PREDICTING THE DAILY OCCUPANCY AT A UNIVERSITY GYM USING TIME SERIES ANALYSIS

ARCHIT ARORA, (aa72344)
ISHAN CHANDEL,(ic6987)
RAVI KIRAN DIVVELA,(rkd465)

Table of Contents

Problem Description

In this project a time series model was developed to predict the daily occupancy at a University gym. A dataset[1] consisting of 485 observations for the period August 2015- December 2016 was used for building the model. The dataset as shown in Fig 1. consisted of total number of people visiting a gym(blue) and average temperature for a particular day(red).

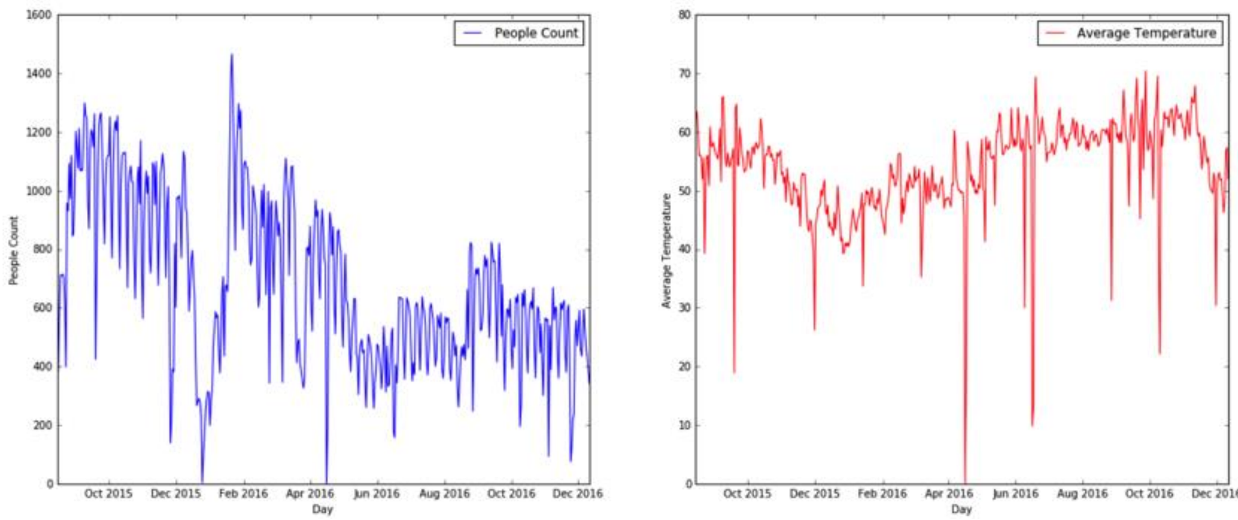

Fig 1: Time series

To train and test the model the dataset was split into training and test set with 455 observations and 30 observations respectively. Two different approaches were tried:

1. A univariate time series model was built using the number of people visiting the gym.
2. A multivariate(ARMAV) model analysing the effect of temperature on the original series was studied and is discussed in the later part of this project.

Modeling Procedure

Fit an ARMA(2n,2n-1) model using F-Test

A simple scalar ARMA model was considered in the starting consisting of the number of people visiting a gym. Successive ARMA(2n,2n-1) models were fitted and F-test was used to check significant reduction in residual sum of squares (using PostulateARMA function). An ARMA(8,7) model was obtained using this procedure. The summary for the model is presented below.

RSS value for ARMA(8,7) is 7.302 X 10^6.

Parameter #	AR(ϕ)	MA(θ)
1	-0.143977903	0.547935056
2	0.113906106	0.558975567
3	-0.05588689	0.408114804
4	0.130002696	0.553114793
5	-0.04712433	0.378420032
6	0.079471343	0.496403635
7	-0.94630719	-0.385926495
8	0.134566947	-

Table 1: Coefficients for ARMA(8,7)

Check Residuals for whiteness

The residuals were then checked for whiteness using autocorr function in MATLAB. Fig 2. below confirmed that the residuals are white and hence ARMA(8,7) model was considered to analyze trends and seasonality.

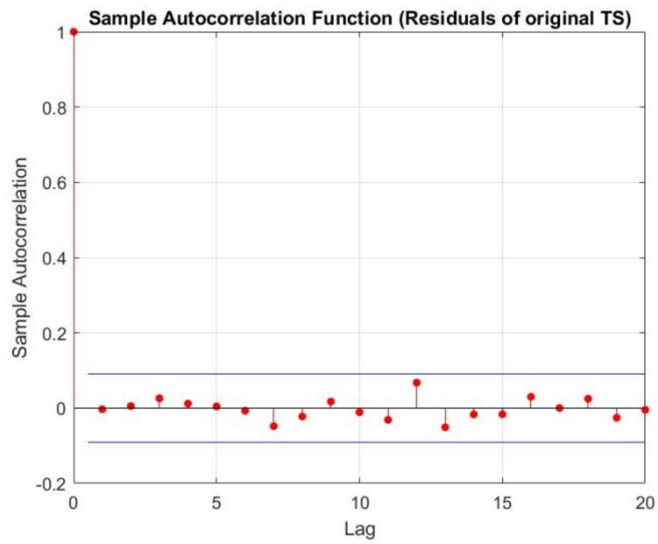

Fig 2: ACF Plot of Residuals of Original Time Series Model

Finding Stochastic Trends and Seasonality

The AR roots of this model were analyzed for any trends or seasonality. The presence of real AR roots close to 1 confirmed a linear trend and complex conjugate pairs on the unit circle indicated stochastic seasonality as shown below.

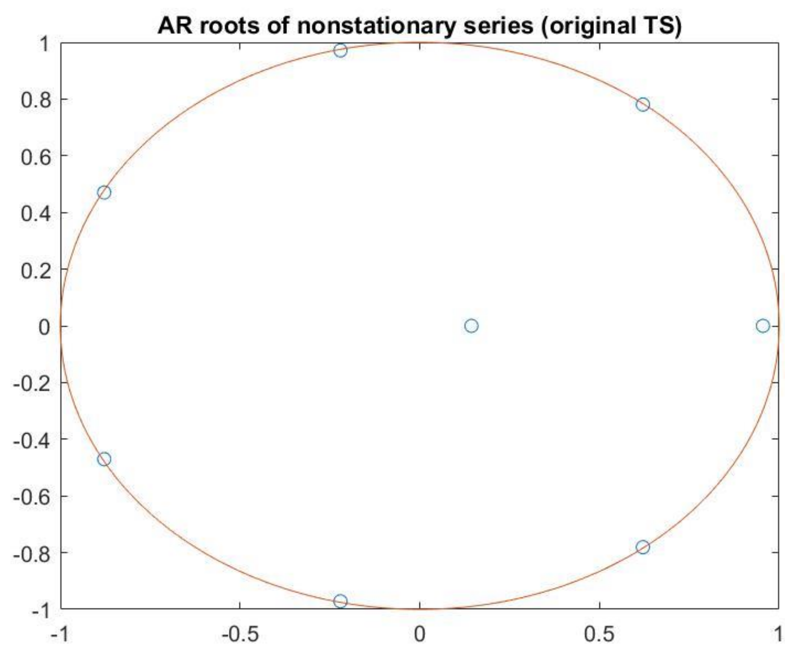

Fig 3: AR Roots of Original Time Series Model

To remove the stochastic trend a linear regression was fitted to the series. With a deterministic approach higher order polynomial fit to the series was also modeled. With the F-test evaluation criteria, linear fit was selected.

Residuals obtained from the linear regression model were modeled (using PostulateARMA) to find the best ARMA model. It was found that, ARMA(8,6) is the suitable model with coefficients shown below.

Parameter #	AR(ϕ)	MA(θ)
1	0.300397133	0.977831838
2	0.163663696	0.922340389
3	0.116168515	0.842474074
4	0.158593352	0.872352981
5	0.120108537	0.811573604
6	0.131598149	0.811686616
7	-0.826664326	-
8	-0.185528767	-

Table 2. Coefficients for ARMA(8,6)

Once the model was built, residuals were checked for the whiteness.

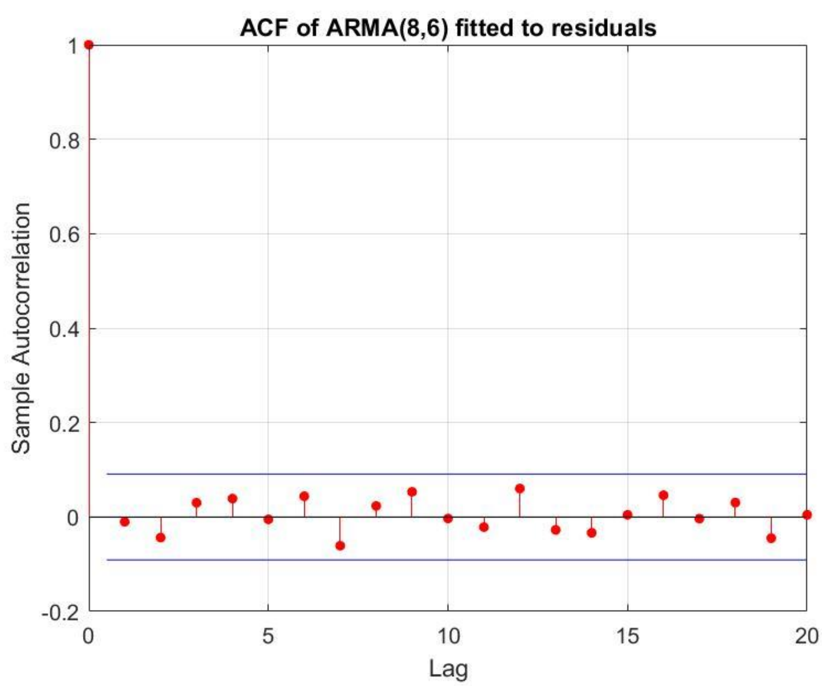

Fig 4. ACF plot Linear model

Since the residuals of the model are white, the AR roots of this model were analyzed for any stochastic seasonality.

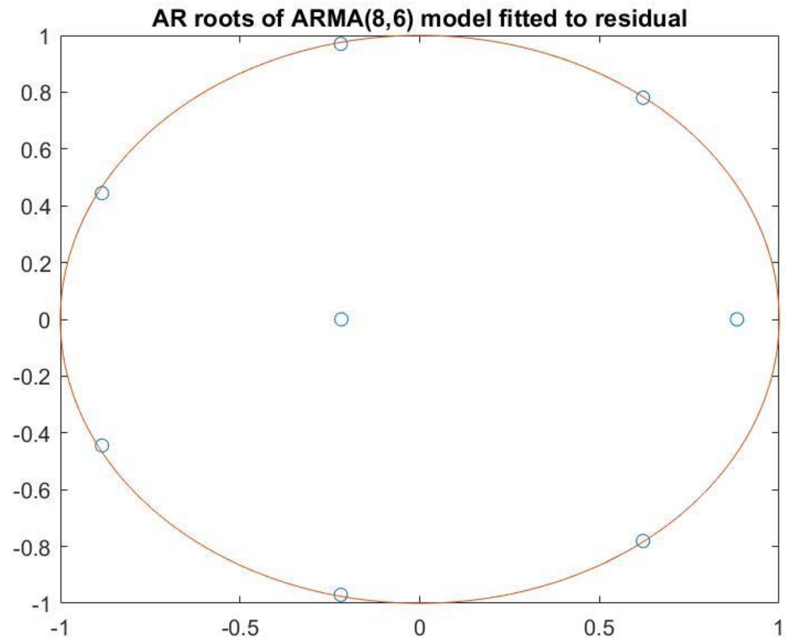

Fig 5. AR roots after removing Linear trend

Since, there were conjugate pairs of roots on the unit circle, this indicated the presence of stochastic seasonality. Using the Parsimonious modeling approach, each of the conjugate pairs residual sum of squares were checked using F-test and it was confirmed that stochastic seasonality of Period 7 and 4 existed. An ARMA(4,6) model was forced on the new series generated by applying seasonality operators on the original series corresponding to period 7 and 4. Coefficients for ARMA(4,6) are shown below.

Parameter #	AR(ϕ)	MA(θ)
1	-2.145254216	-2.001255794
2	2.015690067	2.149906854
3	-0.759173964	-0.968893703
4	-0.065585564	0.035269013
5	-	0.254935918
6	-	-0.13990885

Table 3. Coefficients for ARMA(4,6)

Residuals were checked for white noise.

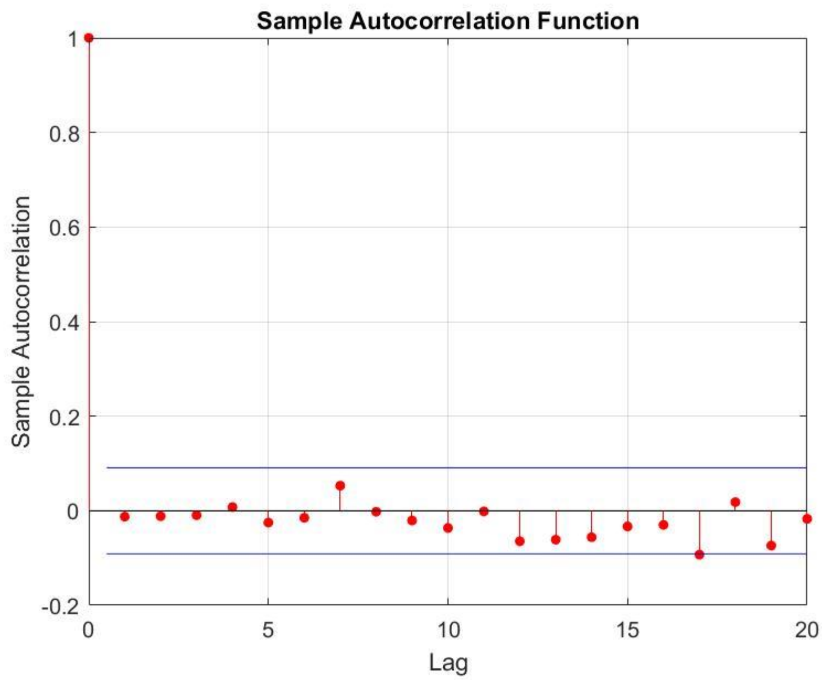

Fig 6. ACF plot ARMA(4,6)

The AR roots were analyzed for any leftover trends or seasonality.

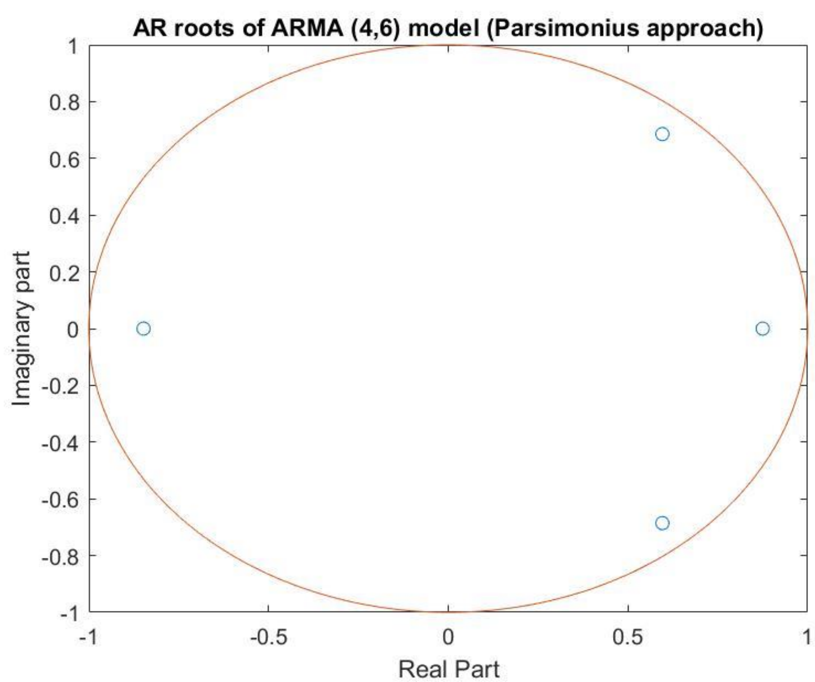

Fig 7. AR Roots for Parsimonious model

Since all the roots were within the unit circle, there was no trend or seasonality left in the final model. This model was used to predict the gym occupancy for the next 30 days.

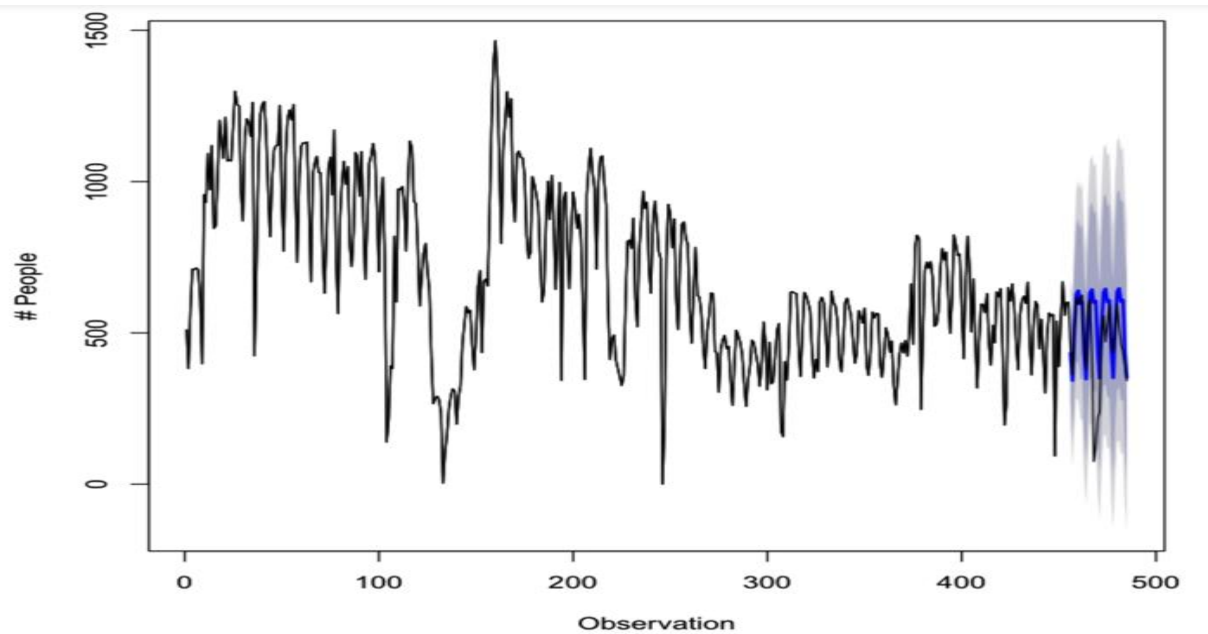

Fig 8. Forecast from ARMA(4,6) for next 30 days

Fitting an ARMAV Model

To better understand the dynamics of the time series, a multivariate time series including daily temperature was analyzed. First series consisted of temperature as input and number of people as response variable. The second series consisted of number of people as input and temperature as response variable. For first series an ARMAV(3,2) was obtained using F-test procedure. For the second series an ARMAV(6,5) model was obtained using F-test procedure. Hence, for the ARMAV model the maximum order model of the two series was taken, that is, ARMAV(6,5) model.

Parameter #	AR(ϕ)	MA(θ)
1	-1.756695807	-1.013162353
2	2.195554366	1.499962765
3	-2.161098389	-1.050470039
4	1.735740403	1.047599611
5	-0.94195918	-0.242902261
6	-0.005459507	-

Table 4. Coefficients for ARMAV(6,5)

Residuals were also checked for white noise before forecasting.

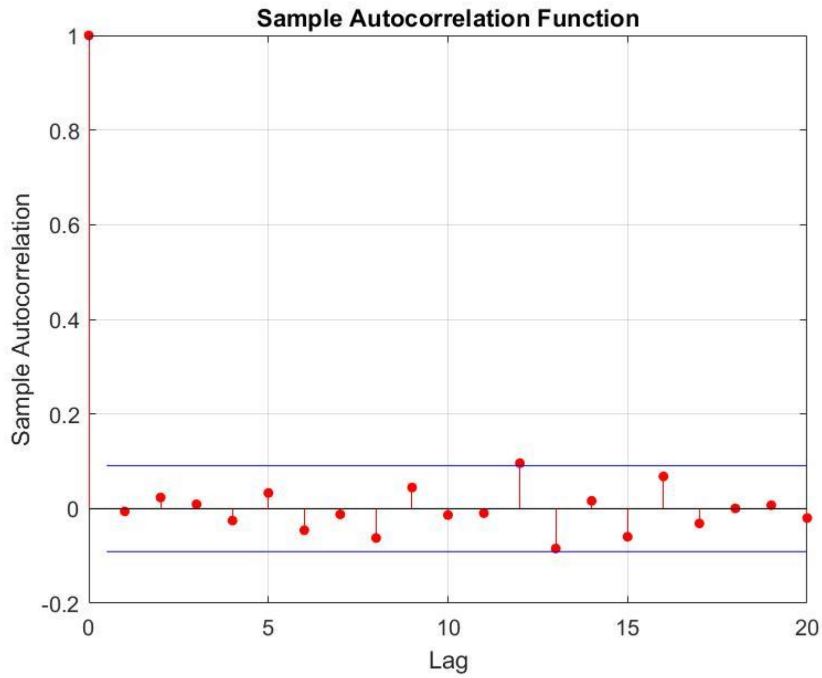

Fig 9. ACF plot ARMAV(6,5)

Forecasting

30 step ahead prediction for the ARMAV(4,6) is as shown below.

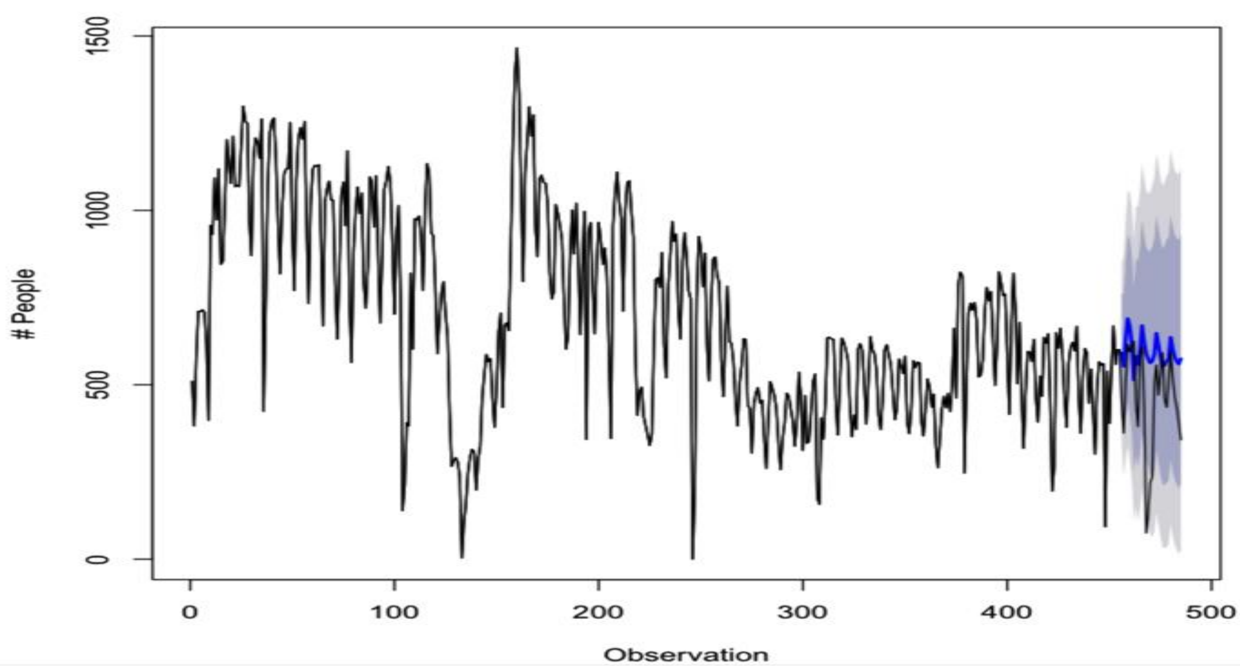

Fig 10. Forecast from ARMAV(6,5) for next 30 days

Conclusion

1. The original model was ARMA(8,7), which was obtained using F-test procedure. After removing trends and seasonality using the parsimonious model approach, an ARMA(4,6) was obtained. This model has been used for making final forecast.
2. The data consisted of seasonality with periods 4 and 7, which were obtained after analysing the AR roots and confirmed using F-test.
3. In ARMAV approach, the effect of temperature was studied on the number of people visiting the gym. However, since the trends and seasonality were not analysed for ARMAV, the forecast accuracy could not be compared with the simple univariate ARMA model.

Future Scope

1. Effect of university holidays could be included in the model to better understand the data.
2. Studying the trends and seasonality in the ARMAV can give improved predictions.

References

1. https://www.kaggle.com/plozano94/gym-data
2. Class Notes from Time Series Class of Dr. Dragan Djurdajnovic.
3. Time Series and System Analysis With Applications by Sudhakar M. Pandit (Author), Shien-Ming Wu (Author)