# Enrollment Forecasting for School Management System

Rabby Q. Lavilles and Mary Jane B. Arcilla

***Abstract*—The electronic School Management System (e-SMS) of Mindanao State University – Iligan Institute of Technology (MSU-IIT) is an information system that supports academic transactions in the university. Admission, curricula, pre-registration, registration, grading, and records management are functions supported by the system. Although these functions helped in the management of the university, enrollment planning specifically student projection is not supported resulting to problems on student complaints about course unavailability, merged courses or dissolved sections. This study explores time series models in projecting the number of students enrolled in a course as forecasting support for e-SMS. A statistical modeling method is adapted in developing the models which includes problem definition and data collection, model formulation, model verification and model implementation. Problem definition and data collection involves defining variables to be considered as well as cleaning the data. Model formulation uses time series models and graphs in analyzing patterns of data in each course. Model verification is done by comparing the result of the projection and actual enrollment while implementation explains the integration of the enrollment forecasting to the e-SMS. Experimental result yields an average of 20% difference between the forecasted and actual values. The resulting forecasts can be used to support in determining the number of sections to be opened before the enrollment commence.**

***Index Terms*—School management system, student projection, enrollment forecasting.**

## I. Introduction

The interactive model of information system emphasizes on the role of information systems to assist and support a user to achieve greater efficiency and effectiveness of decision-making [1]. In industrial and services system setting such as inventory, modeling and analysis techniques, operation research applications are common. However, applications of operations research in state university administration in Philippines have been surprisingly rare. This research explores time series models in enrollment forecasting to project the number of student to enroll in a course. This answers the need of the school administrator to understand enrollment patterns and factors that influences the number of students expected to enroll.

Educational institutions abroad project enrolment for the coming semester uses different statistical models to project the number of students to enrol in a certain course or year level. A study on the enrolment forecasting for an upper division general education component uses regression analysis based on historical enrolment data to forecast the total student enrolment as well as the student demand for seats in theme courses at Grand Valley State University [1].

Manuscript received July 10, 2012; revised August 15, 2012.

The authors are with Mindanao State Univeristy – Iligan Institute of Technology, Philippines (e-mail: rabby.lavilles@g.msuiit.edu.ph).

Another study conducted on the enrolment forecasting models for Community Colleges [3]. Six (6) community colleges were used as a sample data. The study uses three enrolment projection methods: regression, autoregression, and three-component model. Moreover, a study on the application of Markov chain model to forecast undergraduate enrolment was explored [4], [5]. These studies attempt to solve the uncertainty involved in projections and the statistical models used. Moreover, these studies are appropriate on the institutions or context where it were studied.

In this study, a program was developed with a statistical tool that assist the academic program administrators in forecasting the number of students expected to enroll in a subject as well as determining the subjects to be offered. Models are based on the data of MSU-IIT from school year 1998 to 2009 and were evaluated using the actual data of school year 2009-2010.

## II. Student Projection

Mindanao State University – Iligan Institute of Technology (MSU-IIT) uses naïve forecasting as the current practice of the university in determining the number of student for the incoming semester. The request of courses from different department is the basis for the offering of the course as well as in determining the number of students expected to enroll. The request of courses is a document that contains a list of subjects and the expected number of students to enroll sent to the department offering the subject. It is received by the department or college before the enrolment starts. The interview with the department chairpersons and department adviser identified several factors that can be considered to project the number of students expected to enroll in a subject. These factors are past enrolment data, curriculum or program offered, prerequisite and number of students enrolled in the program.

Based on the enumerated factors, the researcher focuses on curriculum and past data. Curriculum is the basis for identification of subjects to be offered and past data will be used to determine the number of students expected to enroll. Major functions supported by the module are build or update a model based on subject, department, college or school, generate forecast for the department or college, and graph the subject history as well as show the departments involved.

## III. Forecasting Models

Three models were evaluated to fit the data for each subject. Among these models single and double exponential smoothing are used to model the subject and the basis for student projection.

### A. Simple Moving Average

Moving average techniques forecast demand by calculating an average of actual demands from a specified number of prior periods. Each new forecast drops the demand in the oldest period and replaces it with the demand in the most recent period [6]. The formula for simple moving average of order 3 is shown in (1).

$$T_t = \tfrac{1}{3}(Y_1 + Y_2 + Y_3) \tag{1}$$

where

$Y_1$ – is the 3rd value from the value to be forecasted
$Y_2$ – is the 2nd value from the value to be forecasted
$Y_3$ – is the 1st value from the value to be forecasted

### B. Single Exponential Smoothing

Exponential smoothing is highly suitable for environments such as inventory systems where forecasts must be made [7]. The equation for single exponential smoothing is shown in (2) [8].

$$F_{t+1} = \alpha y_t + (1-\alpha)(F_{t-1}) \tag{2}$$

where

$F_{t-1}$ is a forecast for the period before current time period t
$y_t$ is the actual data at time period t
$\alpha$ is weight given to the latest data

### C. Double Exponential Smoothing

The formula for double exponential smoothing is shown in (3) and (4).

$$F_t = \alpha y_t + (1-\alpha)(F_{t-1} + T_{t-1}) \tag{3}$$

$$T_t = \beta(F_t - F_{t-1}) + (1-\beta)T_{t-1} \tag{4}$$

where:

$y_t$ is the observed value at time t.
$F_t$ is the forecast at time t.
$T_t$ is the estimated slope/trend at time t.
$\alpha$ - representing alpha - is the first smoothing constant, used to smooth the observations.
$\beta$ - representing beta - is the second smoothing constant, used to smooth the trend.

Initializing the value of the models is dependent on the implementation. In this study the initial value is computed by setting the first $F_t$ to $y_t$, and the initial slope $T_t$ is set to the difference between the first two observations [9].

Moving average is compared with single exponential smoothing to determine which models can represent the data of the university. These models represent the mean of the student enrolled in a given subject. Double exponential smoothing is used in subjects that represent increasing or decreasing trends.

### D. Choosing the Smoothing Constants

The smoothing constants must be values in the range 0.0-1.0. The most appropriate smoothing constant depends on the data series being modeled. In general, the speed at which the older responses are dampened is a function of the value of the smoothing constant. When this smoothing constant is close to 1.0, ignoring old data is quick which means more weight is given to recent observations and when it is close to 0.0, is given relatively less weight is to recent observations. The best value for the smoothing constant is the one that results in the smallest mean of the squared errors or other similar accuracy indicator.

### E. Forecasting Error

Differences or deviation from forecast and the actual data is compared to determine the accuracy of the model. This phase of model building is called model evaluation. The goal is to compare and get the difference between actual and forecasted value using mean absolute percentage error (MAPE). MAPE is a measure of accuracy in a fitted time series value in statistics, specifically trending. It usually expresses accuracy as a percentage, and is defined by the formula [10][11] in (5):

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{5}$$

where

$A_t$ – Actual value
$F_t$ - Forecasted Value

## IV. Methods and Results

Initial analysis reveals that there are subjects in the university which refer to the same subject but used different subject codes. This leads to cleaning the data as preparation for the modeling. Graphical presentation of MA of order three (3) showed that there is a similarity with the SES. The consideration of MA to be integrated in school management system was dropped. Focus on SES and DES and determining the alpha and beta of the models that will produce the least error were the concern of the researcher.

As shown in Fig. 1 and Fig. 2, the effects of the different values of alpha on the service courses shows that an alpha of 0.9 gives more value to the latest values used in the forecast as described in the model. The graph with alpha of 0.1 creates an average value of the data. It smoothes the differences that the actual data is producing and tends to create an average value based on the history of data. Hence, the value of alpha on SES determines the percentage to be extracted from the latest data. Determining the value of alpha therefore is a critical concern of the SES.

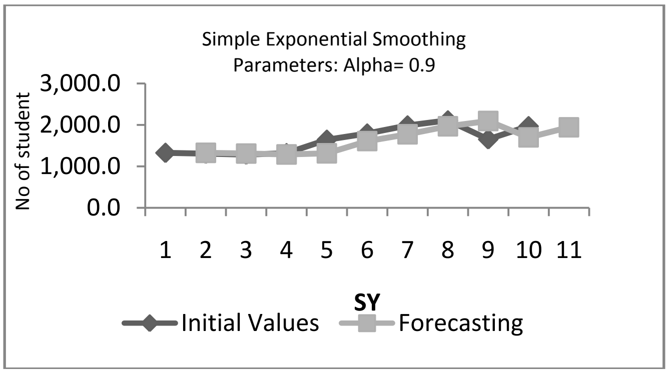


Fig. 1. Single exponential smoothing with alpha of 0.9.

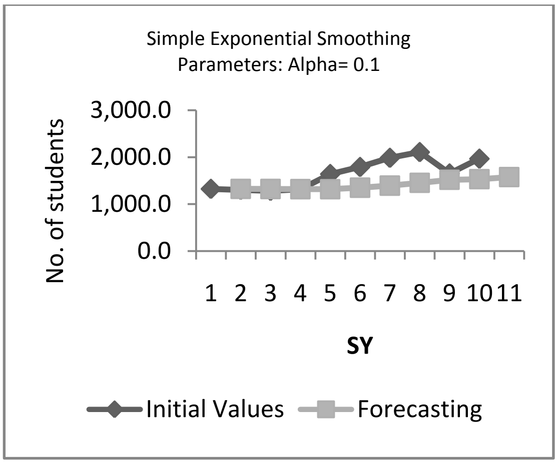


Fig. 2. Single exponential smoothing with alpha of 0.1.

The test result of the smallest average absolute percentage error of the sample subject in the evaluation of alpha is used as the alpha of the single exponential smoothing. Alpha of 0.9 yields the smallest error based on MAPE.

Based on the sample subjects used in the study 80% of the sample shows that on average subjects with alpha of 0.9 in SES yield the smallest MAPE. This means that 80% of the subjects give more value to the recent data rather than on old records on the number of students.

Double exponential smoothing (DES) shows the same results in terms of the value of alpha. It is evident that the latest value in more valuable since it implies the latest student enrolled in a certain subject. Although not all subjects exhibit the same patterns, single and double exponential models are sufficient to explain the patterns of behavior of the subject to be integrated in the school management system.

Although 80% of the subjects showed that 0.9 yields least error in term of MAPE. It is apparent that it is difficult to generalize the value of alpha in SES and alpha and beta for DES. The attempt to formulate a general model that will represent the patterns of behavior of the subject is not practical. Since a subject has its own pattern and there is no definite alpha and beta for both models that will generalize the subjects. A computerized program will determine the appropriate model and also the corresponding alpha and beta.

## V. Modeling Process

Moving average is comparable with naive model of forecast. Including the model as part of the academic program management is not a pragmatic approach. It also exhibits the same characteristics with single exponential smoothing but defers only on the value assigned to past data. Single exponential smoothing shows that on average an alpha of 0.9 using MAPE or MPE shows less error. These subjects show that as the alpha is increasing the error decreases. On the other hand, 20% of the subjects showed the opposite pattern. That means that 80% of the subjects considered the latest observation as major factor in projecting the number of subjects. The remaining 20% gives emphasis on old records of the students enrolled in a subject.

### A. Initial Analysis

Initial analysis on single exponential smoothing can be interpreted that alp ha of the subjects cannot be generalized to an average alpha that will be applied to all subjects. Even though there are similarities of the patterns on 80% of the subjects, it does not guarantee that each subject has the same alpha. The model with alpha that produces the smallest error does not belong to the average because every subject has its own alpha based on data. Even the 20% of the subject does not produce the same alpha. It still ranges from 10 to 50 percent difference which matters on large number of students.

Double exponential smoothing shows varied alpha and beta combination with little differences in error. It shows that part of the result wherein on average, an alpha of 0.9 with beta of 0.1 and alpha of 0.8 with beta of 0.1 has little differences of 16.4 and 16.5 MAPE respectively.

These results show that using double exponential give emphasis on the latest base value and less value is given to the trend part. The beta of 0.1 means that 10% will be derived from the trend component of the subject.

### B. Model Selection

Selecting a model to be integrated in the school management system is based on the evaluation of the smallest percentage error. A MAPE is the main basis since it is not offset by negative numbers in the forecast value [11]. In single exponential smoothing, an alpha with smallest error is compared to the smallest error of the double exponential smoothing. The model with least error was used to forecast on that particular subject given the alpha for single exponential smoothing and beta and alpha for double exponential smoothing.

The idea is implemented into a program so that generation of the alpha for single exponential smoothing and alpha and beta for double exponential smoothing will not be a tedious and computation extensive work for a forecaster.

### C. Model Building

Brute force is used to find the least error based on MAPE. Subjects that exhibit a consistent pattern of increasing or decreasing number of students enrolled were observed to be modeled using double exponential smoothing. On the other hand, single exponential smoothing shows that it is a good model for subjects that follow no particular pattern but more on identifying the mean depending on the value of alpha.

A total of 182 subjects were used in generating the least error model based on the available data. About 58% of subjects generated have least average MAPE using double exponential smoothing with varying alpha. The remaining subjects used single exponential smoothing. The 58% of the subjects that uses double exponential smoothing with varying alpha shows that 58% of the subjects have consistent patterns whether increasing or decreasing patterns of enrolled students. Subjects that follow the double exponential smoothing with beta greater that 0.30 shows a consistent increasing or decreasing values. Those that have small changes on the number of students every school year, has a lower beta.

The remaining 42% uses single exponential smoothing which mean that these subjects have an average number of students throughout its offering. The 42% does not mean that it has consistent number of students but might have patterns that on the first part of the school year is increasing and on the second part have decreasing patterns. This behavior causes the value of alpha to consider the value of the past values or offset some of the errors because of the large error it generated on some parts of school year while the model is in the process of generating the value of alpha and beta with smallest MAPE.

### D. Testing and Evaluation

The result of testing is done by comparing the actual number of students enrolled in first semester academic year 2009-2010 with the result of the forecasted value for the said year and the naive method.

The projected data from the time series models yield less error using MAPE with 20.5 % difference in favor of the time series models. The 20.5% advantage of the model over the current practice which is the naive method doesn't have much effect on subjects with students less than 100 but in subjects with enrolled students which can go beyond one thousand, this value matters. Take for instance ENG 1 which has an enrollee of 2,514, 20% of that number is more or less 502 students. This number of students can open up to 12 or 13 sections with 40 students each section. The 20% difference has major effect on the consideration of the number of sections. The 20.5% advantage over the naive method can already create a difference on estimating the number of students to enroll in a certain subject. Of the subjects selected as part of the sample study, about 87% have enrollees of more than 100. It implies that on service courses, it is a good assistance in terms of determining the number of students. Furthermore, majority of service subjects have more than 100 number of students enrolled.

### E. Integration of Forecasting

The enrollment forecasting module handles the determination of subjects as well as number of students expected to enroll. The model is implemented using postgreSQL. It is programmed in the database server as part of a function of the database. It will then generate a table of the list of subject and the expected students to enroll.

The table is to be retrieved in the graphical user interface of the school management system. This can be used as reference or input when deciding how many sections to open in a particular subject.

## VI. Conclusion and Future Work

The integration of a forecasting module to the school management system is of assistance in projecting the number of students expected to enroll in a subject. The alpha for SES and alpha and beta for DES are critical factors to be considered in using a time series models. These values are dependent on the data to be used. The brute force approach used in the study shows a promising result although in terms of performance other algorithms can also be considered.

Three time series statistical models are considered in the study. The simple moving average of order 3 (MA 3), SES with varying alpha and DES. MAPE is used to evaluate the fit or accuracy of the data to the models. About 58% of subjects generated have least average MAPE using double exponential smoothing with varying alpha and beta. The remaining subjects use single exponential smoothing with alpha having least MAPE.

The result of the projections for academic year 2009-2010 is compared to the naive model used by the university. Initial result yields less error using MAPE with about 20 % difference in favor of the time series models.

The techniques used in this study can be used to other educational institutions with similar setting. It can also be used to enhance the understanding of enrolment patterns. The connection of advising and projection of the number of students can also be used to create a model for forecasting.

Other statistical models can be integrated to the existing model to enhance the accuracy of the model. Furthermore, combination of different statistical models can be considered for further study. One form of combination is using cohort survival models of every program and used it as an independent variable in projecting the number of students to enroll. One category of quantitative models is causal models. This can also be explored by using the factors identified in this study.

## Acknowledgment

This research is funded by PGMASEGS (President Gloria Macapagal-Arroyo Science and Engineering Graduate Scholarship) under Commission on Higher Education of the Philippines.

## References

[1] D. Cecez-Kecmanovic, "The Discipline of Information Systems-issues and challenges," R. Ramsower and J. Windsor (Eds.), in *Proc. of the Eight Americas Conference on Information Systems AMCIS 2002*, Dallas, Texas, USA, August 9-11 2002, pp. 1696-1703.

[2] S. E. Choudhuri, C. R. Standridge, C. Griffin, and W. Wenner, "Enrollment Forecasting for an Upper Division General Education Componentz," *37th ASEE/IEEE Frontiers in Education Conference*, Milwaukee, WI, 2007.

[3] S. Guo, "Three Enrollment Forecasting Models: Issues in Enrollment Projection for Community Colleges," presented at the 40th RP Conference Asilomar Conference Grounds Pacific Grove, California, May 1 –3, 2002

[4] J. M. Fraser, S. Djumin, and J. J. Mager, "The University as Educational Lab," in *Proc. 1999 American Society for Engineering Education Annual Conference & Exposition*,1999

[5] J. C. Segura-Ramirez and W. Chang, "Using Markov Chain and Nearest Neighbor Criteria in an Experience Based Study Planning System with Linear Time Search and Scalability," in *Proc. 2006 IEEE International Conference*, Waikoloa Village, HI, 2006, pp. 395-403.

[6] O. T. Olalekon, B. M. Oyewole, and O. A. Olawande, "Forecasting Demand for Office Space in Ikeja, Nigeria," *Mediterranean Journal of Social Sciences*, vol. 3, pp.323-338, .January 2012.

[7] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge: Cambridge University Press, 2005.

[8] NIST/SEMATECH e-Handbook of Statistical Methods. [Online]. Available: http://www.itl.nist.gov/div898/handbook, U. S. Commerce Department's Technology Administration, 2012.

[9] M. L. Berenson and D. M. Levine, *Basic Business Statistics*, 7th Edition. USA: Prentice Hall, 2006.

[10] J. R. Evans, *Statistics, Data Analysis, and Decision Modeling*, Upper Saddle, New Jersey: Prentice Hall, 2000.

[11] M. Hamburg, *Statistical Analysis for Decision Making*, 6th Edition, Belmont, California: Duxbury Press. 1996.

**Rabby Q. Lavilles** is a graduate of MS in Information Technology at De La Salle University – Manila, Philippines. He is a faculty member of the Information Technology Department of Mindanao State University – Iligan Institute of Technology.

**Mary Jane B. Arcilla** is an Assistant Professor at De La Salle University – Manila, Philippines. She is a graduate of MS in Information Technology from De La Salle University – Manila, Philippines.