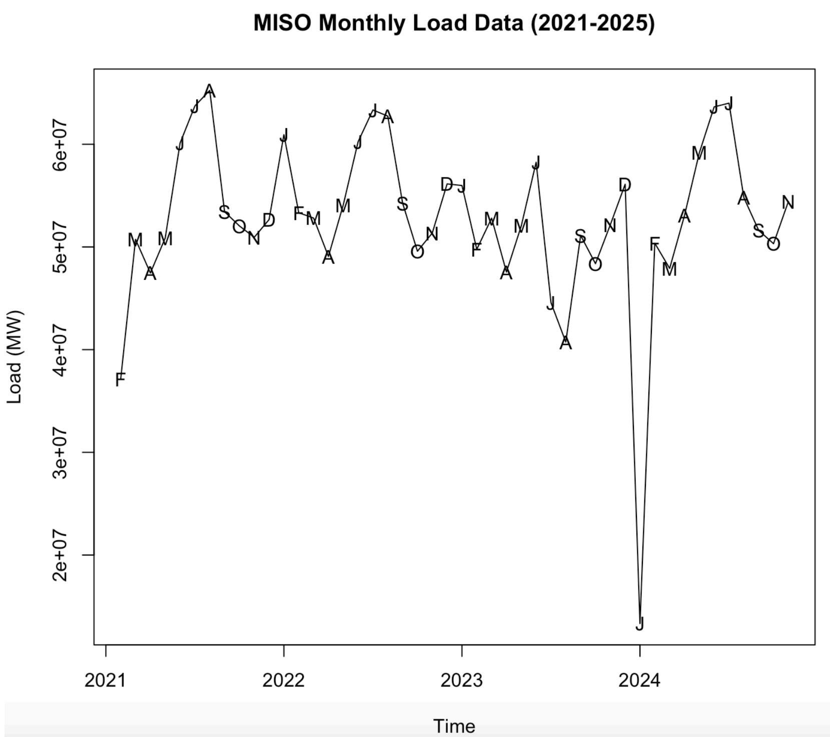


Figure 2 allows us to observe the seasonality of the data, a seasonal component should definitely be included in a model. One note on this time series plot, further examination would be required to determine if the sharp decline in early 2024 was a singular incident or a symptom of a more widespread systemic problem, more research is necessary.

To capture the seasonality observed in the time series plot, two models were fit: a sinusoidal model and a seasonal means model. These models were compared against the actual data in a time series plot below.

Figure 3: Observed vs fitted values

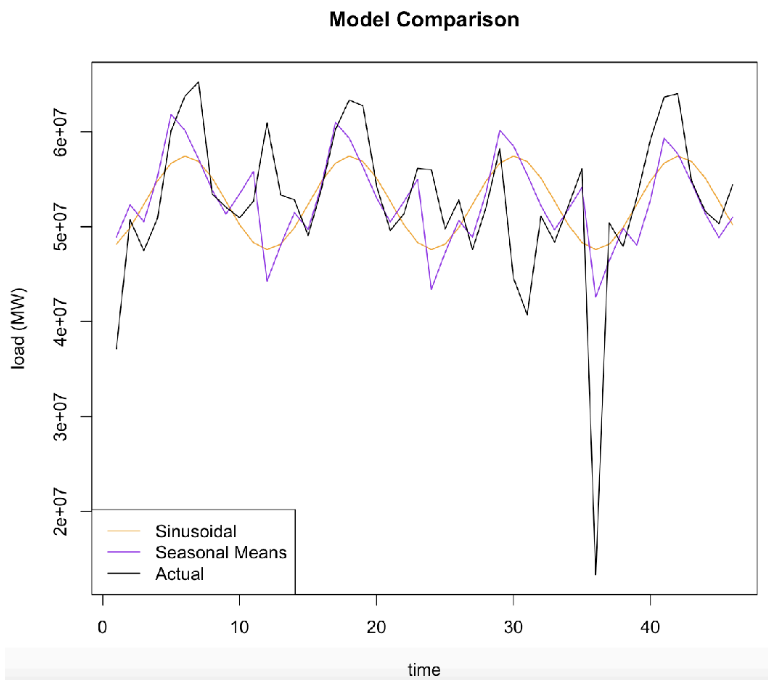


Figure 5 presents how the seasonal variations in real energy usage are attempted to be approximated by the seasonal means model and the sinusoidal model. Although the sinusoidal somewhat understates some peaks and valleys, it accurately depicts the overall seasonal structure. This implies that although the model does a good job of capturing periodicity, they might not be as adaptable when it comes to short-term abnormalities.

These two seasonal models were then compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

Table 1: Sinusoidal model AIC and BIC Comparison

AIC	1596.159
BIC	1603.474

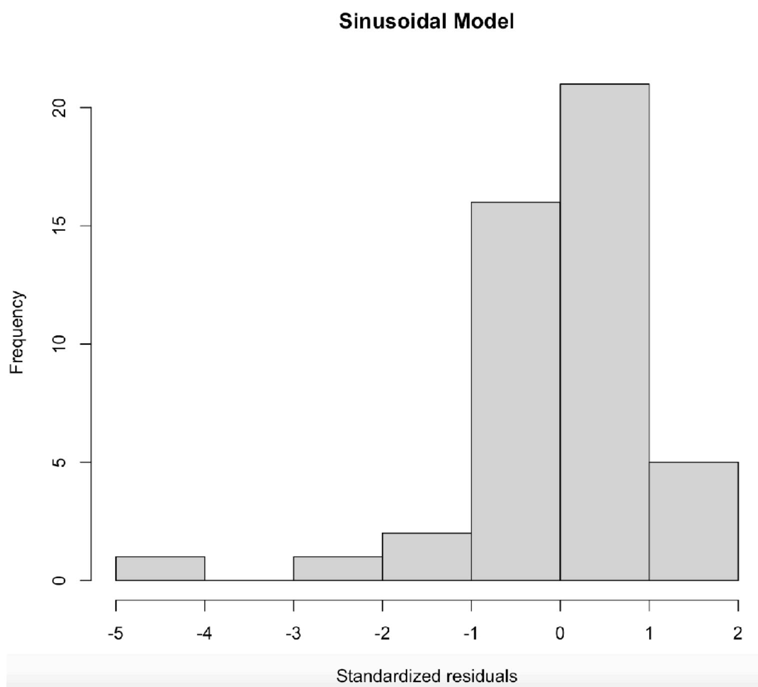
Table 2: Seasonal mean model AIC and BIC Comparison

AIC	1608.482
BIC	1634.083

The performance of two deterministic models—the Seasonal Mean Model, which predicts load using historical monthly averages, and the Sinusoidal Model, which aims to fit a periodic function to the data—is contrasted in Tables 1 and 2. Tables 1 and 2 shows that, in comparison to the seasonal mean model, the sinusoidal model has a lower AIC and BIC. A better model fit is indicated by lower AIC and BIC values. These measurements suggest that the sinusoidal model demonstrated a superior fit based on lower AIC and BIC values, therefore this model was chosen for further investigation.

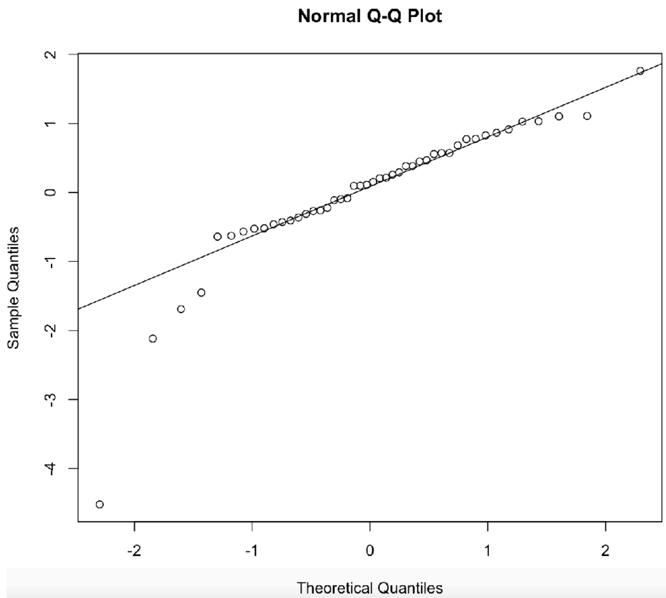
The residuals of the Sinusoidal will now be analyzed to ensure that it satisfies the proper model assumptions.

Figure 4: Histogram of Sinusoidal Model Residuals



This histogram shows the distribution of the sinusoidal model's standardized residuals. The distribution of residuals, which show the discrepancy between actual and anticipated values, offers information about the suitability and accuracy of the model. The distribution of the residuals around zero is approximately symmetrical, but there are a few more intense negative residuals. This points to a distribution that is slightly skewed to the left, suggesting that the model could potentially be biased. The outliers imply that the model bias would occasionally overstate the energy demand. Despite this, the majority of residuals are focused around 0 so the sinusoidal model still appears to be a good fit.

Figure 5: Q-Q plot of SARIMA Residuals



The model's residuals are further examined using the Normal Quantile-Quantile Plot to see if they have a normal distribution. The residuals appear to roughly follow a normal distribution since most of them fall along the straight reference line. This shows that there aren't any significant departures from normalcy in the model. Certain locations in the plot's lower tail (left side) deviate noticeably from the reference line. Although certain extremely negative residuals show sporadic overestimation, the Q-Q plot indicates that the residuals are roughly normal. Although the sinusoidal model does a good job of capturing the majority of the data, it could potentially be made more predictive by managing these outliers and extreme changes with additional predictors or modifications.

Figure 6: Residual diagnostics for the Sinusoidal model

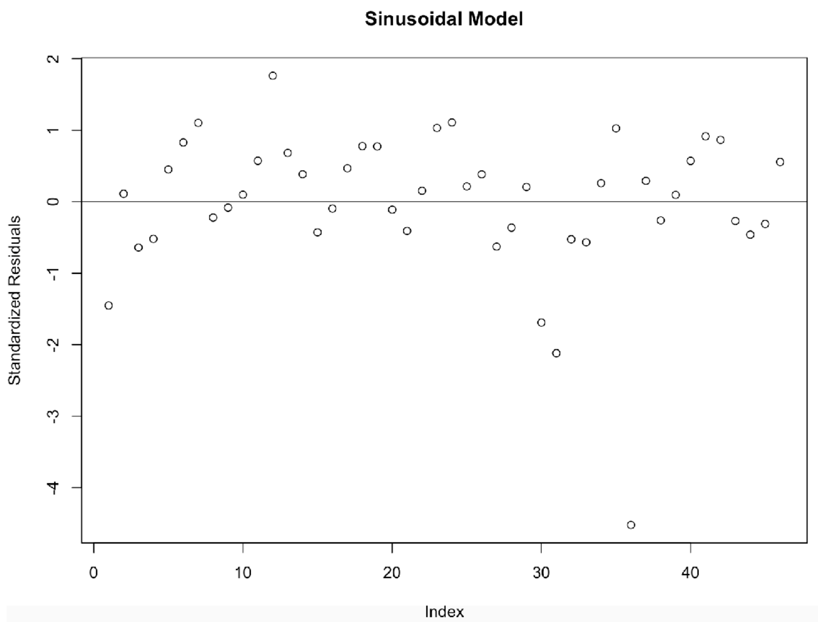


Figure 6 presents the plot of the sinusoidal model's standardized residuals versus their time order. The discrepancies between the actual and anticipated values are represented by residuals. The sinusoidal model's residuals exhibit minor departures despite the singular outlier, suggesting that this model is a good fit to this data.

TIME DEPENDENCY MODEL FITTING

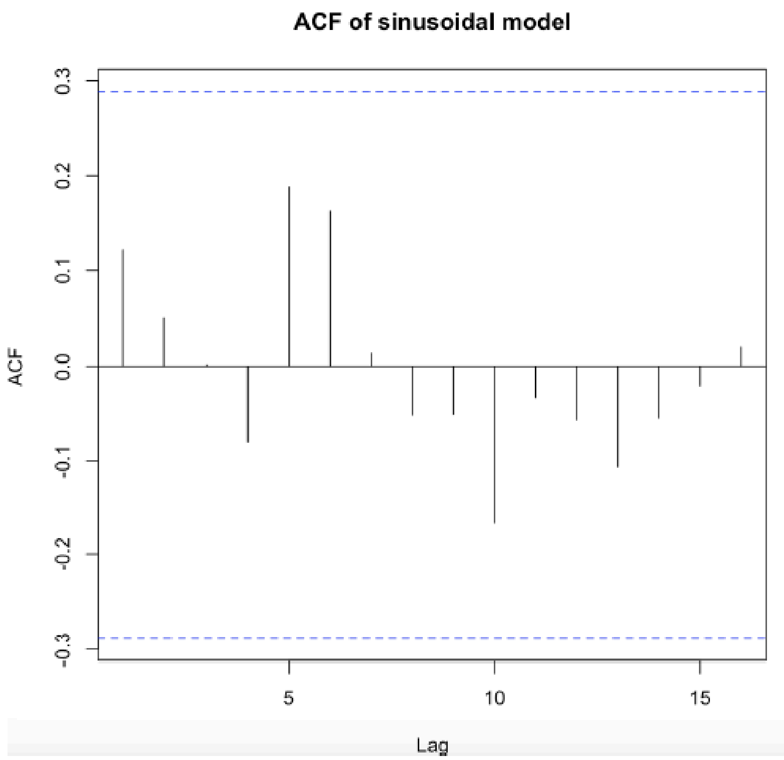
Next, we determined if any ARIMA models would be appropriate for these residuals by performing statistical tests including the Runs Test and Augmented Dickey-Fuller Test, as well as analyzing the ACF and PACF plots. The results can be seen below.

Table 3: Runs Test

P value	0.346
Observed runs	20
Expected runs	23.6087

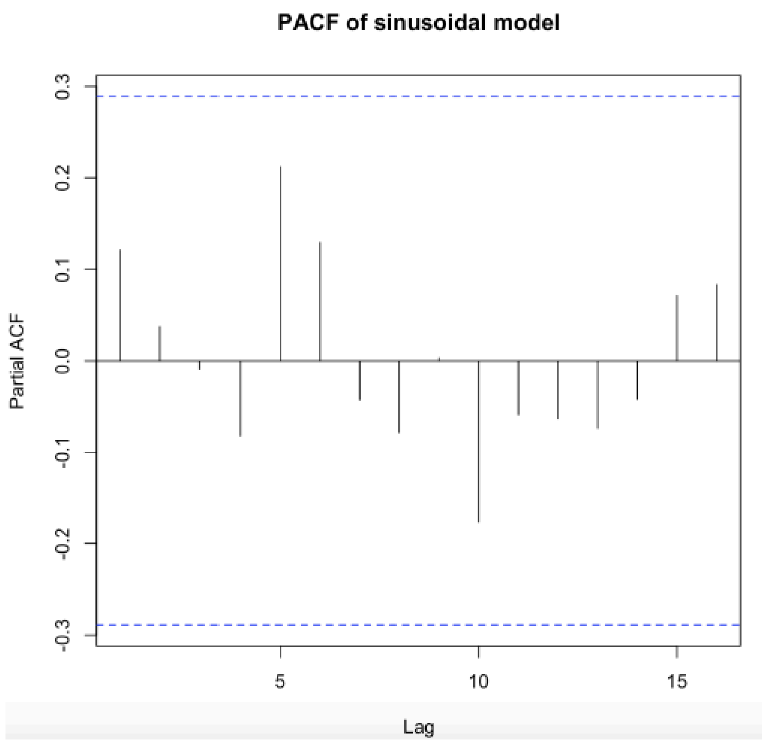
The results of the runs test indicate that the data lacks a dependence on time. There were approximately 24 expected shifts across the mean of the residuals, while only 20 shifts across the mean actually occurred. These values were determined to be close enough together to be insignificant which is exemplified by the large p value of the test.

Figure 7: Autocorrelation function (ACF)



The autocorrelation plot in Figure 7 does not show any significant autocorrelations of the residuals of the sinusoidal model, indicating a lack of time dependence.

Figure 8: Partial Autocorrelation Function (PACF)



The PACF Plot in Figure 8 does not show any significant partial autocorrelations of the residuals of the sinusoidal model, indicating a lack of time dependence.

Table 4: Augmented Dickey Fuller Test

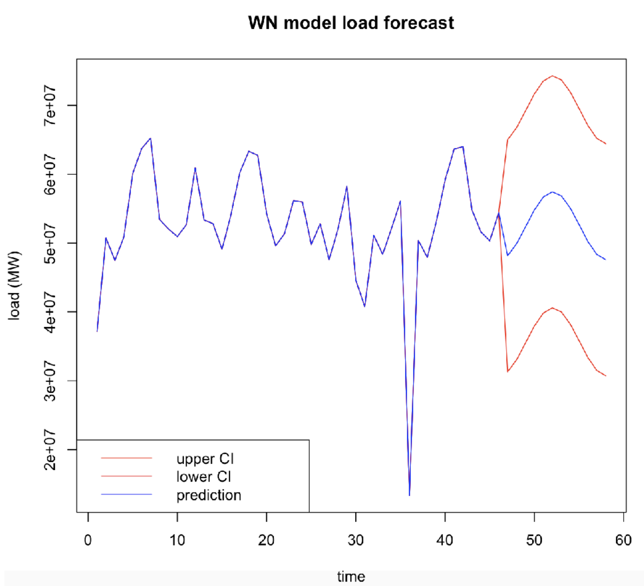
ADF Test Statistic	-6.0353
ADF p-value	0.01

The ADF test was utilized to further exemplify the time series's stationarity. According to Table 2, the Augmented Dickey-Fuller (ADF) Test Statistic was found to be -6.0353 and the p-value was found to be 0.01. Since the null hypothesis of a unit root is rejected, the table shows that the data is stationary and does not need differencing. The residuals are confirmed to be stationary by the ADF test, indicating that there is no unit root behavior in the time series data. This result will allow for improved confidence intervals on the model forecast.

FORECAST

After concluding that a sinusoidal model was the best deterministic trend for the data, and that the residuals were distributed based on a white noise model, the forecast with its confidence interval was plotted to visualize this final model combination.

Figure 9: Sinusoidal Model Forecasting



The actual load is plotted here, with the forecast only starting when the red confidence interval projects outwards. Like previous seasonal patterns, the predicted energy load exhibits a shifting trend. The forecast stays consistent with time due to the white noise assumption of the residuals. This is exhibited by the confidence ranges remaining the same distance throughout the forecast. This plot shows that a sinusoidal model with white noise distributed residuals is a good choice for capturing the load data forecasts.

CONCLUSION

Through our time series plots and model selection residuals, we were able to deduce that the MISO power grid's energy usage shows significant seasonality. There were peaks during months with extreme weather. This seasonal pattern was best described by the sinusoidal model. Once we took the seasonality into consideration with the sinusoidal model, there was no drift exhibited by the residuals in any direction. This lack of drift indicated a lack of macro trends. Lastly, although the deterministic sinusoidal model offers a relatively basic forecast, it proved to be effective for predicting long-term energy consumption as the standard error remained constant due to the white noise model it allowed for the residuals. For utilities to plan for future demand changes and optimize grid operations, these insights are essential. Extreme variances were not fully taken into account by the models in this analysis, indicating the necessity for further predictive elements in the research. To better capture seasonal changes and long-term patterns in energy consumption, variables like economic activity or weather could be included to improve the model.

R Code

```
library(TSA)

library(dplyr)

library(lubridate)

library(tseries)


# Set the working directory to where the data is located

setwd("/Users/crorick/Documents/MS\ Applied\ Stats\ Fall\ 2023/MA5781/group_project")


#read the data in

miso.2021 <- read.csv('miso_load_act_hr_2021.csv', skip = 3)

miso.2022 <- read.csv('miso_load_act_hr_2022.csv', skip = 3)

miso.2023 <- read.csv('miso_load_act_hr_2023.csv', skip = 3)

miso.2024 <- read.csv('miso_load_act_hr_2024.csv', skip = 3)


#choose only columns of interest

miso.2021 <- data.frame(miso.2021$MISO.Total.Actual.Load..MW.,
miso.2021$Local.Timestamp.Eastern.Standard.Time..Interval.Beginning.,
miso.2021$Local.Timestamp.Eastern.Standard.Time..Interval.Ending.)
```

```
miso.2022 <- data.frame(miso.2022$MISO.Total.Actual.Load..MW.,
miso.2022$Local.Timestamp.Eastern.Standard.Time..Interval.Beginning.,
miso.2022$Local.Timestamp.Eastern.Standard.Time..Interval.Ending.)

miso.2023 <- data.frame(miso.2023$MISO.Total.Actual.Load..MW.,
miso.2023$Local.Timestamp.Eastern.Standard.Time..Interval.Beginning.,
miso.2023$Local.Timestamp.Eastern.Standard.Time..Interval.Ending.)

miso.2024 <- data.frame(miso.2024$MISO.Total.Actual.Load..MW.,
miso.2024$Local.Timestamp.Eastern.Standard.Time..Interval.Beginning.,
miso.2024$Local.Timestamp.Eastern.Standard.Time..Interval.Ending.)


#rename columns of interest

colnames(miso.2021) <- c('actual_load', 'end_time')

colnames(miso.2022) <- c('actual_load', 'end_time')

colnames(miso.2023) <- c('actual_load', 'end_time')

colnames(miso.2024) <- c('actual_load', 'end_time')


#concatenate the data

MISO <- bind_rows(miso.2021, miso.2022, miso.2023, miso.2024)


#convert to time series data for manipulation
```

```
MISO$end_time <- as.Date(MISO$end_time)

#converting hourly data to monthly
MISO$end_time <- floor_date(MISO$end_time, 'month')

#sum monthly data
MISO <- MISO %>%
  group_by(end_time) %>%
  summarize(monthly_sum = sum(actual_load))

#convert data into a time series object for ease of analysis
MISO <- ts(MISO$monthly_sum, frequency = 12, start = c(2021, 2))

#plot the data
plot(MISO, type = 'l', xlab = 'Time', ylab = 'Load (MW)', main = 'MISO Monthly Load Data
(2021-2025)')
points(y = MISO, x = time(MISO), pch=as.vector(season(MISO)))

#seasonal means model
```

```
month <- season(MISO)

seasonal.MISO.lm <- lm(MISO ~ month + time(MISO))

AIC(seasonal.MISO.lm)

BIC(seasonal.MISO.lm)


#Use least squares to fit a cosine trend with fundamental frequency 1/12 to the

#percentage change series. Interpret the regression output.

har <- harmonic(MISO, m=1)

MISO.har.lm <- lm(MISO ~ har)

AIC(MISO.har.lm)

BIC(MISO.har.lm)

resid.har <- rstandard(MISO.har.lm)


#Standardized residuals of harmonics model plot

plot(resid.har, ylab='Standardized Residuals', main = 'Sinusoidal Model')

abline(h=0)


#Dickey Fuller Test
```

```
adf.test(resid.har, k=0)

#runs test
runs(resid.har)

#Autocorrelation Function plot
acf(resid.har, main = "ACF of sinusoidal model")

#Partial Autocorrelation Function plot
pacf(resid.har, main = "PACF of sinusoidal model")

#histogram of residuals
hist(resid.har, xlab = 'Standardized residuals', main = 'Sinusoidal Model')

#qqplot of residuals
qqnorm(resid.har)
qqline(resid.har)
```

```
#plot the seasonal means and sinusoidal with the actual

matplot(cbind(predict(MISO.har.lm), predict(seasonal.MISO.lm), MISO),

        type = "l", lty = 1, col = c("orange", "purple", "black"), xlab = "time",

        ylab = "load (MW)", main = "Model Comparison")

legend("bottomleft", legend = c("Sinusoidal", "Seasonal Means", "Actual"),

       col = c("orange", "purple", "black"), lty = 1)


# WN model to forecast the next 12 values of this series.

future.time <- ts(start = c(2024,12), end = c(2028,9), frequency = 12)

future.har <- as.data.frame(harmonic(future.time, m=1))

predictions <- predict(MISO.har.lm, newdata = future.har, type = 'response')

se <- sqrt(var(MISO))


#Plot the series, the forecasts, and 95% forecast limits, and

#interpret the results.

WN.train.pred <- c(MISO, predictions[1:12])

uci <- predictions[1:12] + 2*se*rep(1, times = 12)

lci <- predictions[1:12] - 2*se*rep(1, times = 12)
```

```
WN.train.upper.ci <- c(MISO, uci)

WN.train.lower.ci <- c(MISO, lci)

matplot(cbind(WN.train.upper.ci, WN.train.lower.ci, WN.train.pred),
        type = "l", lty = 1, col = c("red", "red", "blue"), xlab = "time",
        ylab = "load (MW)", main = "WN model load forecast")

legend("bottomleft", legend = c("upper CI", "lower CI", "prediction"),
       col = c("red", "red", "blue"), lty = 1)
```