TABLE I. COST - EFFECTIVENESS INDEX FOR STUDENT PROJECTION MODELS

Index = (Error) (Cost)

Average of By-Level Data Mean Square

Technique	Error	Average Cost Per Run	Index
Judgement Only	17.9	$ .0 (.10)	1.79
Baseline	16.1	.10	1.61
Exponential Smoothing	17.3	.15	2.60
Linear Regression	27.3	.25	5.83
University Model	20.0	1.85	37.00
Special Knowledge	17.7	1.85	32.75
WICHE SFM-IA	24.8	76.67	1,901.42

DOCUMENT RESUME

ED 123 973 HE 007 775

AUTHOR Blake, R. John; Robertson, Leon B.
TITLE Comparison of the Effectiveness of Six Models in Forecasting Student Demand on Academic Departments. Final Report.
INSTITUTION Utah Univ., Salt Lake City.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
BUREAU NO BR-3-1235
PUB DATE Jun 75
GRANT NE-G-00-3-0073
NOTE 58p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Bibliographies; Comparative Analysis; *Departments; Educational Assessment; *Educational Demand; Evaluation; Evaluation Methods; *Higher Education; Models; *Predictive Measurement; *Program Effectiveness; Student Needs; Tables (Data)

ABSTRACT

An accurate forecast of the student demand by level on the academic departments of an institution is vital for budget and financial planning decisions, for faculty workload scheduling, and for physical facility planning. Many methods have been used to forecast this demand, ranging from "seat of your pants" guessing to highly complex computer models. This research project studied six basic models for forecasting student demand at various levels of sophistication and complexity. The models studied included judgment only, a ratio model, a Markov model and a combination model. In addition, the dimension of expert judgment was combined to one model to determine the value of the additional input. The model with the expert judgment added was the best model based on the criterion of least-error using several error analysis indices. The simple models gave as good or better forecasts than the more complex models using the same least-error criterion. Also, using a cost-effectiveness criterion, the simple models were again superior to the costly, sophisticated models. (Author)

FINAL REPORT

COMPARISON OF THE EFFECTIVENESS OF SIX MODELS IN FORECASTING STUDENT DEMAND ON ACADEMIC DEPARTMENTS

National Institute of Education Project No. 3-1235
Grant No. NE-G-00-3-0073

R. John Blake and Leon B. Robertson

University of Utah
Salt Lake City, Utah

June 1975

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

WASHINGTON, D.C. 20202

TABLE OF CONTENTS

TABLES

APPENDIX

FIGURES

COMPARISON OF THE EFFECTIVENESS OF SIX MODELS IN FORECASTING THE STUDENT DEMAND ON ACADEMIC DEPARTMENTS

SUMMARY

An accurate forecast of the student demand by level on the academic departments of an institution is vital for budget and financial planning decisions, for faculty workload scheduling and for physical facility planning. Many methods have been used to forecast this demand ranging from "seat of your pants" guessing to highly complex computer models.

This research project studied six basic models for forecasting student demand at various levels of sophistication and complexity. The models studied included judgement only, a ratio model, a Markov model and a combination model. In addition, the dimension of expert judgement was combined to one model to determine the value of the additional input.

The model with the expert judgement added was the best model based on the criterion of least-error using several error analysis indices. The simple models gave as good as or better forecasts than the more complex models using the same least-error criterion. Also, using a cost-effectiveness criterion, the simple models were again superior to the costly, sophisticated models.

SECTION II: DESCRIPTION OF STUDY

PROBLEM STATEMENT

Ten-year student enrollment forecasts have two primary uses--financial planning and faculty scheduling. One of the basic inputs to a university budget or long range plan is an accurate prediction of the student load by level on each academic department. Student demand by level is so vital because cost per student varies widely by department and also by level within a department. Recent studies on the cost of instruction at the University of Utah[29] have shown that cost per student differs by as much as a factor of 14.5 between departments and by as much as a factor of 46.6 between levels in the same department.

In addition to budget and financial planning decisions, another important use of student demand data is for planning faculty workloads and for scheduling faculty, classroom and advising assignments.

Many complex factors affect the student demand on departments such as changing student expectations, judgments concerning career opportunities, and varying economic and political conditions. These fluctuating student attitudes make the forecasting of enrollment by department increasingly difficult. To compound the problem, these dynamic changes are occurring at a time when budgeting constraints in both the short and long run are demanding more accuracy and when scheduling each faculty member to obtain ~~the complete~~ utilization of his talents is critical.

BACKGROUND

Many attempts have been made over the years to forecast student enrollments at universities. Some early methods were for university budget makers to make an educated guess on next year's enrollment. For small schools with only a few departments and limited students, these rough estimates were accurate enough.

As schools increased in the number of students and in the number of departments, the budget makers had an increasingly difficult time making accurate forecasts. In addition, the budgeting and planning horizon that the administrators were concerned about kept expanding into the future. So the student forecasting problem was compounded beyond the limits of the administration to handle by the "seat of their pants" or by their judgement alone.

The next development was to use some mathematical technique or model for forecasting student enrollments. The models that have proven most successful fall into the following five broad categories:

1. Ratio Techniques
2. Regression Analysis
3. Markov Models
4. Simulative, Branching, Network and Programming Techniques
5. Combination Models

A brief discussion of each of these forecasting methods as applied to student enrollments follows in the next sections.

CATEGORIES OF STUDENT PROJECTION MODELS

Ratio Techniques

Ratio techniques are generally based on the assumption that a ratio adequately describes the probability of passing from one state or classification to another. These "states" usually refer to a category of a classification variable such as student level (undergraduate), major (history) or status (continuing). Some of the models using the ratio method as a basis for forecasting follows: Cohort Survival Technique,[18,28] Class Rate Progression Technique,[18] and Simple Ratio Method.[26]

Regression Analysis

Regression techniques are generally based on the assumption that the trends and relationships observed in the past will continue in the future and that the samples are taken independently. Several models incorporating this technique as part of a larger system are as follows: the CAMPUS Model[16] developed by the Systems Research Group, the Michigan State University Model,[17] the Tulane University Model,[6] and the Peat, Marwick, Mitchell and Company CAP:SC Model.[27] The Trend Line Model[26] analyzed in the ACT research is of the regression type. The University of Utah model[3] is basically a linear regression model. This model will be discussed in more detail in a subsequent section.

Markov Models

Markov models are generally based on the basic assumption that transitions from one state to another during an increment of time depend on the present state and are independent of the past. The Markov model is a discrete time system characterized by a collection of probabilities (transition probability matrix). Each probability represent the likelihood that a student will move from his present state to another state during the next time interval. Another assumption of this model is that this transition probability matrix remains constant over time.

The Markov model is perhaps the most popular model of student flow at the present time, at least in the literature. Some of the published reports of Markov models include the following: Gani,[8] Young and Almond,[32] Oliver,[24] Oliver and Marshall,[25] Marshall, Oliver and Suslow,[20] Orwig, Jones and Lenning,[26] State of Washington,[22] Johnson[11] and Fullerton.[23]

As can be surmised from the number of current models, the Markov model is very popular for forecasting student enrollments. The basic Markovian assumptions, however, tend to be counter-intuitive, as pointed out by Lovell.[19] The stationarity assumption seems inappropriate in the dynamic educational system. Also the assumption that future transitions are completely independent of the past is questionable in the flow of students through the higher education system.

Simulation, Branching, Network and Programming Techniques

Other models can conceivably be applied to the student flow process, such as Monte Carlo simulations,[10] certain classes of branching,[9] percalation processes,[4] network flow theory[7] and linear, non-linear and dynamic programming techniques.[21]

Most of these models have not as yet proven successful in representing student flow but they do offer some promise for future research.

Combination Models

Several models used at institutes of higher education for forecasting student demand are combinations of the above major categories of techniques.

The U.C.L.A. model[30] uses several projection techniques starting with regression methods to estimate the number of students and then projecting enrollment demands using ratio techniques.

The Rensselaer Model[2] uses a Markovian process for projections but derives the transition matrices using regression methods on previous year's data. The model is, in effect, a nonstationary Markovian model.

The University of Colorado simulation of operations model (CUSIM)[2] includes a student flow model which is a mixture of cohort survival ratios and regression smoothing.

PROBLEM DEFINITION

Thus, over the years, new and innovative techniques have beem employed to forecast student enrollments, each one adding a new dimension of sophistication and complexity. Perhaps the best-to-date model for predicting student load on departments is the one that NCHEMS at WICHE is currently testing, SFM-IA. Preliminary documentation[12] indicates that all the latest and proven modeling techniques have been incorporated into their newest model of student flow including such innovations as transition probability matrices and Markov chains.

In another recent publication,[26] the American College Testing Program (ACT) reported a study comparing five methods for projecting enrollment, including ratio, regression and two Markov models. The conclusion of this study was, among other things, that "simple and straight-forward projection models would appear to be just as useful as complex and sophisticated models."

Thus the following questions are raised: Is the NCHEMS model superior for forecasting the student demand on a department by level? Or is there another model that incorporates techniques in addition to the mathematical techniques that would improve the forecast? Is the additional expense of gathering the detail data and running the complex model justified in terms of increased accuracy in projections?

FORECASTING METHODS EXPLORED

There are many different ways in which to describe the methods used in developing forecasting models. A recent article[1] examines two aspects of this problem--the method and the type of information. The first deals with the method used to analyze the data and is labeled the "subjective-objective" dimension. The second deals with the type of information and is labeled the "naive-causal" dimension.

Subjective vs. Objective Methods

Subjective methods are those in which the process used to obtain the forecasts has not been well specified. These are the judgmental, intuitive, "seat of the pants" methods mentioned earlier.

Objective methods are those in which the process used to obtain the forecasts has been extremely well specified. These methods lend themselves well to computer processing and are so tightly defined that other researchers can replicate the method and obtain exactly the same forecasts.

Naive vs. Causal Methods

Naive methods are those which use data on only the dependent variable (e.g., number of students enrolled). Typically, an analysis is carried out to see whether the dependent variable shows any regularities over time. The time pattern is then projected into the future as shown in Figure 1 (a).

Causal methods go beyond the dependent variable to consider also variables which may cause changes in the dependent variable. An attempt is made to determine what causal variables are important, then to forecast the causal variables, and, finally, to infer values for the dependent variable on the basis of the changes in the causal variables. This process is outlined in Figure 1(b). The key assumptions are that the causal variables can

Figure 1. An Illustration of Naïve and Causal Forecasting Methods

(a) Naïve Methods

$$\boxed{Y_{t-h} \cdots Y_{t-2},\ Y_{t-1},\ Y_t} \longrightarrow \boxed{Y_{t+f}}$$

(b) Causal Methods

$$\begin{array}{ccc} \boxed{X_{t-h} \cdots X_{t-2},\ X_{t-1},\ X_t} & \longrightarrow & \boxed{X_{t+f}} \\ \downarrow & & \downarrow \\ \boxed{Y_{t-h} \cdots Y_{t-2},\ Y_{t-1},\ Y_t} & & \boxed{Y_{t+f}} \end{array}$$

Key to Symbols:

Y is the dependent variable

X is the set of causal variables

h is the number of years of historical data

f is the number of years in the future

t is the present year

From Reference (1).

be measured and projected rather accurately in comparison to a projection of the dependent variable and that the relationships will remain constant over time.

Theoretical Forecasting Models

Consider each of these two aspects of the forecasting problem as a dimension in forecasting model space, Figure 2, with subjective-objective on one axis and naïve - causal on the other. Armstrong and Grohman have labeled the intersection of the extreme points of these dimensions as shown in Figure 2.

Armstrong's and Grohman's paper draws the following conclusions:

1. Objective methods are more accurate than subjective methods.
2. Causal methods are more accurate than naïve methods.
3. The superiority of objective and causal methods increases as the forecast horizon increases.
4. Therefore, econometric methods will produce more accurate long-range forecasts than may be obtained from novice judgement, extrapolation or expert judgement.

In the context of the above analysis, the early student forecasting methods were "Expert Judgement" while the various mathematical models are "Extrapolation." According to Armstrong and Grohman, a superior long-term forecasting method would be a combination of these methods. They suggest an econometric model which would be extremely complicated, demand lots of causal analysis, require volumes of data and still not guarantee that the causal relationships can be established for student demand on departments. A realistic compromise from the econometric method would be a mathematical extrapolation modified by expert judgement. One such model has already been developed at the University of Utah and is used extensively in preparing management decision tools for budget allocation and long range planning.

Figure 2. Some Theoretical Types Of Forecasting Methods

Objective	Extrapolation	Econometric
Subjective	Novice Judgement	Expert Judgement
	Naïve	Causal

From Reference (1),

UNIVERSITY OF UTAH MODEL

At the University of Utah, a student projection model[3] has been developed that forecasts the student demand by level by department for ten years. The aggregation levels considered by the model are lower division, upper division and graduate. The model uses a ten-year student credit hour (SCH) history as a base. From this historical data both a ten-year linear regression and a current four-year regression are made on each level for all departments. These regressions are then extrapolated out ten years. The average of the two end point years is fixed as one end of the projection line and the last actual year is fixed as the other end. A linear interpolation is made between these points. Each department's SCH is forecast by this method, and the total SCH is added together by level to determine the total university SCH by level.

An independently derived student headcount ten year total university forecast by level is also input to the model. This forecast takes into consideration such factors as predicted number of high school graduates, students inclination for university training, job market demand for college graduates, drop out rates, credit by examination and other factors that affect student enrollment. This headcount forecast is converted to SCH on the basis of an average student course load by level. Then the departments are each adjusted so that the total aggragated department SCH by level equals the converted university headcount SCH total by level.

The program also has the provision for adjusting the individual departments SCH projection by level to allow for special knowledge the university administrators may have that is not reflected in historical data or in gross, overall trends, such as the phasing out of a department over the next three years, a physical space limitation that will be reached in three years, a leveling off of a high cost program, etc. This special knowledge can be input in one of three ways; actual SCH by level by year, selecting an end point level for the tenth year, or choosing the ten or four year linear extrapolation as the projection line. Of the three methods, the first should not be overlooked in importance due to the significant flexibility it allows.

A check on the validity of the data is obtained by distributing the finalized projections to the department chairmen for comments or amendments.

The above technique of allowing knowledgable administrators to alter certain of the mathematical projections is related to management by exception and is called "forecasting by exception." This added dimension to mathematical forecasting has, among others, the following advantages over pure extrapolation:

1. System makes routine projections for each level of each department. At the University of Utah, this amounts to 144 separate projections.

2. System makes fuller use of knowledge of trends, history and other available data.

3. System is consistent and predictable in its judgements, removes prejudice from forecasts.

4. Management can concentrate efforts on critical problem areas.

5. Allows crises and critical problem areas to be analyzed and adjusted by concerned, knowledgeable people.

6. Enables special knowledge of planned changes to be incorporated into the forecast, such as increasing the freshman class of the medical school by 75 students next fall.

7. Stimulates communication between various segments of the university-administrators, staff, deans, department chairmen and faculty.

PURPOSE AND VALUE OF THE STUDY

The purpose of this study is to compare the accuracy of the forecast of student demand by level on academic departments using six basic methods - judgement, ratio model, moving averages, regression model, Markov model and a combination model. (Student demand on a department is measured in terms of student credit hours SCH In addition the researcher investigated the value of combining expert judgement with the mathematical models in making the student demand forecast.

A cost - efficiency index was developed that combined the cost of running the model with the accuracy of the result. This index measures the relative cost-value of the six models studied.

This study will help institutions determine whether the cost of the increased complexity and sophistication of the forecasting model is compensated for by an increase in the accuracy of the student forecast used for budgeting and planning purposes.

SECTION III: PROCEDURE

DESCRIPTION OF THE MODELS SELECTED

The models selected for the study will now be discussed in their order of increasing complexity.

Judgement Only

The simplest and most straight-forward method of forecasting is for a knowledgeable person to sit down and estimate the numbers based on his own good judgement developed through long years of experience. This "seat of your pants" method sufficed in the past and is still in use today in many institutions.

Even though the definition of the procedure is simple, the execution of the task can be arduous. At the University of Utah, for example, there are 48 credit producing departments. When one considers 3 student levels for each department, that totals 144 decisions. If one is looking 5 years into the future, that explodes to 720 individual decisions. Even though it is time consuming, the task can be accomplished. However, one must consider the problem of human fatigue and inconsistency when estimating what the accuracy of the final results will be.

Never-the-less, the Judgement Only Model is legitimate and must be considered as a possibility, especially in times of cost-value.

Baseline

The Baseline is the next degree of complexity when considering forecasting models. The Baseline Model is an example of a ratio model. This simple ratio model forecasts next year's student level by multiplying the base figure by a ratio. Only one ratio is applied to each level in each department, but a different ratio may be used for each year in the projection horizon. Stated mathematically the Baseline Model becomes:

Yt+1 + Y*R, where Y is the SCH of a level within a department, t is the base year and R is the universal ratio.

The ratio to use for each year is based on the estimated change in total university SCH from one period to the next. This universal ratio is then applied to all levels and all department.

This model could be made more complex by selecting a different ratio for each level, or for each department, or for each level in each department. However, the latter would boil down to a Judgement Only model if applied for just one year. If one selects different ratios for each year, one again is faced with the human fatigue factor.

Exponential Smoothing

The Exponential Smoothing Model was selected to demonstrate the forecasting ability of a simple mathematical technique. This model is an example of a moving average technique and is slightly more complex than the Baseline Model.

Although no reports of using exponential smoothing for projecting student demand by level was found in the literature, many examples of applying this technique to projecting other time series data are found. One such example is the use of exponential smoothing to forecast sales.[31]

The Exponential Smoothing Model requires historical data because it is basically a moving average technique. The Exponential Smoothing Model steps through each historical year and calculates a weighted average and a trend using the exponential smoothing technique developed by Robert G. Brown.[5]

This technique calculates a moving average by using the weighted average of all past observations together with the current value as the basis for predicting the next and subsequent values. The weighted average trend of the data is also considered in predicting future data values.

The determination of how much weight to put on past data versus the current data point in predicting the next value is handled by smoothing constant, a variable with a value between 0 and 1. When the smoothing constant (alpha) is near zero, the historical data has the greatest weight. When alpha is near one, the current data has the greatest weight in predicting the next value.

The equation for calculating alpha is as follows:

alpha = 2/(N+1), where N is the number of historical data points to use in determing the next predicted value. Examples of the relationships between alpha and N are summarized below:

N	alpha
2	0.666
3	0.500
4	0.400
5	0.333
10	0.182
20	0.095

The smoothing constant can be set to have any number of years as the basis for the prediction. The alpha that provided the best results for this study put the greatest weight on the most recent four years of data or on alpha of 0.40. This is consistent with an earlier test at the University of Utah where exponential smoothing was used to predict the total university enrollment for short periods.

The exponential smoothing formula is stated mathematically as follows:

$Zt+1 = \alpha\, Yt + (1-\alpha)\, Zt$, where Yt is the current SCH value, α is the smoothing constant and Zt is the current moving average. The trend of the data is calculated using a similar formula as follows:

$\Delta = \Delta = Zt+1-Zt$

$Tt+1 = \alpha\,\Delta + (1-\alpha)\, Tt$, where Δ is the difference between the updated moving average and the current moving average, α is alpha the smoothing constant, Tt is the current moving average of the trend and Tt+1 is the predicted next value for the trend.

The next predicted value for SCH by level is calculated as follows:

$Yt+1 = Zt+1 + [(1-\alpha)/\alpha]\, Tt+1$, where Yt+1 is the next predicted value.

Linear Regression

Linear Regression is one of the class of regression analysis models. This technique is more complex than the Exponential Smoothing Model and requires a good deal of accurate historical data.

In this model, the trend of the past data is fitted to a straight line by the method of least squares. It is assumed that future values will follow the same historical trend. The trends may be short or long range.

In this study, although ten year's historical data were available, only the most recent four years of SCH by level by department were used in the regression model.

Combination - Linear Regression, Ratio and Judgement

The University of Utah Model is a combination model as detailed in a previous section. It begins with two linear regressions, adjusts these by a ratio to meet an externally determined limit based on several causal factors. These projections may also be tempered by human judgement, experience and special knowledge. These judgement factors are used on an exception basis only where deemed necessary.

The data from the University of Utah model is labeled in this report as the "University Model." When the results of the University Model output were modified by inserting judgement into the model for selected levels in selected departments, the data is labeled the "Special Knowledge" Model.

Markov

As explained earlier, the WICHE Student Flow Model is basically a Markov model using transition probability matrices and Markov chains. This is one of the most complex models currently in use, using data edits, historical analysis of data, admissions criteria and transition logic. The SFM system uses 17 programs and 14 sorts to complete its task of predicting student demand on departments by level.

Detail descriptions of the system, the inputs required, the processing logic and the reports are found in the WICHE documentation: "Introduction to the Student Flow Model SFM-IA,"[13] "Student Flow Model SFM-IA System Documentation,"[14] and "Student Flow Model SFM-IA Reports."[15]

In a nutshell, the SFM-IA system requires a file of students enrolled at the University for two time periods. The Fall Quarter enrollments were selected for this study. From these files it was determined how many students were new from one period to the next, how many students were continuing and at which level and department, and how many students left the University either by graduation or by drop-out. This data was used by the model to generate the transition probability matrix which specifies what fraction of students from one year move to a different department or level the next year and what fraction stays where they are for the next year.

In addition to the transition probability matrix, the SFM-IA also requires the starting enrollment for each department and level together with the expected number of new students of various categories, i.e., freshmen, transfers, returnees, etc.

A summary of the characteristics of each model is shown in Table I, Summary of Student Forecasting Models. In addition to the model name and type are some indicators of the complexity of the model.

The Data Required tells how extensive a file must be pulled together to operate the model. Average Cost to Run and the Number of Programs refer to the computer programs and the cost to execute the system for one year's projection. These costs are the actual cost on the UNIVAC 1108 system at the University of Utah.

The Control Parameters show what the analyst has control over when using the model. The parameters range from exact control over each number in the Judgement Only Model, to control over a single parameter in the next three models, to control of several parameters that directly effect the results of the model forecast in the last three models.

SUMMARY OF STUDENT FORECASTING MODELS

Model Name	Model Type	Data Required	Average Cost to Run	Number of Programs	Control Parameters
Judgement only	Seat of Pants	Last year's SCH by dept. by level is usually the basis	0	0	Each data point
Baseline	Ratio	Last Year's SCH by dept. by level	$.10	1	Baseline Factor
Exponential	Moving Average	At least the most recent 4 year's's SCH by dept. by level	$.15	1	Smoothing Constant, alpha (α)
Linear Regression	Regression Analysis	At least the most recent 4 year's SCH by dept. by level.	$.25	1	Number of years of historical data to use
University Model	Combination	10 year's SCH by dept. by level. Headcount forecast for total Univ. by level	$1.85	1	Number of years of data used for each of two linear regression lines, total Univ. headcount forecast by level.
Special Knowledge	Expert Judgement plus Combination	Same as Univ. Model	$1.85	1	Same as University Model plus control over any dept.'s SCH by level using one of the following: Each data point End point Trend line-one of two calculated
WICHE SFM-IA	Markov	Actual students in institution for at least 2 periods of time	$76.67	17 Programs +14 Sorts	Many -- see WICHE documentation -- some include -- number of students admitted admission acceptance ratio.

RESEARCH APPROACH

Model Implementation

As was mentioned earlier, the University of Utah model was already developed. It was originally written as a FORTRAN program and subroutines for the UNIVAC 1108 in 1968. The model was extensively modified from 1969 to 1971, but has remained fairly stable since that time. The model is used by the University several times each year to make projections for budget planning and resource allocation. The data from the system is used throughout the year.

The data from the University Model and the Special Knowledge Model are by-products of these annual runs made by the Office of Academic & Financial Planning.

The Judgement Only Model got its data from the Director of the Office of Academic and Financial Planning. During Fall Quarter each year he wrote down his experienced judgement of what the SCH for each level of each department would be for the academic year. He is perhaps the best qualified person on campus for this task because he is intimately involved with the budgets and planning of all academic departments at the University.

The Baseline Model, the Exponential Smoothing Model and the Linear Regression Model were written especially for this project. They are all in one program that has access to the same ten year historical SCH data base that is used by the University Model. The FORTRAN programs and subroutines were all thoroughly tested and checked-out prior to the experimental runs.

Several choices of the control parameters were experimented with before the ones finally used were selected as the best for the particular forecasting task.

The WICHE SFM-IA Model caused the most problems, used the most time and cost the most money to develop of any of the models. The major problems encountered are summarized in Appendix I, Implementing the WICHE Student Flow Model on the UNIVAC 1108.

The SFM-IA Model also used a different data base than the other models, requiring the file of students enrolled at the University for four consecutive years. These files were available but had to be converted to be used on the UNIVAC 1108 machine. A special COBOL program was written to take these files in consecutive pairs and create the student records specified by the SFM-IA Model.

The SFM-IA Model was implemented on the UNIVAC 1108 and was checked-out using the test data and reports provided by WICHE. All 17 programs and reports checked-out with the published results.

Experimental Plan

The research plan was to run each model using the data for three academic years. The projected SCH was to be compared to the actual SCH for that year by department and by level.

Data Analysis

A special data analysis program was written to subtract the actual SCH from the projected SCH by department by level and to examine the difference or the error. Several statistics were calculated using these differences, including the following:

a. Mean Square Error

b. Mean Absolute Error

c. Mean Cumulative Error

The Mean Cumulative Error gives an estimate of the average bias of the projection technique, i.e., is the projection always over or always under the actual value.

The Mean Absolute Error gives an estimate of the average degree of projection error.

The Mean Square Error gives an estimate of the average variance of the projection.

These differences for each department were summed into colleges and into the total University where the above statistics were calculated for the college and for the University by level. Figure 3 shows an example of the computer printout for the TOTAL UNIVERSITY using the University Model for Academic Year 1975. Each of the above statistics is shown calculated for this model.

Figure 3. Sample Report from the Student Projection Data Analysis Program

PROJECTION PROGRAMS ERROR ANALYSIS DATE 06/21/75 PAGE 30

PROJECTION PROGRAMS ERROR ANALYSIS

UNIVERSTY MODEL TECHNIQUE

CODE		DEPARTMENT NAME	ACTUAL	PROJCT	ERROR	ABS ER	ER SQUARE	LEV	MEAN CUM ER	MEAN ABS ER	MEAN SQR ER
30	0	TOTAL UNIVERSITY	398337.	390267.	-8070.	8070.	65124900.	LOW			
			251185.	241570.	-9615.	9615.	92448225.	UPR			
			108761.	146067.	37306.	37306.	1391737632.	GRD			
			758283.	777904.	19621.	54991.	384983640.	TOT			
		UNIVERSITY - SUM OF COLLEGES		17.	-16140.	38492.	193669440.	LOW	-949.	2264.	11392320.
				17.	-19230.	41974.	217201226.	UPR	-1131.	2469.	12776543.
				17.	74612.	75054.	1603595216.	GRD	4389.	4415.	94329130.
				17.	39242.	155520.	705722552.	TOT	2308.	9148.	41513091.
		STANDARD DEVIATION						LOW			3375.
								UPR			3574.
								GRD			9712.
								TOT			6443.
		UNIVERSITY - SUM OF DEPARTMENTS		48.	-8070.	53398.	2354572588	LOW	-168.	1112.	4905359.
				48.	-9615.	46827.	824247709.	UPR	-200.	976.	1717181.
				48.	37306.	41126.	1313588830.	GRD	777.	857.	2736642.
				48.	19621.	141351.	3374742704.	TOT	409.	2945.	7030713.
		STANDARD DEVIATION						LOW			2215.
								UPR			1310.
								GRD			1654.
								TOT			2652.

SECTION IV: RESULTS

ANALYSIS OF RESULTS

Mean Square Error

The statistics from the various data analysis computer runs are summarized in tables in Appendix II. Table A shows the Mean Square Error for the total University when summed by departments. This table indicates that the Baseline and Special Knowledge Models have the smallest error for different years.

Table B shows the Mean Square Error by level for the total University when summed by departments. This table indicates that generally the error is larger with lower division than upper and is larger with upper division than graduate. The Judgement Only, Baseline, Exponential Smoothing and Special Knowledge Models each performed best on at least one level during the three years.

Since the Mean Square Error values don't really mean much by themselves, it is enlightening to look at them relative to a standard. Choosing the Judgement Only Model as the standard, Table C shows the factor that each model is of that standard for the total SCH for the total University. The Baseline Model is consistently lower than Judgement Only as is the Exponential Smoothing Model.

When the factor is taken on the Mean Square Error data by level, Table D is the result. This table shows that except for the following number of times, the Judgement Only Model had the smallest mean square error out of the nine independent projections for each model:

Model	Number of Times out of Nine that Mean Square Error was less than Judgement Only Model
Baseline	3
Exponential	4
Linear Regression	2
University	2
Special Knowledge	3
WICHE SFM-IA	0

To try to get a feel for the overall comparative effectiveness of the models in relation to the Judgement Only Model, a product of the factors was taken. Table E summarizes the results from both Tables C and D.

When the three factors for each model are multiplied it produces a number that indicates the relative size of the mean square error of the model when compared to the mean square error of the Judgement Only Model as shown in Table E, all models are much greater than or are fairly close to the Judgement Only Model except for the following:

Model	Relative size of Mean Square Error with Respect to the Judgement Only Model -- Total SCH
Baseline	.25
Exponential Smoothing	.58
Special Knowledge	.43

When the three factors for each level of each model are multiplied together, the product is a number that represents the relative magnitude of.

the mean square error with respect to the Judgement Only Model. These values are shown for each Model in Table E. Note that Baseline and Exponential Smoothing both have two values less than 1.0 and Special Knowledge has one.

To produce a single number that represents the relative value of each model, the four products in Table E were multiplied to produce an overall factor. This factor measures the relative size of the mean square error of each model when compared to the mean square error of the Judgement Only Model. A value less than 1.0 would mean that most of the time the model's error was less than the error of the Judgement Only Model. The larger value would show the degree that the model's error was greater than the error of the Judgement Only Model.

Based on this overall criteria, the ranking of the models would be as follows:

Model	Rank	Overall Factor (Table E)
Baseline	1	.129
Exponential Smoothing	2	.508
Special Knowledge	3	.838
Judgement Only	4	1.000
University	5	3.947
WICHE SFM-IA	6	54.777
Linear Regression	7	66.948

In addition to the mean square error, several other measures of model accuracy were used. The following paragraphs discuss the mean cumulative error, the mean absolute error and the percentage error.

Mean Cumulative Error

The error of the projection for each level by each model was calculated by taking the projected value and subtracting the actual value, i.e.,

$$\text{Error} = \text{Projected} - \text{Actual} .$$

An average of these errors gives a measure of the degree of bias in the model. This average is the mean cumulative error. If the mean cumulative error is always negative, this means that the model is projecting smaller values than needed. If the mean cumulative error is always positive, this means that the model is projecting values larger than actual too much of the time. The ideal is to have small values for the mean cumulative error that are about equally positive and negative.

Table F is a summary of the mean cumulative error for the models in this experiment. The average absolute value of these errors, the average possible errors, the average negative errors, and the number of positive and negative values is given below:

Mean Cumulative Error

Model	Positive	Negative	Avg. Positive	Avg. Negative	Avg. Absolute
Judgement Only	4	5	342.5	120.8	219.3
Baseline	4	5	330.5	260.4	291.6
Exponential Smoothing	5	4	377.6	220.3	307.7
Linear Regression	5	4	586.0	220.3	423.4
University	5	4	237.0	211.3	225.6
Special Knowledge	5	4	304.0	182.8	216.3
WICHE SFM-IA	5	4	506.6	187.5	254.1

The above table shows that all of the models are evenly matched with positive and negative values, therefore no model bias is suspected. However, the average positive values are higher than the average negative values for each model indicating a tendency to forecast high.

The average absolute values would rank the models as follows:

Model	Rank	Average Absolute Mean Cumulative Error
Special Knowledge	1	216.3
Judgement Only	2	219.3

University	3	225.6
WICHE-IA	4	251.1
Baseline	5	291.6
Exponential Smoothing	6	307.7
Linear Regression	7	423.4

Mean Absolute Error

The mean absolute error is calculated by taking the difference between the projected value and the actual value by level and then by averaging the absolute value of these differences. This statistic gives a measure of the degree of error of the projection from the actual. The advantage of this statistic over the mean cumulative error in measuring the magnitude of the projection error is that in the mean cumulative error a large negative error and a large positive error would cancel each other out and make the mean cumulative error look small. In the mean absolute error, all deviations from the actual value are taken into account.

Table G is a summary of the mean absolute error for the models in this experiment. The average mean absolute error for each model is given below:

Model	Average Mean Absolute Error
Judgement Only	812.3
Baseline	738.8
Exponential Smoothing	787.3
Linear Regression	993.1
University	822.2
Special Knowledge	769.2
WICHE SFM-IA	967.8

The average mean absolute error values would rank the models as follows:

Model	Rank	Average Mean Absolute Error
Baseline	1	738.8
Special Knowledge	2	769.2
Exponential Smoothing	3	787.3
Judgement Only	4	812.3
University	5	822.2
WICHE SFM-IA	6	967.8
Linear Regression	7	993.1

Percentage Error

The percentage error reveals what percentage of the total value the error is, e.g., for lower division SCH in the accounting department, the error was 2% of the lower division SCH for that year. This measure of error is valuable because it reports the error relative to the actual value. An error of 4 compared to a value of 10 is much more significant than an error of 4 compared to a value of 100. (The difference is 40% error compared with 4% error).

Table H is a summary of the percentage error for the models in this experiment. The range of error for the models goes from 0.11% to 37%. The table quickly reveals that all the models had trouble predicting the 1975 graduate level SCH. Excluding these values, as outliers, the average percentage error for each model is given below:

Model	Average Percentage Error
Judgement Only	4.68
Baseline	5.43
Exponential Smoothing	5.52
Linear Regression	7.40
University	3.00
Special Knowledge	2.74
WICHE SFM-IA	8.54

The average percentage error values would rank the models as follows:

Model	Rank	Average Percentage Error
Special Knowledge	1	2.74
University	2	3.00
Judgement Only	3	4.68
Baseline	4	5.43
Exponential Smoothing	5	5.52
Linear Regression	6	7.40
WICHE SFM-IA	7	8.54

CONCLUSIONS

The purpose of this study was to compare the accuracy of the forecast of student demand by level on academic departments using six basic methods and to select the best model. The study was also to determine if the cost of the more complex models can be justified in terms of increased accuracy of the forecasts.

The results of the study presented in the previous section are summarized below showing how each model was ranked by the various error analysis criteria:

	Rank			
Model	Mean Square	Mean Cumulative	Mean Absolute	Percentage
Judgement Only	4	2	4	3
Baseline	1	5	1	4
Exponential Smoothing	2	6	3	5
Linear Regression	7	7	7	6
University	5	3	5	2
Special Knowledge	3	1	2	1
WICHE SFM-IA	6	4	6	7

To get a feel for the overall rank, add the rankings above and compare. The results are shown below:

Model	Accumulated Rank	Summary of Rankings
Special Knowledge	1	7
Baseline	2	11
Judgement Only	3	13
University	4	15
Exponential Smoothing	5	16
WICHE SFM-IA	6	23
Linear Regression	7	27

The above table suggests three groupings for the models. The first is the Special Knowledge Model which appears superior to the others. The second grouping is the Baseline Judgement Only, University Model and Exponential Smoothing where there seems to be no significant difference between the models. The third grouping includes the WICHE SFM-IA and Linear Regression which appears to be inferior to the other models studied based on the chosen criteria.

To determine the relative cost-effectiveness of the models, an index was developed which is the product of a measure of effectiveness and a measure of cost. The effectiveness measure chosen was the average of the by-level mean square error data. This is the prime accepted measure of error for the study.

The measure of cost selected was the computer operating cost. It is recognized that all the models require some human time to set up and prepare the data for a run. Even the Judgement Only Model requires some human time. It was assumed that for the models the human time was not significantly different among the models and could be eliminated from the selection criteria. The remaining cost is the computer cost which can be obtained directly off the run sheet and is, thus, readily available.

The average mean square error for each model and the average cost are summarized in Table I. The calculated cost-effectiveness index is also shown there.

Model	Rank	Index
Baseline	1	1.61
Judgement Only	2	1.79
Exponential Smoothing	3	2.60
Linear Regression	4	6.83
Special Knowledge	5	32.75
University	6	37.00
WICHE SFM-IA	7	1,901.42

From the above table, the WICHE SFM-IA model is definitely not a contender in the cost effectiveness contest having both a relatively large error and a large cost.

The conclusion of the study, then, is that a model that combines the objective extrapolation with the causal expect judgement produces the best results in forecasting the student demand on academic departments by level as demonstrated in the Special Knowledge Model.

The study has also shown that simple models provide just as good if not better forecasts than do complex models. If the cost of executing the model is considered, the simple model should be chosen over the complex one every time.

RECOMMENDATIONS

The research reported here showed the value of combining expert judgement with mathematical models in forecasting student demand. The research also pointed out the value of simple models.

The next logical step to experiment with is to use expert judgement in some simpler models such as Baseline and Exponential Smoothing. Earlier in the report it was suggested that the universal factor in the Baseline Model could be replaced by different factors for each level or by different factors for each department. Perhaps different factors could be used for selected departments or selected levels as is done in the Special Knowledge Model.

Quite a bit of work has been done with the Exponential Smoothing Model where the control parameter is alpha, the smoothing constant. Several models have been developed to change alpha based on certain calculated criteria. Perhaps a model could be developed to change alpha based on expert judgement for certain departments or for certain levels.

BIBLIOGRAPHY

1. Armstrong, J. Scott and Grohmon, Michael C., "A Comparative Study of Methods for Long-Range Market Forecasting," Management Science - Application, Vol. 19, No. 2, October, 1972, pp. 211-221.

2. Bailey, D. E., "Proposal for a simulation Model of the University of Colorado," University of Colorado, Boulder, Colorado, 1970.

3. Blake, R. John and Parks, Steve L., Student Credit Hour Projection at the University of Utah, Academic and Financial Planning, University of Utah, Salt Lake City, Utah, January 1973.

4. Broadbent, S. R. and Hammersly, J. M., "Percolation Processes I and II." Proceedings of the Cambridge Philosophical Society, Vo., 53, pp. 629-645.

5. Brown, Robert G. and Mayer, Richard F., "The Fundamental Theorem of Exponential Smoothing," Operations Research, Vol. 9, No. 5, Sept. - Oct. 1961, pp. 673-685.

6. Firmin, Peter A.; Goodman, Seymour S., Henricks, Thomas E.; and Linn, James J., University Cost Structure and Behavior, Tulane University, New Orleans, Louisiana, 1967.

7. Ford, L. R., Jr. and Fulkerson, D. R., Flows in Networks, Princeton, New Jersey, 1962.

8. Gani, J., "Formulae for Projecting Enrollments and Degrees Awarded in Universities," Journal of the Royal Statistical Society, Series A, Vol 126, Part 3, 1963, pp. 400-409.

9. Harris, Theodore E., The Theory of Branching, Prentice-Hall, New York, 1964.

10. International Business Machines Corporation, "A General Purpose Digital Simulator and Examples of its Application" Parts, I, II, III, IV, IBM Systems Journal, No. 3, No. 1, Armonk, New York, 1964, pp. 21-56.

11. Johnson, Richard S., NCHEMS Student Flow Model SFM-IA: An Introduction, National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, Boulder, Colorado, May 1972.

12. Johnson, Richard S., NCHEMS Student Flow Model SFM-IA: Users Guide, National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, Boulder, Colorado, May 1972.

13. Johnson, Richard S., Introduction to the Student Flow Model SFM-IA, Technical Report No. 41A, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education, Boulder, Colorado, May 1974.

14. Johnson, Richard S. and Busby, John C., Student Flow Model SFM-IA. System Documentation, Technical Report No. 41B, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education, Boulder, Colorado, May 1974.

15. Johnson, Richard S., Student Flow Model SFM-IA Reports, Technical Report No. 42, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education, Boulder, Colorado, June 1974.

16. Judy, Richard W.; Levine, J. B.; and Center, S. I. Campus V. Documentation, Vols. 1-6, Systems Research Group, Toronto, Canada, 1970.

17. Koenig, Herman E.; Keenry, M. G., and Zemach, R. "A Systems Model for Management Planning and Resource Allocation in Institutions of Higher Education." Michigan State University, East Lansing, Michigan, 1968.

18. Lins, L. J. Methodology of Enrollment - Projections for Colleges and Universities, Madison, Wisconsin, University of Wisconsin.

19. Lovell, C. C., Student Flow Models: A Review and Conceptualization, Technical Report 25, National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, Boulder, Colorado, August 1971.

20. Marshall, K. T., Oliver, R. M. and Suslow, S. S., "Undergraduate Enrollments and Attendance Patterns, "University of California Administrative Studies Project in Higher Education, Report No. 4, Berkeley, 1970.

21. Menges, G. and Elstermann, G. "Capacity Models in University Management," XVII International Conference of the Institute of Management Science, Vol. 17, London, July 1970.

22. Office of Program Planning and Fiscal Management, "Higher Education Enrollment Projection Model," State of Washington, Olympia, 1970.

23. O'Grady, William D. and Feddersen, Alan P., A Student Flow Model for California State College, Fullerton: A Feasibility Study, Division of Analytic Studies, Office of the Chancellor, The California State Colleges, Los Angeles, California, January, 1972.

24. Oliver, R. M., "Models for Predicting Gross Enrollments at the University of California," Ford Foundation Research Program in University Administration, Report No. 68-3, University of California, Berkeley, 1968.

25. Oliver, R. M. and Marshall, K. T., "A Constant Work Model for Student Attendance and Enrollment," Ford Foundation Research Program in University Administration, Report No. 69-1, University of California, Berkeley, 1969.

26. Orwig, M. D.; Jones, Paul K.; Lenning, Oscar T.; Enrollment Projection Models for Institutional Planning, The American College Testing Program, Research and Development Division, ACT Research Report No. 48, Iowa City, Iowa, January 1972.

27. Peat, Marwick, Mitchell and Company, CAP: SC - Computer Assisted Planning for Small Colleges, Project Report - Phase I, Peat, Marwick, Livingston and Company, New York, May 1969.

28. Perl, Lewis, J., and Katzman, Martin T. "Student Flows in California System of Higher Education," Office of the Vice President, Planning and Analysis, University of California, Berkeley, California.

29. Robertson, Leon B.; Andrew, Loyd D.; Blake, R. John; Instructional Cost Report, Academic and Financial Planning, University, Utah, Salt Lake City, Utah, February 7, 1973.

30. Smith Wayne; A Student Flow Model, Office of Advanced Planning, University of California at Los Angeles, Los Angeles, 1970.

31. Winters, Peter R., "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, Vol. 6, No. 3, April 1960, pp. 324-342.

32. Young, Andrew and Almond, Gwen, "Predicting Distributions of Staff," Computer Journal, No. 3, Jan. 1961, pp. 246-250.

APPENDIX

I. Implementing the WICHE Student Flow Model on the UNIVAC 1108.

II. Data Tables

A. Mean Square Error for Total SCH of the Total University Summed by Department.

B. Mean Square Error by Level for SCH of the Total University Summed by Department.

C. Factor of Each Model's Mean Square Error for total SCH of the Judgement Only Model.

D. Factor of each Model's Mean Square Error for SCH by Level of the Judgement Only Model

E. Product of the Factors in Tables C and D for Three Years.

F. Mean Cumulative Error by Level for SCH of the Total University Summed by Department.

G. Mean Absolute Error by Level for SCH of the Total University Summed by Department.

H. Percentage Error of SCH Projection for SCH of the Total University Summed by Department.

I. Cost - Effectiveness Index for Student Projection Models.

APPENDIX I

IMPLEMENTING THE WICHE STUDENT FLOW MODEL ON THE UNIVAC 1108

The WICHE Student Flow Model was implemented on the University of Utah UNIVAC 1108 computer in the Spring of 1975. Several actions are required of the systems analyst to get the series of programs up and running. Some of these actions were anticipated and documented by WICHE and some were unexpected.

Below is outlined the types of problems encountered in getting the 17 ANS COBOL programs of the Student Flow Model to run on the UNIVAC 1108 machine:

I. Software and documentation delivered late

II. Actions documented by WICHE

  A. Modification to the ENVIRONMENT DIVISION

  B. Requirement of 14 sorts

III. Unanticipated problems

  A. Corrections required to get a clean compile

    1. TOP-OF-PAGE

    2. REDEFINES

  B. Corrections required to get an error-free execution

    1. Improper labeling of sort control fields
    2. Referencing data items in closed files
    3. Remove Segment Feature from some programs.

A detail description of each of the above actions required to get the system running is given below:

I. Software and Documentation Delivered Late. The NCHEMS Student Flow Model SFM-IA was originally scheduled by WICHE to be released in October 1973. I received a copy of the "Type II : NCHEMS early Release Programs" and "Preliminary Draft for Review Purposes Only" system documentation in mid-

Appendix I (continued)

September 1974, almost one full year after the promised date. The difference between the original and the actual delivery date has raised havoc with the scheduled completion of this project.

In addition to the late delivery, these "Early Release Programs" contain a warning that "software will be programs that have not yet been adequately tested or documented for release as Type I software." The following paragraphs discuss in detail the problems encountered using these programs. Tracking and correcting the errors caused at least a two month delay in this project completion.

II. Actions Documented by NCHEMS

A. Modification to the Environment Division. Part of the WICHE documentation warned that certain changes must be made in the ENVIRONMENT DIVISION. The following are examples of cards that were input to correct each of the 17 programs.

```
SOURCE-COMPUTER                 UNIVAC - 1108
OBJECT-COMPUTER                 UNIVAC - 1108

SELECT    REPORT-FILE       ASSIGN TO PRINTER
SELECT    OLD-SFM-FILE      ASSIGN TO UNISERVO NEWFOL.
SELECT    USER-MAIN-FILE    ASSIGN TO CARD-READER
```

B. Requirement of Fourteen Sorts. The SFM-IA system requires the user to supply 14 sorts to operate. A standard UNIVAC 1100 STANDALONE SORT/MERGE package was used for this task. Following is an example of the sort control cards:

```
@ RUN
@ HDG SORT10     STANDALONE SORT   HISTORY MODULE
@ ASG,A   NEWF10., F2
@ ASG,T   NEWS10., F2
```

```
@ UUCC* . SORT MERGE. SORT
    FILEIN = NEW F10
    KEY = 1, 37, A, A, 1
    FILEOUT = NEWS 10
@ FIN
```

III. Unanticipated Problems.

In addition to the above actions documented by NCHEMS that are required to get the system operating, several actions were required to get a clean compile and others were required to achieve a successful execution of the programs.

A. Corrections Required to get a Clean Compile. The following cards were changed to get an error free compile using the @ ANSCOB compiler on the UNIVAC 1108:

1. TOP-OF-PAGE

Delete the following card from the ENVIRONMENT DIVISION:

```
SPECIAL-NAMES          IS    TOP-OF-PAGE.
```

Add the following card to the DATA DIVISION:

```
77 TOP-OF-PAGE 77    PICTURE 99    VALUE 60.
```

Change the following card in the PROCEDURE DIVISION:

```
WRITE REPT-RECORD-OUT    AFTER ADVANCING TOP-OF-PAGE 77
```

2. REDEFINES

In programs SFM 40, 45, 50, 70 and 75.

An 05 level REDEFINES statement did not reference back to the original 05 level data name but instead referenced a data name which was a redefinition of the original 05 level data name.

46

Appendix I (continued)

The following is an example of the changes that were made to fix this problem flagged by the compiler:

```
05  OLD-SPP1-DATA   REDEFINES   OLD-SFM-DATA
05  NEW-SPP1-DATA   REDEFINES   NEW-SFM-DATA
```

B. Corrections Required to Get an Error Free Execution. Once a clean compile was achieved, more energy was expended trying to get a successful run using the WICHE test data set. Following are examples of this kind of problem:

1. Improper Labeling of Sort Control Fields. NCHEMS mailed out a program change in December 1974 updating program SFM01 by moving spaces to NEW-SFM-RECORD. However, when executing SFM02 after SORT02, a vital control record was flagged as missing and SFM02 aborted. This record is the one that assures that SFM01 executed properly.

   It was found that the record was created by SFM01 but it was located in the file out of the sort sequence expected by SFM02. The NCHEMS program change put spaces in the sort control fields when the system required that vital sort data such as "iteration" code and "term number" be in those fields.

   Once these data were inserted back in the records, the sort (SORT01) pulled all the control records to the front of the file where program SFM02 expected them to be located SFM02 now execute properly.

2. Referencing Data Items in Closed Files. In programs SFM20, 25, 40 45, 55, 60, 70, 75, and 80, the program referenced a data name in a

file prior to the file being opened and it also referenced a data name in a file after the file was closed. Both of those conditions caused program execution errors and aborted the run. The programs were changed so that the UPDATE-FILE in SFM20 and the OLD-SFM-FILE SFM25, SFM40, SFM45, SFM55, SEM60, SFM70, SFM75 and SFM80 are opened first thing in the program and closed the last thihg. These changes allowed a successful execution of these programs.

3. Remove Segment Feature. In several programs mysterious execution errors were occurring that the programming and systems people of the University could not explain. After the segmentation feature was taken off these programs, they executed using the WICHE test data and printed the same reports as the documentation illustrated.

APPENDIX II

DATA TABLES

TABLE A. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN SQUARE ERROR* - TOTAL

$*(10^5)$

Technique	Academic Years Projected 1973	1974	1974
Judeement Only	37.5	41.4	74.0
Baseline	20.4	23.6	60.4
Exponential Smoothing	31.0	34.5	62.4
Linear Regression	80.6	40.9	88.9
University Model	51.2	30.7	70.3
Special Knowledge	46.1	29.9	36.6
WICHE SFM-IA	56.6	44.2	82.6
Model with Smallest Error	Baseline	Baseline	Special Knowledge

TABLE B. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN SQUARE ERROR* BY LEVEL

*(10^5)

Technique	Level	Academic Years Projected 1973	1974	1975
Judgement Only	Lower	24.9	16.1	47.8
	Upper	6.2	23.7	18.1
	Grad	1.5	5.8	16.9
Baseline	Lower	10.9	19.4	39.8
	Upper	6.3	25.0	13.4
	Grad	1.6	6.1	22.5
Exponential Smoothing	Lower	19.1	26.1	37.2
	Upper	7.0	31.1	9.9
	Grad	1.6	5.5	18.5
Linear Regression	Lower	53.3	30.6	67.4
	Upper	11.8	28.9	17.7
	Grad	3.1	6.5	26.5
University Model	Lower	27.4	19.5	49.1
	Upper	9.0	21.2	17.2
	Grad	1.8	7.2	27.4
Special Knowledge	Lower	28.6	18.2	33.1
	Upper	7.9	19.9	17.4
	Grad	1.7	7.1	25.7
WICHE SFM-IA	Lower	39.8	28.6	57.9
	Upper	12.7	33.4	20.4
	Grad	2.5	8.1	19.5
Model with the Least Error	Lower	Baseline	Judgement	Knowledge
	Upper	Judgement	Knowledge	Smoothing
	Grad	Judgement	Smoothing	Judgement

TABLE C. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN SQUARE ERROR - TOTAL FACTOR OF JUDGEMENT ONLY BY YEAR

Technique	Academic Years Projected		
	1973	1974	1975
Judgement Only	1.00	1.00	1.00
Baseline	.54	.57	.81
Exponential Smoothing	.83	.83	.84
Linear Regression	2.15	.99	1.20
University Model	1.37	.74	.95
Special Knowledge	1.23	.72	.49
WICHE SFM-IA	1.51	1.07	1.12

TABLE D. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN SQUARE ERROR BY LEVEL FACTOR OF JUDGEMENT ONLY BY YEAR BY LEVEL

Technique	Level	Academic Years Projected 1973	1974	1975
Judgement Only	Lower	1.00	1.00	1.00
	Upper	1.00	1.00	1.00
	Grad	1.00	1.00	1.00
Baseline	Lower	.44	1.20	.83
	Upper	1.02	1.05	.74
	Grad	1.07	1.05	1.33
Exponential Smoothing	Lower	.77	1.62	.78
	Upper	1.13	1.31	.55
	Grad	1.07	.95	1.09
Linear Regression	Lower	2.14	1.90	.78
	Upper	1.90	1.22	.98
	Grad	2.07	1.12	1.57
University Model	Lower	1.10	1.21	1.03
	Upper	1.45	.90	.95
	Grad	1.20	1.24	1.62
Special Knowledge	Lower	1.15	1.13	.69
	Upper	1.27	.84	.96
	Grad	1.13	1.22	1.52
WICHE SFM-IA	Lower	1.60	1.78	1.21
	Upper	2.05	1.41	1.13
	Grad	1.67	1.40	1.15

TABLE E. PRODUCT OF THE FACTORS FOR THREE YEARS IN TABLES C AND D

	Table C	Table D		
Technique	Total	Level	Product	Overall
Judgement Only	1.000	Lower	1.000	1.000
		Upper	1.000	
		Grad	1.000	
Baseline	0.249	Lower	.438	.129
		Upper	.793	
		Grad	1.494	
Exponential Smoothing	0.579	Lower	.973	.508
		Upper	.814	
		Grad	1.108	
Linear Regression	2.554	Lower	3.171	66.948
		Upper	2.271	
		Grad	3.640	
University Model	0.963	Lower	1.371	3.947
		Upper	1.240	
		Grad	2.411	
Special Knowledge	0.434	Lower	.900	.838
		Upper	1.024	
		Grad	2.095	
WICHE SFM-IA	1.810	Lower	3.446	54.777
		Upper	3.266	
		Grad	2.689	

TABLE F. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN CUMULATIVE ERROR BY LEVEL

(Error = Projected - Actual)

Technique	Level	1973	1974	1975
		Academic Years Projected		
Judgement Only	Lower	529	112	-47
	Upper	-201	261	-184
	Grad	-37	-135	468
Baseline	Lower	180	99	-67
	Upper	-381	333	-299
	Grad	-69	-486	710
Exponential Smoothing	Lower	411	520	9
	Upper	-256	417	-165
	Grad	-46	-414	531
Linear Regression	Lower	873	434	-377
	Upper	-75	640	-91
	Grad	268	-338	715
University Model	Lower	-168	60	-168
	Upper	-309	195	-200
	Grad	28	125	777
Special Knowledge	Lower	-112	52	-267
	Upper	-238	216	-114
	Grad	58	43	847
WICHE SFM-IA	Lower	716	638	-235
	Upper	-194	374	-187
	Grad	113	-134	692

TABLE G. TOTAL UNIVERSITY - SUM OF DEPARTMENTS MEAN ABSOLUTE ERROR BY LEVEL

Technique	Level	Academic Years Projected 1973	1974	1974
Judgement Only	Lower	1111	1096	1201
	Upper	629	1168	716
	Grad	273	497	620
Baseline	Lower	701	718	1026
	Upper	590	1146	879
	Grad	242	558	789
Exponential Smoothing	Lower	992	1045	1042
	Upper	622	1265	717
	Grad	247	515	641
Linear Regression	Lower	1522	1143	1509
	Upper	782	1249	957
	Grad	395	587	794
University Model	Lower	988	812	1112
	Upper	693	1090	976
	Grad	293	579	857
Special Knowledge	Lower	895	789	1021
	Upper	711	968	915
	Grad	260	461	903
WICHE SFM-IA	Lower	1364	1114	1486
	Upper	681	1189	1012
	Grad	382	614	868

TABLE H. TOTAL UNIVERSITY - SUM OF DEPARTMENTS PERCENTAGE ERROR

Technique	Level	Academic Years Projected 1973	1974	1975
Judgement Only	Lower	6.35	4.75	5.62
	Upper	3.81	5.28	3.96
	Grad	1.48	6.19	10.11
Baseline	Lower	2.15	1.23	.81
	Upper	7.22	6.89	5.72
	Grad	2.78	16.66	31.35
Exponential Smoothing	Lower	4.93	6.44	.11
	Upper	4.84	8.63	3.16
	Grad	1.86	14.19	23.45
Linear Regression	Lower	10.47	5.38	4.55
	Upper	1.41	13.23	1.75
	Grad	10.82	11.57	31.56
University Model	Lower	2.02	.75	2.03
	Upper	5.86	4.04	3.83
	Grad	1.14	4.29	34.30
Special Knowledge	Lower	1.96	1.62	3.21
	Upper	4.91	3.11	2.18
	Grad	2.08	2.87	37.38
WICHE SFM-IA	Lower	8.91	6.85	5.64
	Upper	4.06	15.12	3.01
	Grad	9.37	15.38	24.67

TABLE I. COST - EFFECTIVENESS INDEX FOR STUDENT PROJECTION MODELS

Index = (Error) (Cost)

Average of By-Level Data Mean Square

Technique	Error	Average Cost Per Run	Index
Judgement Only	17.9	$ 0 (.10)	1.79
Baseline	16.1	.10	1.61
Exponential Smoothing	17.3	.15	2.60
Linear Regression	27.3	.25	5.83
University Model	20.0	1.85	37.00
Special Knowledge	17.7	1.85	32.75
WICHE SFM-IA	24.8	76.67	1,901.42