**Project: Forecasting**

*Objectives*

To use various forecasting algorithms to determine the best model for a time series that will be created using your student id (SI) number. For this analysis you can use any statistical package, such as MatLab, R, SAS, or use Python with all the available statistical functions.

*Data Set*

You will use the data set provided to you that is completely randomized based on your student id. The set of all data sets for all the students in the class is posted in the schedule under the name: DataSets. Your data set consists of 2000 observations. Partition your data set into two parts, the training set and the testing set. For the training data set use the first 1500 observations, and keep the last 500 for the testing data set.

*Tasks*

*Simple moving average model*

1.1 Apply the simple moving average model, i.e., $x_t = \frac{1}{m}\sum_1^m x_{t-i}$ to the training data set, for a given *m*.

1.2 Calculate the error, i.e., the difference between the predicted and original value in the training data set, and compute the the root mean squared error (RMSE).

1.3 Repeat above two steps by varying *m* and calculate the RMSE.

1.4 Plot RMSE vs *m*. Select *m* based on the lowest RMSE value. For the best value of *m* plot the predicted values against the original values.

2. *Exponential smoothing model*

2.1 Apply the exponential smoothing algorithm, i.e., $s_t = ax_{t-1} + (1-a)s_{t-2}$ to the training data set, for a given $\alpha$.

2.2 Repeat steps 1.2 to 1.4 in order to find out the best value for $\alpha$.

2.3 Select $\alpha$ based on the lowest RMSE value. For the best value of $\alpha$ plot the predicted values against the original values.

3. AR(*p*)

3.1 Apply the autoregressive algorithm AR(*p*) to the training data set for a given value *p*.

3.2 Select *p* by plotting the partial autocorelation function.

3.3 Estimate the parameters of the AR(*p*) model. Provide RMSE value and a plot the predicted values against the original values.

4. Run all three models on the test data, and chose the best one.

*What submit*

Submit all the requested results, your conclusions, and the code that you wrote to obtain the results. It is very important that you provide enough results to support your conclusions. Conclusions without insufficient results will make you lose points. Also, it is important that you develop your own code. Sharing code is not allowed and constitutes cheating, in which case both students (the

one that aids and the one that receives) will get a zero for the project and will be reported to the student conduct office.

*Grading*
The TA will first verify that your code works and produces the results you submitted. The break down of the grades will be as follows:

Task 1: 30 points
Task 2: 35 points
Task 3: 25 points
Task 4: 15 points

Remember that you will be graded mostly on your ability to interpret the results