AN ENROLLMENT FORECASTING MODEL FOR A STATEWIDE SYSTEM OF HIGHER EDUCATION

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

VERNON RUSSELL HOFFNER, JR.

1975

This is to certify that the

thesis entitled

An Enrollment Forecasting Model for a

Statewide System of Higher Education

presented by

Vernon Russell Hoffner, Jr.

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Management

Phillip C. Carter
Major professor

Date 10/2/75

O-7639

ABSTRACT

AN ENROLLMENT FORECASTING MODEL FOR A
STATEWIDE SYSTEM OF HIGHER EDUCATION

By

Vernon Russell Hoffner, Jr.

This research is addressed to the problem of providing accurate forecasts of the statewide enrollments needed in all levels of higher education planning. Financial planning for higher education at the institutional and state levels requires some estimate of future enrollments. In addition, enrollment projections are used as a basis for the general planning and coordination of educational activities at both the institutional and state decision making levels.

A Markov student flow model was used to generate the enrollment forecasts described in this dissertation. This method has two advantages over other types of enrollment forecasting. The first is that in addition to generating total statewide enrollment forecasts, the model can be easily expanded to differentiate between groups of institutions or classifications of students. For example, forecasts were generated for several different groups of baccalaureate institutions and community colleges. Forecasts of community college enrollments were also separated into transfer and non-transfer classifications. These classifications were used to distinguish between the group of

students most likely to transfer to a baccalaureate institution and those students enrolling in the growing area of vocational-technical programs.

The second advantage is that it uses aggregate enrollment data in order to estimate the substochastic Markov transition matrix for the enrollment forecasting model. The estimation technique utilizes quadratic programming to obtain a least squares estimate of the transition probabilities. This is an improvement over the maximum likelihood technique which is normally used to estimate Markov transition probabilities. The maximum likelihood technique requires individual student data for the estimation process, which would be very expensive and impractical--if not impossible--to obtain when applied to a statewide system of education. The historical data required by the quadratic programming procedure has been and is currently collected by several agencies, one being the Higher Education General Information Survey. This availability of data means that no additional data collection activities are required in order to accurately forecast statewide enrollments.

The Markov student flow enrollment forecasting model was validated by using it to generate forecasts that could be compared to actual historical enrollments. The quality of the forecasts was evaluated with the use of three measures of the discrepancy between the predicted and actual enrollments. The error measures were the relative error, the average error in units of the forecast, and Theil's Inequality Coefficient.

Vernon Russell Hoffner, Jr.

The forecasting model is divided into two phases. Phase I forecasts the number of students entering the system of higher education by applying the Markov student flow concept to the statewide elementary and secondary school system. The application of the model to the forecasting of first-time entrants begins with the forecasting of statewide elementary and secondary enrollments. The flow concept is used to forecast the first-time entrants by tracing the flow of students through the elementary and secondary grade levels and into the system of higher education. The mean difference between the predicted enrollments and the actual enrollments was zero at the .05 level of significance for the two states tested, Michigan and New York.

Phase II uses the forecast of the first-time entrants from Phase I as the input into the Markov flow process used to forecast the statewide higher education enrollments. This phase was also validated by comparing predicted enrollment levels with actual historical enrollment levels. In addition, two other methods for forecasting higher education enrollments, the population ratio method and a multiple linear regression model, were evaluated and compared with the Markov model. The forecasts generated by the population ratio method were the least accurate. The forecasts generated by the Markov student flow model and the multiple linear regression model were of comparable accuracy and both were more accurate than the population ratio forecasts.

In addition to forecasting enrollments, the model can also be used to aid educational administrators in the areas of planning and policy making. The model can be used to explore and answer various kinds of "What if?" questions that might be useful in the process of developing educational policies.

AN ENROLLMENT FORECASTING MODEL FOR A STATEWIDE SYSTEM OF HIGHER EDUCATION

By

Vernon Russell Hoffner, Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Management

1975

ACKNOWLEDGMENTS

It is not possible to name all of the many individuals who aided in this research effort or to state precisely what each person contributed to the final product, this dissertation.

The author would like to thank his dissertation guidance committee for their valued contributions: Dr. Phillip Carter, Associate Professor of Management Science, Co-Chairman of the committee, for his suggestions and guidance in the final stages of the research and the writing of the dissertation; Dr. Patrick Toole, Assistant Professor of Engineering Design and Economic Evaluation at the University of Colorado, Co-Chairman of the committee, for his introduction of this area of research to the author and his insight into the methodological tools required in this research activity; and Dr. Richard Gonzalez, Chairman of the Department of Management, for his advice and interest.

The author would like to express his appreciation to the staff, and particularly to Dr. James Hatcher, of the Michigan Department of Education, Higher Education Planning and Coordination for their assistance to the author in gathering and understanding the data needed for the research project.

Finally, the author expresses his deepest appreciation to his wife, Nancy, for her patience and support during the long academic career culminating in this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table		Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Purpose of the Study

The objective of this research is the development and validation of a Markovian student flow model, based on aggregate, historical, cross-sectional enrollment data. The model can be used to trace the flow of students into and through a statewide system of higher education. A primary use of the model will be to forecast future enrollments of the undergraduate populations of the public institutions of higher education within any given State. The enrollment forecasts can be computed by flowing the enrollment populations through the Markov process and calculating the expected number of students in each of the states of the system at successive points in time.

The structure of the model can be as detailed as the data availability permits. At one end of the scale, the system could be aggregated into only three states, one state for the community colleges, and one each for the upper and lower divisions of the baccalaureate institutions. A more detailed structure for the model is presented in Figure 1. The model in this example consists of twelve states, one state for each of the class levels at the complex universities, the state colleges, and the community colleges. The community college is separated into two parts, one for the students taking an academic

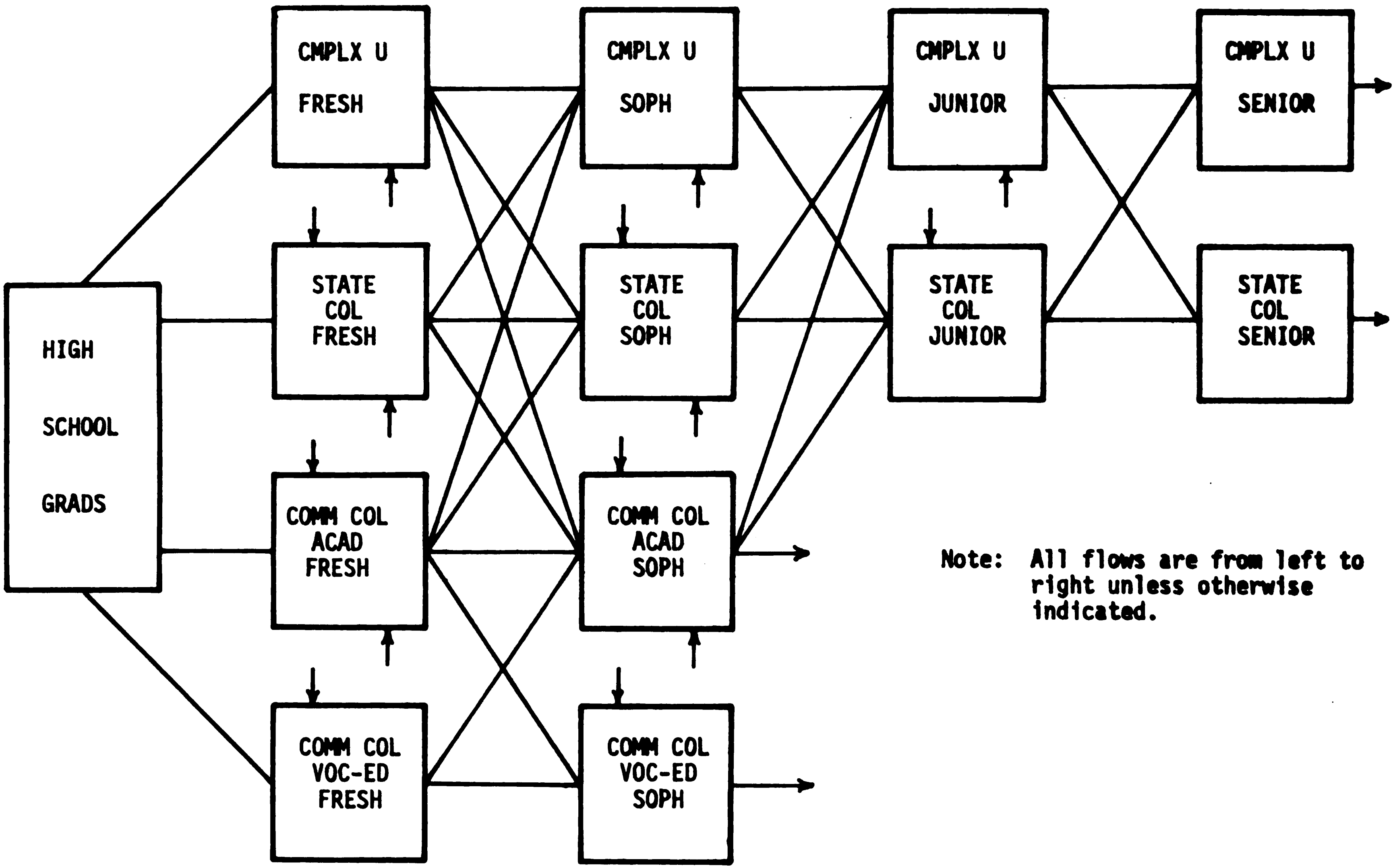

Figure 1. Student flows and enrollment levels.

curriculum with the option of transferring to a baccalaureate institution and one for the students on a terminal, vocation or occupational program. The addition of the vocational education states to what might be considered a strictly academic model is the result of the growing number of students enrolled in these terminal programs.

There are two kinds of states discussed in this dissertation. The first is the classification states of the Markov student flow model. These states refer to classification of students by class level, institution of attendance, or academic program. The second is the political or geographical State, which is capitalized to differentiate it from the Markov state.

The terms applying to institutions of higher education are defined in the following manner. A baccalaureate institution is any organization that confers a baccalaureate degree for an appropriate amount of study in a particular field of knowledge. A complex university is a baccalaureate institution that offers master's, doctoral, and professional degrees in many fields of knowledge. The primary academic emphasis at the complex university is on research and providing an academic environment for the students studying for advanced degrees. The state colleges are baccalaureate institutions that do not offer doctoral and professional degrees. They do offer some master's degrees, but the primary academic emphasis is upon teaching at the undergraduate level and teaching at the master's level where appropriate. A community college does not offer baccalaureate degrees, only associate degrees and certificates of study. Community colleges offer programs of study requiring two years, or less, of full time study.

These programs may be in the traditional fields of academic study and be acceptable toward a baccalaureate degree at any baccalaureate institution, or the programs may be in the vocational areas, such as; automobile mechanics, electronics, nursing, and dental hygiene. The vocational programs provide the students with immediately salable, occupationally related skills.

Conceptually, the model could contain one state for each of the class levels--freshman, sophomore, junior, and senior--at each of the institutions within the system. In addition, the community college classifications could be divided into academic or transfer programs and terminal programs for each of the two class levels at each institution. This would be an important additional classification since it would clearly identify those students which will probably transfer to a baccalaureate institution. In addition, it identifies the portion of the community college enrollment which has experienced the greatest growth during the last few years--the terminal, vocational-technical programs. The resulting model could be quite large with four state classifications for each institution. For example, Michigan has 29 public community colleges and 13 public baccalaureate institutions. Using the more finely divided classification system, the model for public higher education in Michigan would contain 168 states. Other States would have correspondingly large numbers of classification states in the Markov student flow system, depending on the number of institutions of higher education within the State.

The flow of students into a Markov student flow model is exogenous to the model. This is indicated in Figure 1 with the box labeled "High School Grads" and the dashed lines represent the flow of high school graduates into the various academic entry states. In order to forecast the number of students enrolled in the system of higher education, the number of students entering the system must be known or forecasted. The flow of students through the elementary and secondary school system of a State can also be modeled with the Markov student flow process. The Markov student flow model will also be used with elementary and secondary enrollments to predict the number of students that will enter the system of higher education.

The process of forecasting the number of students enrolled in the statewide system of higher education will be divided into two phases. Phase I forecasts the number of first-time entrants into the system of higher education. This is accomplished by applying the Markov student flow process to the statewide elementary and secondary enrollments. The result is a flow of students through the elementary and secondary school system, through the point of graduation and into the system of higher education. Phase II, the higher education phase, uses the forecast of students entering the higher education system with the Markov flow process to forecast the statewide enrollments. Both phases forecast enrollments on an annual basis, with forecasts being of the Fall enrollments.

Limitations

The development of a full scale model requires an extensive amount of data concerning the students enrolled at each institution. Due to the amount of data required in the estimation of the model parameters, only data for the State of Michigan were used in validating the higher education phase of the forecasting model. The data which were available for Michigan are quite limited and had to be gathered from various sources. There are three sources of institutional enrollment data for all of the institutions of higher education in Michigan. The Michigan Association of Collegiate Registrars and Admissions Officers (MACRAO), an organization of institutional administrative officers involved in the admission and registration of students at their respective institutions, has been collecting and publishing institutional enrollment data for more than twelve years. The Higher Education General Information Survey (HEGIS), a comprehensive survey of institutions of higher education conducted by the United States Office of Education, has been conducted annually since its initial survey in 1966. The Michigan Bureau of Programs and Budget (BPB) has enrollment data for each of the public institutions of higher education as a part of each institution's budget request for the last four budget request cycles. The BPB institutional budget requests provide enrollment data beginning with the 1968-69 fiscal year.

There are several problems with the data available from these three sources. This is particularly true in regard to the development of a full scale student flow model. The first problem is that each source has different definitions or methods of counting enrolled students. This results in different enrollment figures from each source for a given Fall enrollment period. In addition, the counts had different deadlines for submittal and in the case of the MACRAO data the submission date was so early that an accurate count could not be supplied. In many instances, the count of extension students at various institutions were only estimates because of the early MACRAO reporting date.

Another definitional problem results from the changes each source has made over the years in the definitions, methods, and data response forms for the collection of enrollment data. The changes in reporting methods of each information source provides for the possibility of error within the data that would not be detectable.

The amount of detail within the data or the lack of detail is also a problem in the development of a model. The enrollment data by class level are available from the BPB for only three years. However, the information is not complete for all institutions. Several institutions did not respond with the full detail requested. The HEGIS data are divided into lower division and upper division for undergraduates for only the last four years. Only total institutional enrollment data were collected by HEGIS in the first survey. The enrollment data for 1967 and 1968 were collected on the basis of undergraduate and graduate student enrollments. The MACRAO data are separated into lower division and upper division for all years since 1966. However, extension students are counted as a separate category and some portion of these students are undergraduates.

The most difficult data problem concerns the number of transfer students and the institutional sources and destinations of these students. No data have ever been collected on a statewide basis concerning transfer students. The only method of obtaining these data would be to go directly to each institution. This would be very expensive and time consuming since many institutions only have manual files on their students. As a consequence, the procedure to be followed in regard to transfer students will be to estimate the transfer rate, based on the available data.

The above problems with data availability may mean that a fully developed forecasting model of student flows for a statewide system will not be possible. However, the data which are available is sufficient for the estimation of the parameters of a smaller, more aggregated model of the system. The amount of scaling down of the model will depend on the accuracy with which the model parameters can be estimated from the available data.

The data limitations may also result in a problem in validating the forecasting model. The underlying assumptions and the structure of the system can be examined from a logical viewpoint. However, with only eight annual observations of the system, it is unlikely that full statistical validation can be accomplished.

The Need for Enrollment Forecasts

The manner in which students move through the system of higher education--changing majors, leaving the system for undertermined periods of time and then returning, failing and repeating courses, continuing on for advanced and professional degrees, switching enrollment from one institution to another--can and does have a significant impact on the planning and management of individual institutions and on the planning and coordination of groups of institutions. Changes in student preferences, economic and social conditions, and requirements for manpower are influencing the enrollment patterns in higher education. The emergence of community colleges as an important sector of higher education, with an increasing emphasis on occupationally related curricula, is a major influence in the changing educational patterns. In addition, many institutions are making academic changes to accommodate the changed preferences and goals of the incoming students.

The projection of student enrollments has its basis in the budgetary processes used for colleges and universities. Implicitly, if not explicitly, the development of a budget involves some estimate of the enrollment for each institution, and the enrollment in each institutional subunit, during the coming fiscal year. In addition, the flow of students and enrollment projections are used as a basis for the general planning and coordination of activities at both the institutional and state decision making levels.

In the past two decades, the planning and management emphasis has been in placing facilities where they are needed most, managing a continually expanding curriculum and enrollments, and obtaining the necessary qualified faculty and staff. In many instances, a significant amount of the financial support for the growth and expansion of higher education has come from the state governments. The education sector of the economy has experienced a tremendous amount of growth during the 1960's, but the 1970's have brought a termination to this exceptional rate of growth. As a result, some universities and colleges are finding themselves with facilities which cannot be fully utilized, faculties with fewer than anticipated students, and unanticipated gaps between resources and demands. This situation has led to a more critical analysis of the system of higher education. At the state level the analysis has shifted to the development of a statewide picture and on long range needs. This is in addition to the analysis, and is used to place in perspective the operational needs and current financial requests of individual institutions. Higher education has become a complex system in which the institutions find themselves interacting at an increasing rate, both by design and by chance. This trend is not likely to be reversed. A large number of students begin their college education at a community or junior college and later transfer to a baccalaureate institution. Programs arise which must be allocated to a few institutions because enrollment is too small for the programs to be economically feasible in all colleges. In this context it is very important to consider the system as a whole, as well as the individual institutions, for to do otherwise entails great risks of creating unnecessary conflicts and ignoring potential complements within the system. Long range projections are especially important when planning capital outlay needs, but they are also useful in a broader sense. To the extent that the state can forecast its future needs in the area of higher education, it can more adequately determine what priorities must be set and what policies must be implemented in order to satisfy those priorities.

The structure of the system of higher education has undergone some significant changes during the past two decades. Higher education underwent a period of slow growth during the 1950's, followed by a period of rapid growth during the 1960's, which was followed by a period of almost no growth in the early 1970's. During this time the number and types of institutions grew and changed. In addition, not only did the enrollment distributions change within the system of higher education, but the goals of the students also changed. Initially, the purpose of the community college was to provide the first two years of a student's education toward his baccalaureate degree. Now the community colleges are beginning to provide more terminal educational opportunities, particularly in the vocational and technical areas, rather than areas of transfer education.

The pressures of these changes in the system of higher education and the greater competition for scarce financial resources have caused many institutions and state planning agencies to turn to planning and management systems. These planning and management systems provide a method of organizing information, utilizing the information with planning tools and techniques for educational administrators, and maintaining a framework for economic decision making in the area of higher education. One of the key inputs in all of the planning and management systems is the projection of student enrollments.

Although student enrollment forecasts are important to all planning, there has been little recent work in the area of statewide enrollment forecasting models. The primary reason appears to be that until recently the summation of the individual institutional enrollment projections was an accurate forecast of the enrollment within the statewide system. Typically, these projections reflected the constraints on the capacity of the system and the individual institutions. Enrollment is more a function of capacity than it is of availability of students for most professional and graduate programs within the system of higher education. Because of the capacity constraints, only a limited number of students can be enrolled in the professional and graduate programs. In most instances, there is sufficient educational demand to fill all of the vacant positions. This used to be true for the undergraduate programs, however, it is no longer true.

Notwithstanding the difficulties in making accurate enrollment projections during this period of change in the system of higher education, there remains a need to organize the data and make as good enrollment forecasts as possible. Without reasonably good enrollment forecasts, the limited financial resources of the state governments could be budgeted and allocated to educational areas where an educational demand might not materialize. The model proposed in this dissertation is an attempt to bring to bear new concepts and techniques on statewide enrollment forecasts in a systematic way.

Overview

Chapter II describes several of the techniques that have been used to analyze and forecast student enrollments. The early research was directed toward identifying the causitive factors influencing higher education enrollments. The more recent research has been directed toward the development of institutional enrollment forecasting models. Many of these later models have a Markov structure.

Chapter III discusses the application of a Markov student flow model to a statewide system of higher education. The statewide Markov student flow model is formulated. The use of quadratic programming for the estimation of the Markov transition probabilities is also described.

Chapter IV applies Phase I, the first-time entrants forecasting phase, of the model to forecasting elementary and secondary enrollments. The results indicate that Phase I is a valid forecasting procedure. This phase is then used to forecast the first-time enrollments for the State of Michigan.

Chapter V completes the development of the higher education student flow model with Phase II, the higher education enrollment projection phase. Forecasts of the total enrollments for the State of Michigan are generated up to 1980. These forecasts are compared to the forecasts generated by two other methods, the population ratio method and a multiple linear regression model. Following the comparison and evaluation, the Markov model is expanded to differentiate between several different groupings of institutions. The enrollment forecasts for these institutional groups are compared with actual historical enrollments and evaluated.

CHAPTER II

HIGHER EDUCATION ENROLLMENT PROJECTION MODELS

The importance of enrollment models and projections is the result of the need for accurate forecasts in the planning at all levels of higher education. As a consequence, many individuals and organizations have constructed enrollment projection models. The Federal Government, several educational research organizations, probably every state, and nearly every college make some attempt to forecast enrollments. This chapter will review the models that have been used to forecast enrollments and those models that have been proposed as enrollment projection models. The review will be organized into four sections: 1) simple linear models, 2) institutional models, 3) national models, and 4) statewide models. In addition, some of the forecasting models are submodels within larger resource allocation models. The remainder of the models are used to provide inputs into some resource allocation decision making process.

Simple Linear Models

Ratio Models

The model used by the United States Office of Education to project enrollments is representative of the early models used to forecast enrollments.[1] The ratio of enrollments to an age group considered to be the most typical among college students, usually the 18 to 21 year old age group in the case of undergraduates, is computed for several recent years. The trend in this ratio is then assumed to continue at a constant pace. This assumption, in combination with a projection of the population of the age group, is the basis for the enrollment forecasts. The resultant forecast is the product of the projected population multiplied by the projected rate of enrollment. This approach is also used by the Michigan Department of Education.[2] This method has several benefits: the computation of the forecast is quite easy, the method is simple to explain to policy makers, and the data requirements are minimal. However, the simplicity of this method does not take into consideration any factors that may cause the enrollment to vary other than the age group most likely to attend college.

Linear Regression Models

Economists have identified several factors that have had an influence on the level of enrollments.[3] Linear regression analysis was the primary method used in the identification process. In many of the studies, the observations were from national data over periods of twenty or more years. The results of the various investigations were not used to forecast future college enrollments, but just to explain the changes in enrollment levels. In addition to confirming the size of the relevant population group as a significant factor, it was determined that aggregate demand for higher education in the United States was positively related to disposable income per family and negatively related to tuition rates.[4] It was also verified that the discharges and changes in the size of the armed forces also affected the total United States college enrollments.[5] Other factors relevant to the decision to enroll in an institution of higher education that have been identified are parental characteristics, student ability, student financial aid, unemployment, and student location.[6]

Although the above reported studies were not oriented toward the development of enrollment projection models, their results were used in the development of a linear regression forecasting model of college enrollments for the State of Michigan.[7] The factors considered important within the forecasting model were the number of people in the 18 to 21 year old age group, the number of people discharged from the armed forces, income levels, and the unemployment rate.

Institutional Models

Most of the enrollment projection models are based on the analysis of cross-sectional enrollment data. The structure of the model and the values of the model parameters are the result of the observation of the number of students in particular class levels at a given point in time. The flow rates of the students from one class level to another is only observed over one time interval. No attempt is made to track the flow of students over time periods of more than one time interval.

Another basis for enrollment projection models is the analysis of cohort data. The model parameters are estimated by observing the paths of students in a selected cohort as they progress through their academic programs. A cohort is a group of students all having a common characteristic. Usually the characteristic that is common to the group of students is the time of entry into the academic system. For example, the students entering Michigan State University as freshmen during the Fall, 1975 term would be a cohort. The first institutional model described below is based on the use of cohort data; the following two institutional models use cross-sectional data for estimating the model parameters.

The Oliver Models

The cohort survival technique is used in the model developed over a period of years at the University of California by Oliver, *et al.*[8] The cohort survival technique consists of determining the pass and failure probabilities for a given cohort of students as they progress through the educational system. The assumption made in using this approach is that if the cohort is selected correctly, the resulting pass and fail probabilities will be constant from one time period to the next. The earliest form of this model used was the Grade Progression Ratio method. The method generated progression ratios for each level within the university. The model has the following mathematical form.[9]

$$z_1(t+1) = a_{1,1}\ z_1(t) + y_1(t+1)$$

$$z_{j+1}(t+1) = a_{j,j+1}\ z(t) + y_{j+1}(t+1)$$

where $a_{1,1}$ is the fraction of freshmen who return to that level in the next time period, $a_{j,j+1}$ is the ratio of progression for the students moving from the j^{th} to the j+1st level in the next time period, and $y_j(t)$ is the number of new admissions to grade j during the t^{th} period.

$y_j(t)$ is the number of new admissions to grade j during the t^{th} period.* These ratios are used to predict the enrollments, $z_j(t)$, for the future by the above first order difference equations.

The application of the Grade Progression Ratio method to the four undergraduate class levels results in[10]

$$z(t+1) = Az(t) + y(t+1), \qquad (2.1).$$

where z(t) is the enrollment vector, y(t) is the admissions vector, and A, the progression ratio matrix is given by

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{12} & 0 & 0 & 0 \\ 0 & a_{23} & 0 & 0 \\ 0 & 0 & a_{34} & 0 \end{bmatrix}$$

The original Grade Progression Ratio method does not include explicitly the possibility of a student remaining in the same class level. Oliver extended the model to include this possibility. The model structure remains the same as Equation 2.1, except that all the diagonal elements of A will be nonzero. This change in the model structure results in

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{12} & a_{22} & 0 & 0 \\ 0 & 0_{23} & a_{33} & 0 \\ 0 & 0 & a_{34} & a_{44} \end{bmatrix}$$

---

* The notation used in various equations through this dissertation will consist of capital letters for matrices and lower case letters for vectors. A matrix element will be indicated by a double subscript. Vector elements will be indicated by a single subscript. The letter t will always be used in parenthesis to signify a point in time for the associated vector or matrix.

Later versions of the model have the same form as Equation 2.1, but they have been expanded from four states to five states and later to eight states. The concept of the amount of work that the student has performed toward his degree is also included in these models.[11] The student makes a decision to either attend, vacation, or drop out of school, conditioned on the amount of work completed toward his degree. As a result of this expansion, the five state model contains the four class levels of the earlier model plus one state for those students that are vacationing.[12]

The last formulation of this model was expanded to include the graduate levels at the university.[13] In addition, the model was formulated so it could be used for planning the resource requirements at the university. The eight cohorts identified in the eight state flow model for the whole university were:[14]

1. Students who complete studies in the lower division.
2. Students who graduate with a bachelor's degree.
3. Students who graduate with a master's degree.
4. Students who graduate with a doctoral degree.
5. Lower-division students who drop out.
6. Upper-division students who drop out.
7. Students working toward a master's degree who drop out.
8. Students working toward a doctoral degree who drop out.

The output of this sector of the model was used as input to a second sector that determined the number of instructional staff required in each of the following categories:[15]

1. Teaching assistants--these must be registered graduate students and may therefore come from cohorts 3, 4, 7, or 8.

2. Nontenured regular faculty--these may be doctoral graduates (cohort 4) or they may come from external sources.

3. Tenured regular faculty, none of whom come directly from student cohorts.

The NCHEMS Model

The Oliver models started from the point of developing an adequate student enrollment projection model and then progressed toward the development of a resource requirements prediction model. Several groups have taken the reverse approach. They have developed resource requirements prediction models and one submodel or sector of the model predicts the level of enrollments. A widely publicized resource requirements prediction model has been developed at the National Center for Higher Education Management Systems (NCHEMS) at the Western Interstate Commission for Higher Education.[16] The enrollment sector of this model separates the student flow process into two major components.[17] The first component produces, from historical institutional data, estimates of the number of applicants for admission and the number of admitted students who enroll at the institution. The second component models the flow of the students through the institution. It uses the new enrollment provided by the admissions component and the previous enrollment to project the future enrollment by level and major. The model has the general form

$$x(t+1) = P\,x(t) + y(t+1),$$

where $x(t)$ is the enrollment vector, $y(t)$ is the admissions vector, and P is the Markov transition matrix. The NCHEMS student flow model was specifically designed for short term projections and to be used as a submodel within the resource requirements prediction model.

Michigan State University Model

The model developed at Michigan State University by Koenig, *et al.* is also a resource allocation model, with the enrollment

projection portion as only one sector of the model.[18] This particular model is a state-space systems model of the university. The student sector of the model predicts the student population by class level and major area of study at future points in time. This computation is based on past and present enrollments, available financial aids for the students, and the prediction of the incoming student population. This student forecasting model has the same general structure of

$$x(t+1) = P\ x(t) + y(t+1).$$

The specific formulation of the MSU model of the student sector of the university is:

$$s(t) = P(t)\ s(t-1) + a(t)\ n(t) + K_1(t)\ g(t) + K_2(t)\ h(t).^{19} \qquad (2.2)$$

Where s(t) is the state vector of the student sector whose components represent the number of students in the various levels of study and areas of education. The vector s(t) depends on the enrollment in the previous year, s(t-1), the enrollment choices of the new students--represented by the product, a(t) n(t), of the number of new arrivals entering their respective categories--the available assistantships g(t), and the available fellowships, scholarships, and the other financial aid h(t). The matrix P(t) is the transition matrix that represents the proportion of students moving from one field and level classification to another during one time period and $K_1(t)$ and $K_2(t)$ are the matrices that represent the effectiveness of the various financial aids in attracting and retaining students. The parameters of Equation 2.2 were estimated and verified with the use of the standard linear regression model.

National Models

The Gani Model

The earliest work on Markov student flow models was done by Gani in studying the enrollments of the universities of Australia.[20] Gani's purpose was to develop a model of the university system in Australia that would accurately project the total enrollments in the system and the number of graduates of the system. The assumption for this Markov model is that the pass and repeat rates at the universities can be determined. The basic structure of the Markov model developed by Gani has the same structure as the preceding Markov models,[21]

$$x(t+1) = P\ x(t) + f(t).$$

Cohort data for the years 1955 to 1960 were used to determine the transition probabilities of P.

Manpower Models

Two other national Markovian student flow models have been developed. A model of the British educational system was developed by Armitage, et al. and a model of the Norwegian system was developed by Thonstand.[22] Both of these models describe more than just the school enrollments and the educational process. The Armitage model includes state classifications for teachers at all levels of the educational system and a classification for people working outside of the educational system. In this respect it can be classified as a simplified manpower model. The Thonstad model can more readily be classified as a national manpower model, with a sophisiticated educational sector.

The Armitage description also includes a detailed discussion of the effects of bottlenecks or constraints within the educational system.[23] The bottleneck is the result of not having enough positions for the number of students applying for admission. An example of this is the situation with medical schools in the United States where there are more qualified applicants than there are position available. The effect of this problem on the model is to cause those applicants not accepted to go to the next lower opening within the system. The resulting model, including the bottleneck processes, can be used in the planning and decision making activities for the whole educational system. However, at this point the model is no longer Markovian in structure.

Statewide Models

The HEEP Model

A statewide student flow model has been developed by the Office of Program Planning and Fiscal Management in the State of Washington.[24] The Higher Education Enrollment Projection Model (HEEP) is purportedly a Markov model concerned with undergraduates. However, the description of the model is not that of a Markov process. It uses matrix notation to describe the status of the system at different points in time, i.e., the number of students in the various states. The use of a matrix structure does not necessarily result in a Markov student flow model.

The HEEP model includes the public and private baccalaureate institutions and all the community colleges within the State. The state classifications within the model are one for each of the community colleges and one for each of the four undergraduate classes at the baccalaureate institutions. The model is separated into three components. The entrance component computes the number of new students in the Washington system for each of the classifications. The transfer component calculates the number of transfers from each institution to every other institution within the system and the retention component computes the number for each institution that will remain at that institution. The results of the computations of these components are the number of students enrolled in each classification and for the total system for a given time period.

The Perkins and Paschke Model

A second model is reported by Perkins and Paschke.[25] This model is oriented toward the projection of the cost of operating the statewide higher education system. The model is:[26]

> . . . concerned with the development and the use of a simulation model to estimate the costs of various policy alternatives involving post-secondary education in a state. The model has been used to investigate the effect upon expenditures of changes in exogenous variables . . . .

The model was used to predict the operating and construction costs for each college and university in the State of Indiana. All public and private baccalaureate institutions were included. The student enrollment projection subsector of the model generates projections for each institution. The enrollments are projected using one of three methods, based upon the classification of the institution. The first method uses regression analysis to identify the explanatory variables for each institution in the first group. The second method uses the historical trend to predict enrollments at a second group of institutions. The projections for the remaining colleges and universities were taken from an outside source.[27] These enrollment predictions are then used in projecting the future costs of the system of higher education under several different policy alternatives.

Summary

Many techniques have been used to forecast student enrollments at the various levels of application. Figure 2 displays a matrix classification of the techniques and areas of application. Although work has been done on each type of model, much work remains to be done in using the models for forecasting student enrollments. Many of the models have a Markov like structure or are Markov student flow models. The algebraic structure of the Markov model appears to fit well the conceptualization of flow of students through the academic system. The application of the Markov student flow model to the forecasting of statewide enrollments still remains to be pursued. This void is evident from Figure 2, where the vacant intersection of the Markov row and State application column indicates that no reported research in this area was found. The remainder of this dissertation develops and applies the Markov student flow model to forecasting statewide enrollments. The validity of the model is tested with enrollment data from academic institutions within Michigan.

Technique	National	State	Institutional
	APPLICATIONS		
Ratio	U.S. Dept. HEW	Mich. Dept. of Ed.	
Linear Regression	References 3,4,5 and 6	Mich. Dept. of Ed.	
Markov	Gani Thonstad		NCHEMS MSU
Other	Armitage	HEEP Perkins and Paschke	Oliver

Figure 2. Application/Technique matrix of enrollment forecasting models.

Chapter II Notes

[1]U.S., Department of Health, Education, and Welfare, Office of Education, National Center for Educational Statistics, Projections of Educational Statistics to 1970-1980 (Washington, D.C.: Government Printing Office, 1971).

[2]Michigan, Department of Education, State Plan for Higher Education in Michigan (Lansing, Michigan: 1969).

[3]Robert Campbell and Barry N. Siegel, "The Demand for Higher Education in the United States," American Economic Review, Vol. 57 (June, 1967), pp. 482-494.

Harvey Galper and Robert M. Dunn, Jr., "A Short-Run Demand Function for Higher Education in the United States," Journal of Political Economy, Vol. 77 (September-October, 1969), pp. 765-777.

[4]Campbell and Siegel, "The Demand for Higher Education."

[5]Harvey Galper and Robert N. Dunn, Jr., "A Short-Run Demand Function."

[6]A. J. Corazzini, D. Dungan, and H. G. Grabowski, "Determinants and Distributional Aspects of Enrollment in U.S. Higher Education," Journal of Human Resources, Vol. 7, No. 1 (Winter, 1972), pp. 39-59.

Stephen A. Heenack and Paul Feldman, "Private Demand for Higher Education in the United States," The Economics and Financing of Higher Education in the United States, a Compendum of Papers submitted to the Joint Economic Committee, Congress of the United States (Washington, D.C.: Government Printing Office, 1969), pp. 375-395.

R. Radner and L. S. Miller, "Demand and Supply in U.S. Higher Education: A Progress Report," American Economic Review, Vol. 60 (May, 1970), pp. 326-334.

[7]Michigan, Department of Education, Financial Requirements of Public Baccalaureate Institutions and Public Community Colleges (Lansing, Michigan: 1971).

[8]R. M. Oliver, Models for Predicting Gross Enrollments at the University of California, Report No. 68-3, Ford Foundation Research Program in University Administration (Berkeley, California: 1968).

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

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

R. M. Oliver, D. S. P. Hopkins, and R. Armascost, An Academic Productivity and Planning Model for a University Campus, Report No. 3, University of California Administrative Studies Project in Higher Education (Berkeley, California: 1970).

R. M. Oliver and D. S. P. Hopkins, "An Equilibrium Flow Model of a University Campus," Operations Research, Vol. 20, No. 2 (March-April, 1972), pp. 249-264.

[9]Oliver, Models for Predicting, p. 5.

[10]Ibid., p. 17.

[11]Oliver and Marshall, A Constant Work Model, p. 1.

[12]Marshall, Oliver, and Suslow, Undergraduate Enrollments, p. 22.

[13]Oliver and Hopkins, "An Equilibrium Flow Model," p. 253.

[14]Ibid.

[15]Ibid.

[16]David G. Clark, et al., Introduction to the Resource Requirements Prediction Model 1.6, Technical Report No. 34A, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education (Boulder, Colorado: 1973).

[17]C. C. Lovell, Student Flow Models: A Review and Conceptualization, Technical Report No. 25, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education (Boulder, Colorado: August, 1971).

[18]Herman E. Koenig, M. G. Keeney, and R. Zemack, A Systems Model for Management, Planning, and Resource Allocation in Institutions of Higher Education (East Lansing, Michigan: Michigan State University, 1968).

[19]Ibid., pp. 21-26.

[20]J. Gani, "Formulae for Projecting Enrolments and Degrees Awarded in Universities," Journal of the Royal Statistical Society, A126 (1963).

[21]Ibid., p. 406.

[22]Peter Armitage, Cyril Smith, and Paul Alper, Decision Models for Educational Planning (London: Allen Lane the Penguin Press, 1969).

Tore Thonstad, Education and Manpower (Edinburgh: Oliver and Boyd, 1969)).

[23]Armitage, Decision Models, p. 30.

[24]Office of Program Planning and Fiscal Management, *Higher Education Enrollment Projection Model* (Olympia, Washington: 1970).

[25]William C. Perkins and Paul E. Paschke, "A Simulation Model of the Higher Education System of a State," *Decision Sciences*, Vol. 4, No. 2 (April, 1973), pp. 194-215.

[26]*Ibid*., p. 195.

[27]*Ibid*., p. 198.

CHAPTER III

MODEL FORMULATION

The models described in the preceding chapter provide examples of the methods that have been used to project future levels of higher education enrollments. Each model could be used in a decision process where funds and resources are allocated to various institutions. However, the decision process relies primarily on the classification of the student population within the system of higher education. This classification of the student population can be defined more specifically as: the modeling of the students' movement through the educational system, the determination of the distribution of the students throughout the educational system, and the description of the students within the educational system.

The Markov student flow model satisfies the needed classification of the student population. The Markov process identifies the flow of students through the system. Each state indicates the description of the students within it and the total number of states identifies the distribution of the students. In addition, the internal structure of the particular application can change without changing the mathematical formulation of the Markov student flow model. Within the basic formulation of the model, the states of the model can be enumerated in any way necessary to accurately and adequately describe a particular statewide system of higher education. One method of enumeration could be the classification of the students within the state by their class levels, regardless of the institution they are attending, i.e., freshman, sophomore, junior, senior, or graduate. This would be a rather simple classification scheme. The other extreme of detail could separate the student states into the following types of classifications: (1) class level; (2) institution of attendance; (3) current credit hour load; (4) State of home residence; and (5) major area of study.

The class level classifications can be freshman, sophomore, junior, senior, master, doctoral, and other. Applying the model to the State of Michigan, and including the 47 private colleges, the 29 community colleges, and the 13 public colleges and universities, 89 different institutional classifications are obtained. The student credit hour load can be used to separate the students into part-time and full-time classifications. The residence classifications could be divided into 52 distinct elements, one for each of the United States and the District of Columbia, and one for all foreign students. The major areas of study could result in many different classification elements, a typical breakdown for areas of study could be business, engineering, science, liberal arts, education, and other. If a model containing all of the above detail were developed, it would consist of 388,752 different states.*

A model that is highly simplified may not provide enough information for decision making. On the other hand, a model that is highly

---

*This number of classification states is the product of the 7 class levels, the 89 institutions, the 2 credit hour classifications, the 52 States of home residence, and the 6 areas of study.

detailed may provide more information than can be meaningfully used in the decision process. The highly detailed model may not be relevant to the decision process or it may be impossible to collect all the required data. In addition, the cost of data collection for the very detailed model may be prohibitive and far outweigh the benefits that can be derived from the model.

Model Structure

The mathematical formulation of a statewide student flow model has the same structure as the Markov models in the preceding chapter. The formulation of the system of higher education can be stated as

$$x(t) = x(t-1)\,P + f(t), \qquad (3.1)$$

where $x(t)$ is the state vector at time t, $f(t)$ is the vector of entrants to the system at time t, and P is the transition matrix. Each element of P, p_{ij}, is the probability of moving from state i at time $t-1$ to state j at time t. The elements of the state vector are the numbers of students in a particular class level and institution or group of institutions at a particular point in time. The transition probabilities of P are the probability of movement from one class level/institution state at time t, to another class level/institution at time $t+1$. In addition, the total number of students in the system can be calculated by

$$N(t) = \sum_i x_i(t),$$

where $N(t)$ is the total enrollments at time t.

Phase II of the model is now in a form that can be used to project student enrollments at the different institutions and class levels. If the values of f(t) are known or can be forecast, then the successive values of x(t) can be easily computed, assuming that P is known.

The forecasts of f(t) can be carried out with a model similar to the one just discussed for the system of higher education. The model for forecasting the first-time entrants has two stages. The first stage forecasts the number of high school graduates. This can be achieved by using the student flow model to flow the students through the elementary and secondary school system. The model structure consists of fourteen states, representing the grades kindergarten through twelve and the high school graduates. The model formulation is

$$y(t) = y(t-1)\, Q + k(t) \qquad (3.2)$$

where y(t) is the fourteen element state vector at time t, Q is the 14 x 14 transition probability matrix, and k(t) is the vector of entries into the school system. The result of this process is the potential entrants into the system of higher education. The forecast number of high school graduates is

$$g(t) = y_{14}(t).$$

The next stage of the model projects the number of first-time entrants as a function of the high school graduates. This can be accomplished by letting

$$f(t) = r\, g(t)$$

where r is the probability vector for high school graduates to enter each of the entrant states of the higher education system.

Phase I of the model can now be used to forecast the first-time entrants necessary for Phase II. The structure of the two phases is illustrated in Figure 3.

There are several simplifying assumptions that could be made in the process of estimating the values of various parameters of the model. However, there are at least two assumptions that have to be made. The first is that the system is closed. The data is available only for individual States, but there will be entrants coming from outside the State, and some of the graduates will leave the State to attend college elsewhere. It will be assumed that these two flows counterbalance each other. The second assumption relates to the fact that some students do not enroll in college directly after graduating from high school. It will be assumed that this lagged portion of entrants does not have a significant effect on the estimation of future entrants. This is due to the assumed balancing effect of the delayed entrances over succeeding time periods. These assumptions are necessary since the reported figures for first-time enrollments only indicate the total first-time enrollments at each institution.

The basic assumptions of Equation 3.1 are that all flows not explicitly stated will be assumed to be either not significant or counterbalanced by a flow in the opposite direction. The result of this assumption will be that the transition probabilities for these flows can be assumed to be zero. For example, the flow of students dropping out of school temporarily will be balanced by those students

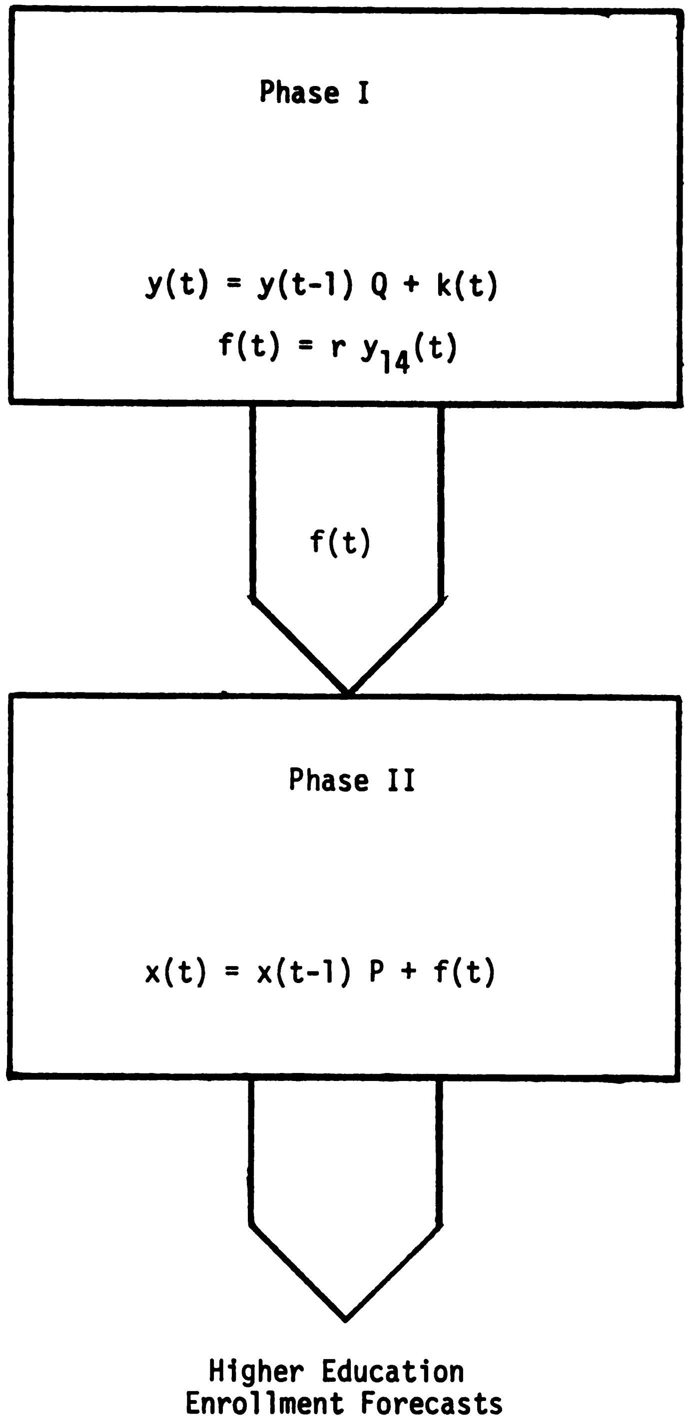

Figure 3. The two phase Markov student flow model.

resuming their studies after a temporary dropout. An example of a flow that would not be significant would be the flow of students from the community college vocational education programs to the state universities.

Two assumptions must also be made in the estimation of the parameters of Equation 3.2. The first assumption is that the elements of the vector k(t) are zero except for kindergarten students. The second assumption is that either the students can pass to the next grade level, or they must remain in their current grade level. This implies that Q is a diagonal and off diagonal matrix, i.e.,

$$q_{i,j} = 0 \text{ except where } j = i \text{ or } j = i+1.$$

These assumptions are necessary since the school enrollment numbers only indicate the number of students in each grade level.

Although, the above assumptions limit the detail of the forecasts by the model, they allow the development of a forecast from the data. The assumptions and aggregation that are necessary in the development of the model result from the paucity of detailed and accurate data.

Parameter Estimation

Entrant Distribution

There are three parameters in the model that have to be estimated from historical data. These parameters are the transition matrices, P and Q, and the probability vector r. The value of r can be estimated by,

$$r = \frac{1}{T} \sum_{t=1}^{T} \frac{f(t)}{g(t)}$$

The probability vector r is just the average of the ratio of flow of the high school graduates into the various entry states for the system of higher education.

Transition Probabilities

The problem with using the Markov process as a forecasting technique lies in the estimation of the transition probabilities. There are two methods of estimating the Markov transition probabilities, the maximum likelihood technique and the least squares method.

The maximum likelihood method requires the collection of data at the level of the individual student. The formulation for the estimation procedure for the maximum likelihood technique is

$$q_{i,j} = \frac{\sum_{t=2}^{T} \frac{y_{i,j}(t-1,t)}{y_i(t-1)}}{T - 1}$$

where $y_{i,j}(t-1,t)$ is the number of students that move from grade i at time t-1 to grade j at time t, $y_i(t-1)$ is the total number of students in grade i at time t-1, and $q_{i,j}$ is resulting estimate of the transition probability, averaged over the observed time periods.

The use of the maximum likelihood method requires a count of the number of individuals that move from any particular grade level to any other grade level in the succeeding time period. At the statewide level, the collection of the data needed to perform a maximum likelihood estimate of transition probabilities would be very difficult and expensive. This makes the use of the maximum likelihood technique for estimating the Markov transition probabilities highly infeasible.

Using the least squares technique of estimating the Markov transition probabilities from aggregate data would appear to be the reasonable approach to take. This approach does not use the actual counts of individuals in the various states at different points in time, but rather uses the relative frequency or proportion in each state at the different points in time to estimate the transition probabilities.[1] However, the underlying assumption is that a transition can be made from any one state to any other state. This assumption is invalid when considering student flow systems where the student remains in his current class level or passes onto the next class level. Neither the maximum likelihood method nor the least squares method is applicable for estimating the Markov transition probabilities when only aggregate data is available and there is a restriction on the possible transitions.

The approach that is used in this study to estimate the transition probabilities is a modified least squares method. The actual number in each state of the system is used in the estimation procedure and restrictions on the allowed transitions are permitted. This method utilizes quadratic programming to estimate the Markov transition probabilities. Quadratic programming minimizes a least squares objective function subject to the Markov probability constraints as constraints. The resulting quadratic programming formulation for Q is

$$\min \sum_{t=2}^{T} \sum_{i} [x_j(t) - \sum_{i} x_i(t-1)q_{ij} - f_j(t)]^2$$

$$\text{s.t.} \quad \sum_{j} q_{ij} \leq 1 \quad \text{for all } j$$

$$q_{ij} \geq 0 \quad \text{for all } i,j,$$

where Q is a diagonal and off diagonal matrix with each row having only nonzero elements, $q_{i,i}$ and $q_{i,i+1}$. Solving the above quadratic programming formulation results in the Markov transition probabilities for the transition matrix Q.

The quadratic programming formulation for P is

$$\min \sum_{t=2}^{T} \sum_{j} [x_j(t) - \sum_{i} x_i(t-1)\, p_{ij} - f_j(t)]^2$$

$$\text{s.t.} \quad \sum_{j} p_{ij} \leq 1 \quad \text{for all } j$$

$$p_{ij} \geq 0 \quad \text{for all } i,j.$$

The structure of P will depend on the structure of the system of higher education being modeled.

The quadratic programming problems for P and Q can be solved with the quadratic programming algorithm developed by Wolfe.[2] The resulting Markov transition matrix may be substochastic.[3] The row probabilities are not required to sum to one, but are allowed to sum to less than one. This structure of P and Q ignores the absorbing state for students dropping out of the educational system.

The quadratic programming method for estimating the transition probabilities has two advantages over the previously described methods. The first is that aggregate data can be used to estimate the transition probabilities. This eliminates the need to collect individual transition data necessary for the maximum likelihood technique. This is a particularly important advantage when estimating transition probabilities for a large system like a statewide system of higher education. The cost of collecting the individual transition data from each of the schools in the system would be very prohibitive to the use of the model. In many cases it would be necessary to manually search a school's records to obtain the required data.

The second advantage is that the transition directions can be limited by the quadratic programming constraints. The least squares method assumes transitions from any one state to any other state. However, a flow model has only a limited number of transitions and these transitions cause the individuals in the system to flow through a particular sequence of states. The maximum likelihood technique could be used to estimate the flow model transition probabilities if the individual data were available. The two advantages of the quadratic programming method make it a usable procedure for estimating the transition probabilities for large flow systems where only aggregate flow data are available.

Model Validation

The general approach to the validation of forecasting models is to use the model to forecast actual observed historical values. This is the approach that will be used in the validation of each of the two phases of the forecasting model. This approach will also be used in the comparison of each phase against other forecasting techniques.

When comparing forecasted results against actual observations, it is desirable to be able to establish the quality of the forecast. Three measures will be used in the evaluation of the forecasts. The first measure will be the relative discrepancy between the predicted and actual values, where P_i and A_i are the predicted and observed values, and R_i is the relative error for that pair of values.

$$R_i = \frac{P_i - A_i}{A_i}$$

The second measure is the prediction error in units, $P_i - A_i$, of the forecast for each pair of predicted and actual values. The third measure is Theil's Inequality Coefficient.[4] The inequality coefficient (U) of the pairs (P_i, A_i) is

$$U = \left(\frac{\sum(P_i - A_i)^2}{\sum A_i^2}\right) 1/2$$

Forecasts generated by both phases of the higher education enrollment forecasting model will be evaluated with the above three measures which provide an indication of the quality of the forecasts. The mean for each of the first two measures is computed for each forecast year. Each mean is tested to determine if it is significantly different from zero. Theil's Inequality Coefficient is also computed for each forecast year. The Coefficient is only used as a method of evaluating the forecasts generated by different methods for a given year.

Chapter III Notes

[1]Henri Theil and Guido Rey, "Quadratic Programming Approach to the Estimation of Transition Probabilities," Management Science, Vol. 12, No. 9 (May, 1966), pp. 714-721.

T. C. Lee, G. G. Judge, and A. Zellner, Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data (Amsterdam: North-Holland Publishing Co., 1970).

[2]P. Wolfe, "The Simplex Method for Quadratic Programming," Econometrica, Vol. 27, No. 3 (July, 1959), pp. 382-398.

George Hadley, Nonlinear and Dynamic Programming (Reading, Massachusetts: Addison-Wesley Publishing Co., Inc., 1964), Chapter 7.

[3]William Feller, An Introduction to Probability Theory and Its Applications, Vol I (3rd ed.; New York: John Wiley & Sons, Inc., 1968), p. 400.

[4]Henri Theil, Applied Economic Forecasting (Amsterdam: North-Holland Publishing Co., 1966), Chapter 2.

CHAPTER IV

FORECASTING FIRST-TIME ENTRANTS

The Markov models described in Chapter II did not consider forecasting the first-time entrants within the model. They either assumed the first-time entrants would be known, or they forecast the first-time entrants from some other source outside of the Markov student flow model. This can be a problem. Without having a valid forecast of first-time entrants, the use of a Markov type forecasting model, through which students would flow, would be an incomplete type of forecasting. In order to accurately forecast the enrollments for universities and community colleges on a statewide basis, knowledge of the number of students entering the system is a necessary requirement. As Oliver[1] has shown, the total enrollment for the baccalaureate institutions depends upon the number of entrants to the institutions at a time prior to that particular year. That is, the enrollment in a particular grade or class level in a given year depends upon the enrollments in the preceding year for the preceding grade. For example, the number of juniors in a selected year is primarily dependent on the number of sophomores in the preceding year. In addition, the number of freshmen in any given year depends significantly upon the number of first-time entrants for that year. This result indicates that in order to accurately forecast the enrollments of a particular institution or group of institutions, the number of first-time entrants into those institutions is of critical importance.

When considering a full scale statewide model, there are two approaches that can be taken in forecasting the first-time entrants. One approach would be to forecast the individual first-time entrants at each institution and then to sum these to determine the number of first-time enrollments for the total statewide system. The second approach would be to forecast the total number of possible entrants into the statewide system of higher education and then to allocate those entrants amongst the schools on some prorated basis. This pro-rational basis could be determined on the basis of past history plus forecasted growth for each of the various institutions. The second approach appears to be the best when considering a statewide system of higher education. The underlying rationale is that there is a limited number of individuals available to enter the system at any given time. The number of students entering a system of higher education is a random variable. The sum of the independent estimates of the first-time entrants into the individual institutions could be greatly different from the single estimate of the number of students entering the total system of higher education. Also, we are considering here a somewhat competitive situation between the various institutions. They are all competing in the same market, i.e., they are all competing for the same students who will be entering the system of higher education. In this respect, the estimate of the number of students entering a given institution is not independent of the estimates of the number of students entering the other institutions.

Use of the Model

The procedure used for estimating first-time enrollments is similar to the procedure for estimating the total enrollment in the system of higher education. This approach entails the estimation of a substochastic, Markov transition matrix for the flow process of the elementary and secondary school system. Using this approach, we can forecast the number of students in each grade as they progress through a statewide elementary and secondary school system. This method allows us to collect aggregate data at a very low cost which should lead to accurate forecasts of first-time enrollments.

The first-time enrollments phase of the enrollment forecasting model is tested with enrollment data from Michigan. The statewide enrollments for public and non-public schools have been gathered in the State of Michigan for some period of time. These numbers can be used to estimate the transition probabilities needed for the Markov process in order to forecast the flow of students through the system of elementary and secondary education within the state and this will give us an estimate of the number of students that will be entering the system of higher education.

Methods of Forecasting Elementary and Secondary Enrollments

Phase I of the student flow model could be used to forecast the elementary and secondary enrollments for a statewide system. The process of validating Phase I of the model compares the results of the elementary and secondary forecasts with the forecasts generated by a different procedure.

The method that is currently used for forecasting enrollments on a statewide basis for elementary and secondary schools is the progression ratio technique. This technique has the same basic mathematical structure as the student flow models in Equation 2.1 and Equation 3.2.

The progression ratio technique is the method used by various agencies to estimate the transition matrix for the student flow model. It has the form

$$q_{i,i+1} = \frac{\sum_{t=2}^{T} \frac{y_{i+1}(t)}{y_i(t-1)}}{T-1}$$

where $q_{i,i+1}$ is the estimated flow rate from grade i to grade i+1. This technique uses several years of enrollment data to estimate the ratio, $q_{i,i+1}$ between the preceding year's enrollment for a particular grade level $y_i(t-1)$, and the enrollment for the next grade level in the succeeding year, $y_{i+1}(t)$. The average of the T-1 ratio observations is the estimate of the transition ratio. For example, using five or six years of data for the second and third grades, the ratio of third to second grade enrollments is calculated for each pair of years. The average of these ratios for this time period is calculated and becomes the estimate of the progression ratio for the flow of students from the second to the third grade. Using this method the progression ratios for all grades, kindergarten through the twelfth, can be estimated. The progression ratio technique has the advantage that it is easy to understand, easy to implement, and provides usable forecasts.

An alternative method that can be used to forecast elementary and secondary enrollments is Phase I of the higher education enrollment forecasting model. Eliminating the step that forecasts the first-time entrants into the system of higher education, Phase I becomes a forecasting model for statewide elementary and secondary enrollments.

These two methods of forecasting statewide elementary and secondary enrollments are compared and evaluated using the procedure outlined in Chapter III.

Validation and Evaluation

The two forecasting methods were tested with Michigan and New York elementary and secondary public school enrollment data. The primary reason this data is used in the testing procedure was that the data is readily accessible. The Michigan enrollment data is from the Michigan Statistical Abstract.[2] The New York enrollment data is from the New York Statistical Yearbook.[3]

Michigan Elementary and Secondary Enrollments

The first test will compare the results of the two methods using the Michigan public school data. Michigan public school enrollment data from the years 1964 to 1969 were used to estimate the values of Q. The results of the quadratic programming estimation process appears in Table 1. These values are used in Equation 3.2,

$$y(t) = y(t-1)\,Q + k(t), \qquad (3.2)$$

along with 1969 enrollment data to forecast the enrollments for 1970, 1971, and 1972. These forecasts appear in Table 2.

TABLE 1. Markov student flow model transition matrix generated from aggregate Michigan public school enrollments, 1964 to 1969.

.000	.930											
	.000	.886										
		.059	.941									
			.043	.957								
				.046	.954							
					.047	.953						
						.049	.951					
							.079	.921				
								.083	.917			
									.145	.855		
										.157	.843	
											.086	.914
												.000

TABLE 2. Markov student flow model forecasts of Michigan public school enrollments for the years following 1969.

YEAR		K	1	2	3	4	5	6
1969	A	183918	173173	166998	162471	161131	163882	161641
1970	P		171044	163284	164131	162897	161421	164051
	A	180938	172142	165576	165876	163005	161168	165365
	R		-.006	-.014	-.011	-.001	-.002	-.008
1971	P		168272	161179	160708	164567	162990	161873
	A	169197	169601	167075	164340	166707	163658	162636
	R		-.008	-.035	-.022	-.013	-.004	-.005
1972	P		157353	158599	158579	161368	164657	163262
	A	163248	162600	167319	165653	166410	169475	166705
	R		-.032	-.052	-.043	-.030	-.028	-.021

YEAR		7	8	9	10	11	12
1969	A	160486	154533	158640	154843	139503	122082
1970	P	165448	160634	164710	159948	142530	127506
	A	166511	161261	164343	160266	142200	126063
	R	-.006	-.004	.002	-.002	.002	.011
1971	P	169083	165710	171184	165938	147093	130272
	A	169535	166567	170830	165144	147924	129142
	R	-.003	-.005	.002	.005	-.006	.009
1972	P	167299	169479	176778	172415	152536	134443
	A	169779	171427	177474	172744	154386	134884
	R	-.015	-.011	-.004	-.002	-.012	-.003

P=Predicted Values
A=Actual Values

R= Relative Error $\frac{P-A}{A}$

The forecasts were generated in the following manner. Since the number of first-time entrants is not known, the number of kindergarteners in the preceding year is used for k(t), the number of entrants into the first through twelfth grade system. The enrollments for 1969 are the known starting point for the forecast. The enrollments for 1970 are forecast through the use of Equation 3.2, the estimated values of Q, and the known enrollment from 1969. Iterating through Equation 3.2, the resulting forecast for 1970 is used to forecast the 1971 enrollments and then the 1971 forecast is used to predict the 1972 enrollments. The following equations illustrate the iterative process of the forecast.

$$\begin{aligned} \hat{y}(1970) &= y(1969)\ Q + k(1970) \\ \hat{y}(1971) &= \hat{y}(1970)\ Q + k(1971) \\ \hat{y}(1972) &= \hat{y}(1970)\ Q + k(1972) \end{aligned} \qquad (4.1)$$

where y(1969) is the known enrollments from 1969 used to start the forecasting process. $\hat{y}(t)$ is the predicted enrollment for the forecast year. This forecasting process is used in each of the forecasts displayed in this and the next chapter.

When the predicted enrollments are compared with the actual enrollments for the first year forecast, line one and line two of Table 2, the resulting relative error for each grade level, line three, is quite small. The size of the relative errors for the following years of the forecast remain small, although the errors have been compounded through the iterative process of the forecasting procedure.

The quality of the forecast can be evaluated in part by determining if the average forecast error for the grades in each of the forecast years is equal to zero. Three tests were performed in the evaluation process. The three tests were: (1) a t-test on the relative error, (2) a t-test on the prediction error in units of the forecast, and (3) a runs test for the randomness of the twelve enrollment predictions for a given year. Each test is performed on the comparison of the predicted and actual values for a given forecast year. The null hypothesis tested in each case is that the average prediction error is zero.

$$H_0: \text{ Average forecast error } = 0.$$

The alternative hypothesis is that the average prediction error is not zero.

$$H_A: \text{ Average forecast error } \neq 0.$$

Table 3 summarizes the statistical error measures for the Michigan public elementary and secondary enrollments forecast. Selecting a level of significance of .05, the null hypothesis is not rejected for the first and second years of the forecast for each of the three statistical tests. However, the null hypothesis is rejected for the third year of the forecast by each of the tests.

The information presented in Table 2 and in Table 3 appears to indicate that there is a bias in the forecasts. The predicted values are less than the actual values in almost every case. The alternative hypothesis that there is a negative bias to the forecasts can be tested.

$$H_A: \text{ Average forecast error } < 0.$$

The results of the t-tests and the runs test at the .05 level of significance indicate the rejection of the null hypothesis and the

Table 3. Markov student flow model statistical summary of Michigan public school enrollments for the years following 1969.

	RELATIVE ERROR		
YEAR	AVERAGE	STANDARD DEVIATION	t VALUE
1970	-.003	.007	-1.438
1971	-.007	.012	-2.044
1972	-.021	.016	-4.512*

	PREDICTION ERROR†			
YEAR	AVERAGE	STANDARD DEVIATION	t VALUE	RUNS
1970	-514	1048	-1.701	6
1971	-1191	1954	-2.111	4
1972	-3507	2708	-4.487*	1*

* Accept the alternative hypothesis that the average forecast error is not equal to zero and that there is a negative bias in the forecast enrollments at the .05 level of significance, n = 12.

† Headcount.

acceptance of the alternative hypothesis that there is a negative bias in the predicted enrollments in the third year. The forecast under predicts the enrollments by a small factor in the third year of the forecast. The average under prediction is 2.1 percent in the third year of the forecast. The forecast error in the third year is the result of a compounding process due to the interative forecasting procedure.

Estimating the values of Q by the progression ratio technique for the same Michigan data as above results in the values that appear in Table 4. It should be noted that six of the twelve progression ratios are greater than one. These values indicate that more students progress to a given grade than existed in the preceding grade in the previous year. It is physically impossible for a flow rate to be greater than one, which is implied here. However, these flow rate estimators do provide reasonable forecasts. The enrollment forecasts generated with the progression ratio estimated flow rate appear in Table 5.

The progression ratio forecasts in Table 5 were generated using the same iterative procedure as appears in Equation 4.1. The forecast discrepancies for this method appear to be relatively small and about the same size as those from the Markov student flow forecast. Applying the same statistical tests to this forecast as to the previous forecast, the quality of the forecast can be evaluated. A summary of the statistical error measures appears in Table 6.

The null hypothesis is accepted by all three of the tests at the .05 level of significance in the first year forecast. The null hypothesis is rejected by the two t-tests in the second year, and all three tests reject the null hypothesis in the third year. Testing the second alternative hypothesis that the forecasts contain a negative bias results in the acceptance of the alternative hypothesis for the second and third years. This acceptance is for all statistical tests except the runs test in the second year forecast.

TABLE 4. Progression ratio transition matrix generated from aggregate Michigan public school enrollments, 1964 to 1969.

.000	.930											
	.000	.942										
		.000	.983									
			.000	1.002								
				.000	1.000							
					.000	1.001						
						.000	1.030					
							.000	1.001				
								.000	1.067			
									.000	1.008		
										.000	.919	
											.000	.915
												.000

TABLE 5. Progression ratio forecasts of Michigan public school enrollments for the years following 1969.

YEAR		K	1	2	3	4	5	*6
1969	A	183918	173173	166998	162471	161131	163882	160641
1970	P		170976	163062	164131	162849	161070	163998
	A	180938	172142	165576	165876	163005	161168	165365
	R		-.007	-.015	-.011	-.001	-.001	-.008
1971	P		168206	160993	160262	164513	162787	161184
	A	169197	169601	167075	164340	166707	163658	162636
	R		-.008	-.036	-.025	-.013	-.005	-.009
1972	P		157291	158385	158229	160635	164450	162902
	A	163248	162600	167319	165653	166410	169475	166705
	R		-.033	-.053	-.045	-.035	-.030	-.023

YEAR		7	8	9	10	11	12
1969	A	160486	154533	158640	154843	139503	122082
1970	P	165434	160625	164851	159836	142372	127607
	A	166511	161261	164343	160266	142200	126063
	R	-.006	-.004	.003	-.003	.001	.012
1971	P	168892	165578	171350	166094	146963	130231
	A	169535	166567	170830	165144	147924	129142
	R	-.004	-.006	.003	.006	-.006	.008
1972	P	165993	169038	176633	172642	152717	134431
	A	169779	171427	177474	172744	154386	134884
	R	-.022	-.014	-.005	-.001	-.011	-.003

P=Predicted Values
A=Actual Values

R=Relative Error $\frac{P-A}{A}$

Table 6. Progression ratio statistical summary of Michigan public school enrollments for the years following 1969.

YEAR	RELATIVE ERROR AVERAGE	STANDARD DEVIATION	t VALUE
1970	-.003	.007	-1.577
1971	-.008	.013	-2.207*
1972	-.024	.017	-4.688*

YEAR	PREDICTION ERROR† AVERAGE	STANDARD DEVIATION	t VALUE	RUNS
1970	-580	1087	-1.849	4
1971	-1342	2052	-2.266*	4
1972	-3792	2818	-4.662*	1*

* Accept the alternative hypothesis that the average forecast error is not equal to zero and that there is a negative bias in the forecast enrollments at the .05 level of significance, n = 12.

† Headcount

Comparing the results of the two techniques, the Markov student flow model does not generate significantly better results than the progression ratio technique. Testing the null hypothesis that the mean error measures of the two techniques are equal

H_0: Markov prediction error = Progression ratio prediction error,

against the alternative that the mean error measures are not equal

H_A: Markov prediction error ≠ Progression ratio prediction error, results in the acceptance of the null hypothesis. Table 7 displays the t-values of the equal means tests. In all three years, the t-tests of the means for both the relative error and the prediction error in units of the forecast result in the acceptance of the null hypothesis. Theil's inequality coefficient provides additional support to the conclusion that there is no significant difference between the two forecasting methods for elementary and secondary enrollments.

Table 7. Comparison of the Markov student flow and the grade progression ratio methods for Michigan elementary and secondary enrollment forecasts.

	THEIL'S INEQUALITY COEFFICIENT		EQUAL MEAN PREDICTION ERROR t-VALUES	
Year	Markov	Progression Ratio	Relative Error	Prediction Error
1970	.007	.007	0	.199
1971	.014	.015	.265	.250
1972	.026	.028	.604	.342

New York Elementary and Secondary Enrollments

The result of the first evaluation of the Markov student flow process indicates no significant improvement over the progression ratio technique. A second set of tests and comparison of the two techniques were performed using public school enrollment data from the State of New York. The New York public school enrollment data from the years 1963 to 1968 was used to estimate the values of Q. The results of the quadratic programming estimation process appear in Table 8. Using these values of Q and the enrollment numbers for 1968, the forecasts are iteratively generated for the years 1969, 1970, 1971, and 1972. These forecasts appear in Table 9.

The quality of the New York forecasts can be evaluated with the same statistical tests used to evaluate the Michigan public school enrollments. Table 10 summarizes the statistical error measures for the New York public elementary and secondary enrollment forecasts generated with the Markov student flow model. The result is the null hypothesis of the average error for each annual forecast is zero is not rejected at the .05 level of significance against the first alternative hypothesis. The prediction of the New York enrollments resulted in better forecasts than did the prediction of the Michigan enrollments.

Estimating the values of Q for the New York data by the progression ratio technique results in the values displayed in Table 11. Just as in the Michigan example, several of the student flow rates are larger than one. The enrollment forecasts generated with the use of this Q are displayed in Table 12. These forecasts were generated with the iterative process described previously.

The quality of the progression ratio forecasts can be evaluated in the same manner as the previous forecasts. Inspecting the relative errors for each forecast year, the relative errors appear to be approximately the same size as those resulting from the Markov

TABLE 8. Markov student flow model transition matrix generated from aggregate New York public school enrollments, 1963 to 1968.

.000	1.000											
	.000	.979										
		.000	.979									
			.010	.990								
				.014	.986							
					.013	.987						
						.007	.993					
							.041	.959				
								.000	1.000			
									.124	.876		
										.123	.877	
											.033	.897
												.000

TABLE 9. Markov student flow model forecasts of New York public school enrollments for the years following 1968. (thousands)

YEAR		K	1	2	3	4	5	6
1968	A	285	282	269	263	262	257	247
1969	P		285	276	266	264	262	255
	A	287	281	276	267	264	261	256
	R		.014	.000	-.004	.000	.003	-.002
1970	P		287	279	273	267	264	260
	A	283	284	273	273	267	264	261
	R		.011	.022	-.000	.000	-.001	-.004
1971	P		283	281	276	274	267	262
	A	272	283	277	271	272	268	263
	R		.000	.014	.018	.007	-.005	-.003
1972	P		272	277	278	277	274	265
	A	257	272	275	273	271	271	266
	R		.000	.007	.018	.022	.010	-.003

YEAR		7	8	9	10	11	12
1968	A	249	234	257	252	225	192
1969	P	255	239	266	256	228	202
	A	255	242	270	258	229	200
	R	.002	-.013	-.015	-.007	-.002	.009
1970	P	264	245	272	264	232	205
	A	264	248	275	265	231	203
	R	.000	-.012	-.012	-.002	.005	.009
1971	P	269	253	279	271	240	208
	A	269	257	286	275	240	205
	R	.000	-.015	-.026	-.016	-.002	.016
1972	P	271	258	288	277	245	215
	A	269	262	296	282	248	213
	R	.009	-.015	-.028	-.016	-.011	.009

P=Predicted Values
A=Actual Values

R=Relative Error $\frac{P-A}{A}$

student flow forecasts. However, when the statistical qualities of the four annual forecasts are examined, the evaluation changes. Table 13 summarizes the statistical error measures. The null hypothesis is accepted that the average forecast error is zero for the first two years, but there is a negative bias in the last two forecast years.

Table 10. Markov student flow model statistical summary of New York public school enrollments for the years following 1968.

YEAR	RELATIVE ERROR AVERAGE	STANDARD DEVIATION	t VALUE
1969	-.001	.008	-.564
1970	.001	.009	.500
1971	-.001	.013	-.227
1972	.000	.015	.007

YEAR	PREDICTION ERROR† AVERAGE	STANDARD DEVIATION	t VALUE	RUNS
1969	-.360	2.155	-.579	7
1970	.340	2.514	.469	7
1971	-.330	3.595	-.318	5
1972	-.068	4.168	-.057	5

Note: The null hypothesis that the average forecast error equals zero is accepted at the .05 level of significance, n = 12, for all four years.

† Thousands

TABLE 11. Progression ratio transition matrix generated from aggregate New York public school enrollments, 1963 to 1968.

.000	.967											
	.000	.979										
		.000	.989									
			.000	1.003								
				.000	.998							
					.000	.994						
						.000	1.034					
							.000	.959				
								.000	1.138			
									.000	.997		
										.000	.906	
											.000	.897
												.000

TABLE 12. Progression ratio forecasts of New York public school enrollments for the years following 1968.
(thousands)

YEAR		K	1	2	3	4	5	6
1968	A	285	282	269	263	262	257	247
1969	P		276	276	266	264	262	255
	A	287	281	276	267	264	261	256
	R		-.019	.000	-.004	-.001	.002	-.002
1970	P		278	270	273	267	263	260
	A	283	284	273	273	267	264	261
	R		-.022	-.011	.000	-.001	-.002	-.004
1971	P		274	272	267	274	266	262
	A	272	283	277	271	272	268	263
	R		-.033	-.019	-.015	.007	-.006	-.005
1972	P		263	268	269	268	273	265
	A	257	272	275	273	271	271	266
	R		-.033	-.025	-.015	-.012	.009	-.005

YEAR		7	8	9	10	11	12
1968	A	249	234	257	252	225	192
1969	P	255	239	266	256	228	202
	A	255	242	270	258	229	200
	R	.002	-.013	-.014	-.007	-.003	.009
1970	P	264	245	272	265	232	205
	A	264	248	275	265	231	203
	R	.001	-.012	-.012	.001	.005	.009
1971	P	269	253	279	271	240	208
	A	269	257	286	275	240	205
	R	-.001	-.014	-.025	-.015	.002	.016
1972	P	271	258	288	278	245	216
	A	269	262	296	282	248	213
	R	.006	-.016	-.026	-.015	-.010	.013

P=Predicted Values
A=Actual Values

R=Relative Error $\frac{P-A}{A}$.

Table 13. Progression ratio statistical summary of New York public school enrollments for the years following 1968.

	RELATIVE ERROR		
YEAR	AVERAGE	STANDARD DEVIATION	t VALUE
1969	-.004	.008	-1.769
1970	-.004	.009	-1.569
1971	-.009	.014	-2.229*
1972	-.011	.014	-2.584*

	PREDICTION ERROR†			
YEAR	AVERAGE	STANDARD DEVIATION	t VALUE	RUNS
1969	-1.125	2.059	-1.893	8
1970	-1.166	2.351	-1.718	6
1971	-2.582	3.721	-2.404*	4
1972	-3.029	3.887	-2.699*	6

*: Accept the alternative hypothesis that the average forecast error is not equal to zero and that there is a negative bias in the forecast enrollments at the .05 level of significance, n = 12.

† Thousands

Comparing the results of the two forecasting techniques the Markov student flow model does not generate significantly better results than the progression ratio technique for the first three years of the New York forecast. However, the forecasts for the fourth year are different at the .05 level of confidence. Table 14 displays the statistical comparisons for the New York forecasts. Theil's inequality coefficients are smaller in each of the forecast years, although the difference between the pairs of coefficients for each forecast year is small.

Table 14. Comparison of the Markov student flow and the grade progression ratio methods for New York elementary and secondary enrollment forecasts.

Year	THEIL'S INEQUALITY COEFFICIENT		EQUAL MEAN PREDICTION ERROR t-VALUES	
	Markov	Progression Ratio	Relative error	Prediction error
1969	.008	.009	1.246	1.204
1970	.009	.010	1.851	2.052
1971	.013	.017	1.969	2.042
1972	.015	.018	2.517*	2.437*

*: Accept the alternative hypothesis that the mean prediction errors are not equal at the .05 level of significance, n = 24.

The conclusion from the preceding comparison and validation process is that the Markov student flow model generates accurate enrollment projections for statewide elementary and secondary school systems. The comparison of the Markov student flow model with the progression ratio technique did not indicate a clear superiority of the Markov model. Theil's inequality coefficient was smaller for the Markov student flow model in six of the seven forecast years. The student flow forecasts contained a negative bias for only one year, compared to four years for the progression ratio technique.

In addition, the prediction error measures were smaller in six of the seven forecast years and there was a significant difference between the fourth year forecasts for New York. The Markov student flow model provides marginally better results than the progression ratio technique.

First-Time Enrollments

The final application of this chapter is the generation of the first-time enrollment predictions for Phase II of the higher education enrollment projection model. This forecast will use Michigan public and non-public, elementary and secondary school enrollments,[4] with the first-time entrants to the system of higher education appended to the flow process.[5] The modified flow process includes the transition from the twelfth grade to the state of enrolling in a university or community college. The total statewide elementary and secondary enrollments are used in forecasting the first-time entrants because the non-public students constitute an important factor in school enrollments. This portion of the school enrollment should be included in the prediction process when the first-time entrants are forecast. Including the non-public enrollments identifies the total potential higher education entrants.

The forecast of first-time entrants and total elementary and secondary enrollments is displayed in Table 15. The forecast was generated from the base year, 1970, with the use of a transition matrix Q estimated with enrollment data from the years 1967 to 1970. Although the forecast can be compared with only two years of actual

TABLE 15. Markov student flow model forecasts of Michigan public and private school enrollments and first-time entrants for the years following 1970.

YEAR		K	1	2	3	4	5	6
1970	A	185645	195974	192220	194606	191901	190358	194789
1971	P		192896	189492	187799	193019	189790	189499
	A	173708	191512	190936	189501	193410	190330	189747
	R		.007	-.008	-.009	-.002	-.003	-.001
1972	P		180845	186517	185134	186354	190896	188633
	A	167462	179595	186608	185740	187682	191522	188898
	R		.007	-.000	-.003	-.007	-.003	-.001
1973	P		174153	174900	182227	183666	184304	189606
1974	P			168409	170877	180785	181646	183538
1975	P				164536	169619	178796	180677
1976	P					163275	167753	177844
1977	P						161479	167393
1978	P							160889
1979	P							
1980	P							

YEAR		7	8	9	10	11	12	1st TIME ENTRANTS
1970	A	194155	188125	182744	177400	158602	142758	79689
1971	P	196690	191828	186867	181951	163208	144962	83085
	A	195626	192116	188544	181723	163688	144786	85697
	R	.005	-.002	-.009	.001	-.003	.001	-.030
1972	P	191927	194412	190576	186100	167395	149172	84368
	A	190806	192702	192003	186374	167469	148104	83544
	R	.006	.009	-.007	-.001	-.000	.007	.010
1973	P	190763	190166	193213	189836	171212	152999	86818
1974	P	191588	188817	189395	192557	174650	156488	89045
1975	P	185969	189494	187907	189269	177152	159630	91076
1976	P	182875	184335	188453	187640	174128	161917	92904
1977	P	179995	181140	183656	188011	172629	159153	94236
1978	P	169999	178268	180383	183637	172970	157783	92627
1979	P	163172	168828	177501	180282	168946	158095	91830
1980	P		161902	168496	177369	165860	154417	92011

P=Predicted Values
A=Actual Values

R=Relative Error $\frac{P-A}{A}$

enrollments, the quality of this forecast can be evaluated in the same manner as the previous forecasts. However, the evaluation process is expanded to include the first-time entrants. The statistical evaluation of the predicted enrollments compared with the actual enrollments is summarized in Table 16.

Table 16. Markov student flow model statistical summary of Michigan public and private school enrollments and first-time entrants for the years following 1970.

YEAR	RELATIVE ERROR AVERAGE	STANDARD DEVIATION	t VALUE
1971	-.004	.009	-1.522
1972	.001	.006	.647

YEAR	PREDICTION ERROR† AVERAGE	STANDARD DEVIATION	t VALUE	RUNS	THEIL'S INEQUALITY COEFFICIENT
1971	-502	1129	-1.604	8	.007
1972	98	1005	.351	5	.007

† Headcount

Note: The null hypothesis that the average forecast error equals zero is accepted at the .05 level of significance, n = 13, for each of the forecast years.

We now have a method of forecasting the first-time enrollments for a statewide system of universities and community colleges. As the elementary and secondary students flow through the lower school system, they become potential entrants to the system of higher education. Therefore, we can apply the same flow technique to the estimation of first-time enrollments that we used to estimate the enrollments in the various grades in the lower school system. Now that we have the forecast first-time entrants into the system of higher education, we can use these estimates in the process of forecasting the enrollments for the system of higher education.

Chapter IV Notes

[1]R. M. Oliver, _Models for Predicting Gross Enrollments at the University of California_, Report No. 68-3, Ford Foundation Research Program in University Administration (Berkeley, California: 1968).

[2]Michigan State University, Graduate School of Business Administration, _Michigan Statistical Abstract_ (10th ed.; East Lansing, Michigan: 1974).

[3]New York State, Division of the Budget, Office of Statistical Coordination, _New York State Statistical Yearbook_ (Albany, New York: 1973), p. 201

[4]Michigan State University, _op. cit._, p. 157.

[5]The first-time enrollment data is part of HEGIS data mentioned in Chapter I.

CHAPTER V

FORECASTING STATEWIDE ENROLLMENTS

Forecasting statewide enrollments for higher education can now be attempted using the results of the preceding chapter. Using the forecast of first-time enrollments generated by the flow of students through elementary and secondary school system, future statewide enrollments at the undergraduate level can be projected. Before the results of the flow process are examined, let us look at the results of two other methods. Both methods are applied to the forecasting of the total enrollments of public community colleges and baccalaureate institutions in the State of Michigan. The first is the population ratio method. The second is a multiple, linear regression model used by the Michigan Department of Education for forecasting enrollments.

A time period of five years was selected for the estimation of P in the Markov model. This length of time was selected because there are only eight annual observations of enrollments available for estimation and validation. Using five years for the estimation of P leaves three years of actual enrollments for validation. In addition, the five year period only provides four observations for the quadratic programming estimation procedure. A shorter time period would not provide enough observations for the estimation process. A longer time period would not allow enough time periods for the comparison with the other forecasting methods and for validation of the Markov student flow model. The estimation of the population ratios also use a five year time period to be consistent with the Markov estimation process and to provide comparable forecasts.

Population Ratio Method

The population ratio method is the simplest procedure for forecasting statewide enrollments in the system of higher education. However, it requires data on the 18 to 21 year old age group during the past and a forecast of the size of that age group for the future. This data is not available for the 18 to 21 year old age group or by single years, except for census years. As a result, some other method must be used to predict the size of 18 to 21 year old population.

There are two possible methods that could be employed to predict the 18 to 21 year old, college age population. One would be to use primary or secondary school enrollment by grade and year as the prediction basis for the college age group several years in the future. This method has a major disadvantage; only the public school enrollments are available for the required years. The private school enrollments have only been gathered for a few years. This would cause no problem if the ratio of the number of students in the private schools to the number of students in the public schools was stable. This ratio has been becoming smaller since 1963 when it was 0.193.[1] In 1966, the first year of usable enrollment data for the statewide system of higher education in Michigan, the ratio was 0.174. The ratio drops to 0.100 in 1972. This observed drop is the result of the closing of many private schools and those students entering the public school system. The result of the student transfers has caused the public school enrollments to continue growing after the total statewide enrollments have begun to decline. Therefore, the use of only public school enrollments as a predictor of the college age group would introduce an error attributable to the transfer of students from the private schools to the public schools. The total public and private school enrollments cannot be used because the private school enrollment data is not available by grade level for all of the years needed to compare with the higher education enrollments.

The second method uses the number of live births as the predictor of the future college age population. Since the population ratio method only calculates a ratio from the relevant population, the live births for the time period 18 to 21 years prior to the year in question could be used in the calculation process. For example, the number of live births in the years 1945 to 1948 would be used in place of the 18 to 21 year old population for the computation of the college enrollment ratio for 1966.

Population Ratio--Average

There are two techniques that can be used to estimate the ratios for the population ratio method of forecasting higher education enrollments. One technique calculates the average of the ratios of the higher education enrollment to the college age population for the years of the selected historical period. The second technique

calculates a simple linear regression of the ratio from the ratios of the selected historical period. Both techniques are used with the live birth method to produce forecasts of the higher education enrollments.

The live birth data used in computing the population ratios is displayed in Table 17. The corresponding higher education enrollments are displayed in Table 18. Using the data from these tables, the average ratio is computed for the years 1966 to 1970. The population ratio averages and the resulting enrollment forecasts for the community colleges and university undergraduates are displayed in Table 19.

Population Ratio--Linear Regression

The same data is used to compute a linear regression equation for the population ratio. The resulting linear equations for the community colleges and the baccalaureate institutions and the enrollment forecasts are displayed in Table 20. The regression parameters are all significantly different from zero. Both parameters of the community college equation and the intercept of the university equation are different at the .005 level of significance. The slope parameter of the university regression equation is different from zero at the .025 level of significance. The coefficients of determination for both equations indicate a good fit for the equations over the 1966 to 1970 historical period.

The results of these forecasts are graphically presented in Figures 4 and 5 along with the results of the multiple linear regression

Table 17. Number of live births.

Michigan 1940-1969			
Year	Number	Year	Number
1940	99,106	1955	196,294
1941	107,498	1956	206,068
1942	124,068	1957	208,488
1943	125,441	1958	202,690
1944	113,586	1959	198,301
1945	111,557	1960	195,056
1946	138,572	1961	192,825
1947	160,275	1962	182,790
1948	153,726	1963	178,871
1949	156,469	1964	175,103
1950	160,055	1965	166,464
1951	172,451	1966	165,794
1952	177,835	1967	162,756
1953	182,969	1968	159,058
1954	192,104	1969	165,760

Source: Michigan Department of Public Health, Michigan Center for Health Statistics, Michigan Health Statistics--Annual Statistical Report, (Lansing, Michigan: 1969), p. 2.

Table 18. Public community college and baccalaureate undergraduate enrollments for Michigan, 1966-1973.

Year	Community Colleges	Baccalaureate Undergraduates
1966	69,504	128,208
1967	79,256	135,573
1968	95,065	149,347
1969	115,791	158,130
1970	126,621	165,051
1971	132,189	169,178
1972	136,121	164,178
1973	151,652	163,327

Source: Higher Education General Information Survey, Michigan Association of Collegiate Registrars and Admissions Officers, and the Michigan Bureau of Programs and Budget as described in Chapter I.

Table 19. Enrollment forecasts based on the average ratio of enrollments to live births 18 to 21 years earlier for the years 1966 to 1970.

Year	Community Colleges	Baccalaureate Undergraduates
1971	107,254	163,621
1972	112,213	171,184
1973	115,901	176,811
1974	120,269	183,474
1975	124,217	189,497
1976	125,855	191,995
1977	126,165	192,469
1978	124,462	189,870
1979	122,038	186,174
1980	118,960	181,477

	RATIO	
	Average	Standard Deviation
Community College	.1547	.0267
Baccalaureate Undergraduates	.2360	.0110

Table 20. Enrollment forecasts based on the linear regression of the ratio of enrollments to live births 18 to 21 years earlier for the years 1966 to 1970.

Year	Community Colleges	Baccalaureate Undergraduates
1971	145,318	176,932
1972	165,309	189,754
1973	184,453	200,786
1974	205,631	213,328
1975	227,075	225,469
1976	244,957	233,649
1977	260,486	239,445
1978	271,691	241,361
1979	280,838	241,710
1980	287,826	240,534

Community College Ratio

$$\widehat{CCR} = \underset{(.0070)}{\overset{**}{.0998}} + \underset{(.0021)}{\overset{**}{.0183}}(t-1965) \qquad R^2 = .962$$

Baccalaureate Ratio

$$\widehat{BR} = \underset{(.0054)}{\overset{**}{.2165}} + \underset{(.0016)}{\overset{*}{.0064}}(t-1965) \qquad R^2 = .835$$

*: The parameter below this is significantly different from zero at the .025 level of significance, n = 5.

**: The parameter below this is significantly different from zero at the .005 level of significance, n = 5.

(): The number within the parenthesis is the standard deviation of the estimated parameter which appears above the parenthesis.

model and the Markov student flow model. Figure 4 presents a comparison of the forecasts for the community colleges. Figure 5 presents a comparison of the forecasts of the undergraduate enrollments for the colleges and universities. The wide divergence of the two population ratio techniques is evident for both sectors of the system of higher education. These results are evaluated and compared with the results of a linear regression forecast and the Markov student flow forecast in a later section of this chapter.

Linear Regression

Statewide enrollments have also been forecast through the application of multiple linear regression. One example is a study conducted by the Michigan Department of Education to determine the effects of different variables upon higher education enrollments. The study consisted of a multiple linear regression analysis of the variables affecting Michigan public higher education enrollments over the period from 1951 to 1969.[2] The analysis identified four variables that strongly affected enrollments. These variables were:

(1) the number of young people in the age group most likely to attend college in future years,

(2) the number of people discharged from the armed services,

(3) income levels, and

(4) the employment rate.[3]

The coefficients of these four variables were determined and then used in linear regression enrollment forecasting model. The enrollment forecasts for 1971 to 1980 are presented in Table 21 and

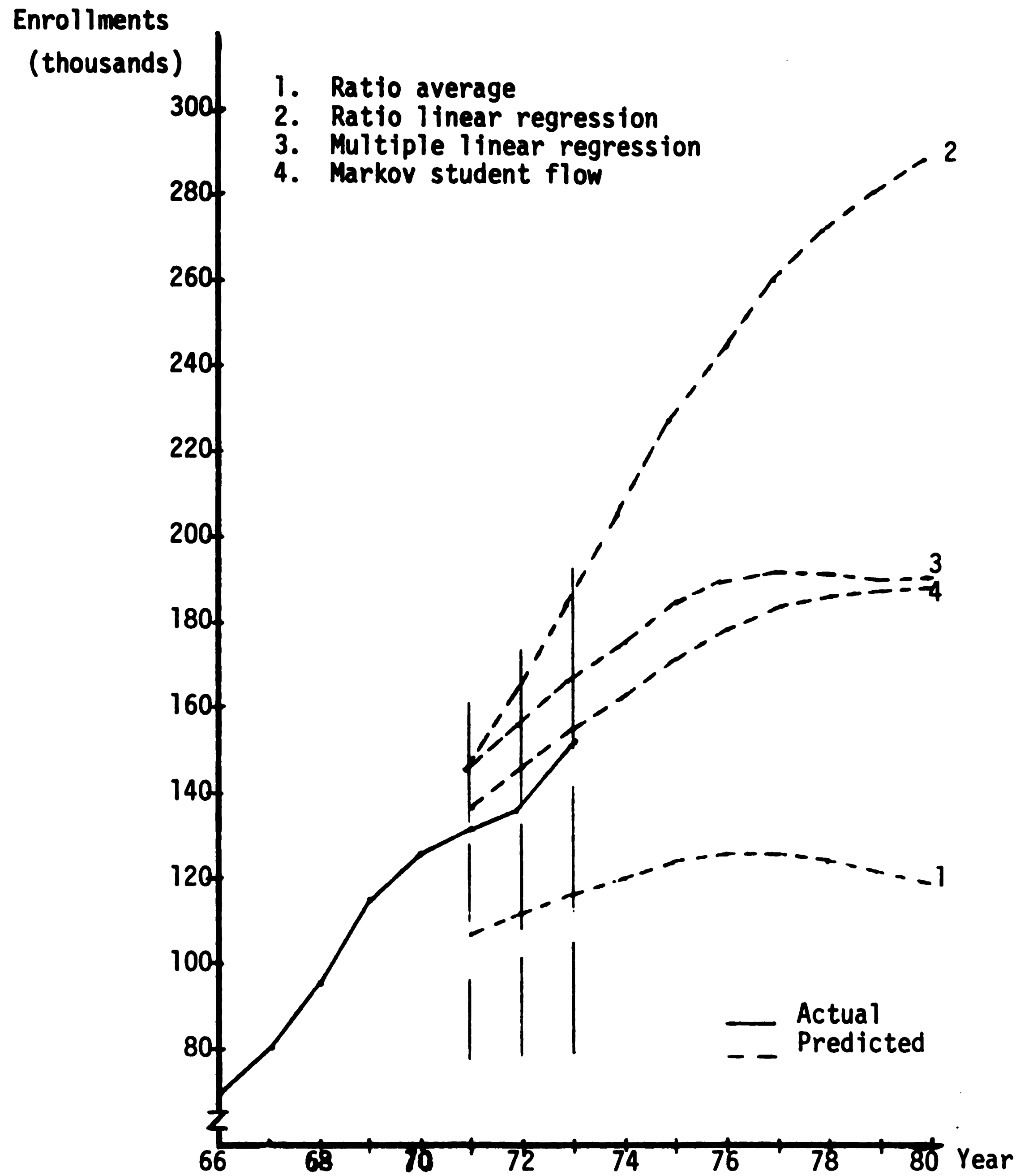

Figure 4. Forecasts of public community college enrollments for the State of Michigan.

Note: See Tables 16, 17, 18, 19, and 20 for numerical values.

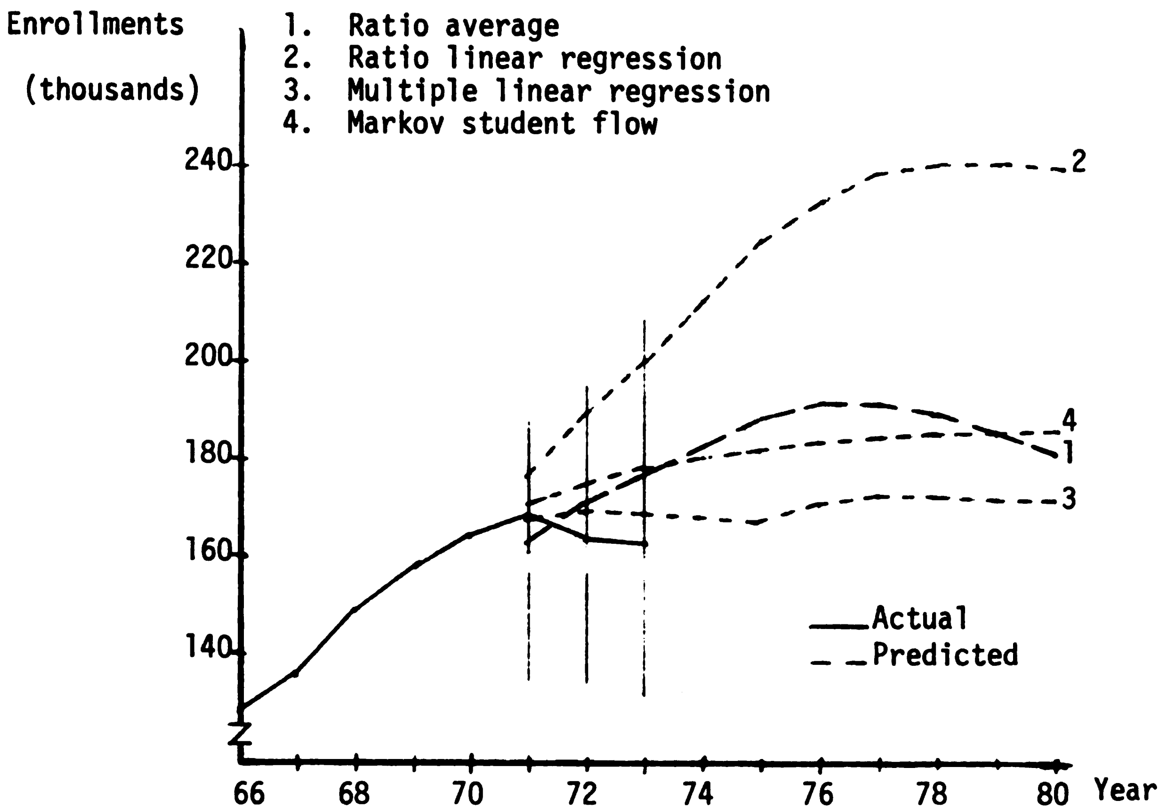

Figure 5. Forecasts of public college and university enrollments for the State of Michigan.

Note: See Tables 16, 17, 18, 19, and 20 for numerical values.

Table 21. Enrollment forecasts from the Michigan Department of Education linear regression model.

Year	Community Colleges	Baccalaureate Undergraduates
1971	145,800	167,700
1972	156,000	168,900
1973	166,000	169,300
1974	175,300	168,400
1975	184,300	166,800
1976	189,300	171,200
1977	191,400	173,100
1978	190,800	172,700
1979	190,300	172,200
1980	189,800	171,700

Source: Michigan Department of Education. Financial Requirements of Public Baccalaureate Institutions and Public Community Colleges. Lansing, Michigan. 1971. p. 10.

graphically displayed in Figures 4 and 5. No indication was given as to the validity of the model or the goodness of fit of the linear equations used to forecast the enrollments.

Statewide Student Flow Model

Forecasting statewide enrollments using the student flow model described in Chapter III, the results displayed in Table 22 are obtained. These results were obtained by using the student flow model in a highly aggregated, three state form. The three states consist of one state for the community colleges, one state for the lower division of the colleges and universities, and one state for the upper division of the colleges and universities. The community college forecast is graphically displayed in Figure 4. The total enrollment forecast for the colleges and universities is graphically displayed in Figure 5.

The statewide student flow model forecasts were generated with the use of Equation 3.1. The transition matrix, P, was estimated with enrollment data from the years 1966 to 1970. Starting with the 1970 Fall enrollments, the Fall enrollments for the years 1971 to 1980 were iteratively generated. The annual enrollment forecasts are produced by advancing the model in steps of one year, using the previous year's forecast as the starting point for the next year's forecast and using the first-time entrant forecasts, f(t), from Chapter IV. This procedure is indicated in the following set of equations.

Table 22. Markov student flow model enrollment forecasts.

Year	Community Colleges	Baccalaureate Undergraduates Fresh Soph	Junior Senior	Total
1971	137,453	87,826	82,961	170,787
1972	146,211	88,418	86,699	175,118
1973	154,704	88,792	89,570	178,362
1974	162,790	89,028	91,749	180,777
1975	170,401	89,177	93,388	182,565
1976	177,481	89,270	94,612	183,882
1977	183,698	89,330	95,521	184,851
1978	186,375	89,367	96,193	185,560
1979	187,428	89,391	96,688	186,079
1980	188,337	89,405	97,051	186,456

$$\hat{x}(1971) = \hat{x}(1970)\,P + \hat{f}(1971)$$
$$\hat{x}(1972) = \hat{x}(1971)\,P + \hat{f}(1972)$$
$$\vdots$$
$$\hat{x}(1980) = \hat{x}(1979)\,P + \hat{f}(1980),$$

where $\hat{x}(t)$ is the forecasted enrollments for the t^{th} year. This procedure is used in all of the student flow model forecasts that follow.

Three Model Comparison

Each of the three models has its advantages and disadvantages. However, the primary requisite for any of the forecasting models is that it predict the future with accuracy. A secondary consideration is the ease and cost of gathering the needed data and estimating the parameter values of the model. In regard to the ease and cost of data gathering and estimating the parameter values, the models can be ranked in the following sequence: (1) population ratio, (2) student flow, and (3) linear regression. However, the important ranking is the one of accuracy. The models can be ranked on the basis of accuracy, in the following sequence: (1) student flow, (2) linear regression, and (3) population ratio. This ranking is based on the comparison of the accurracy of the projection of the enrollments for 1971, 1972, and 1973. Table 23 displays the forecasts for each of the models and the amount of discrepancy between each forecast and the actual enrollment in each year. In addition Theil's inequality coefficient has been computed for each forecast method and year. The coefficient for each year was computed from only two pairs of observed and predicted values, the

TABLE 23. Statistical comparison of the three model forecasts.

	1971		1972		1973	
	COMM COLL	BACH	COMM COLL	BACH	COMM COLL	BACH
Actual	132189	169178	136121	164559	151652	163327
Markov Student Flow						
P	137453	170787	146211	175118	154704	178362
R	.040	.010	.074	.064	.020	.092
U	.037		.075		.089	
Multiple Linear Regression						
P	145800	167700	156000	168900	166000	169300
R	.103	-.009	.146	.026	.095	.037
U	.057		.095		.070	
Population Ratio: Average						
P	107254	163621	112213	171184	115901	176811
R	-.189	-.033	-.176	..040	-.236	.083
U	.108		.116		.171	
Population Ratio: Linear Regression						
P	145318	176932	165309	189754	184453	200786
R	.099	.046	.214	.153	.216	.229
U	.064		.180		.223	

P=Predicted Value
U=Theil's Inequality Coefficient
R=Relative Error $\frac{P-A}{A}$
A=Actual Value

total community college enrollment and the total undergraduate enrollment at the baccalaureate institutions. This will facilitate the comparison of the results of each method on a year-to-year basis.

Inspecting the results of the population ratio method, we find that both the simple average and the linear regression techniques result in relatively large forecast errors. Comparing Theil's inequality coefficient for each year of either population ratio technique with those of either the student flow or linear regression method, in every instance, the inequality coefficient of the population ratio technique is larger. In addition, the size of the coefficient is increasing much more rapidly for the population ratio techniques than the other methods. This implies that of the three methods, the population ratio method is the least accurate over the time span under consideration. The conclusion is that the population ratio method is not as accurate as other models and should not be used when other methods are available. In addition, it appears that the amount of error will increase in the more distant forecasts. This is particularly evident with the community college forecasts displayed in Figure 4.

Comparing the results of the student flow model and the multiple linear regression model does not lead to a clear decision as in the preceding comparison. The amount of error measured and Theil's inequality coefficient for the different observations do not strongly indicate either method as predominately better than the other. In each of the three annual observations, the student flow method forecasts the community college enrollments more accurately than did the linear regression method. However, the linear regression method forecasts the undergraduate enrollments at the baccalaureate institutions more accurately. Theil's inequality coefficient measurements for each of the three years do not indicate either method as being consistently more accurate.

The student flow method produces smaller Theil's inequality coefficients for the years 1971 and 1972, but the linear regression method produces the smaller coefficient for 1973. In addition, the amount of difference between the pair of coefficients for a given year is not significant relative to the magnitude of the coefficients. Neither of the two methods clearly produces more accurate forecasts than the other. Considering the results of each method over the three year period, the aggregate flow method provides marginally better results. The average error is smaller, .050 vs. .069, and the average Theil's inequality coefficient is smaller, .067 vs. .074, for the student flow method. The student flow method and the linear regression method are both clearly better forecasting methods than the population ratio method. However, neither of the two is clearly superior to the other, although the student flow method did perform better than the linear regression method over the 1971 to 1973 time period.

The total forecast undergraduate enrollments for both the linear regression and the student flow methods do not differ greatly over the 1971 to 1980 forecast period. This is clearly evident in the graphical presentation of the total undergraduate enrollments in Figure 6. The first seven years of both forecasts are approximately the same with very little difference in either forecast. The forecast totals begin to diverge after 1976. However, even at the point

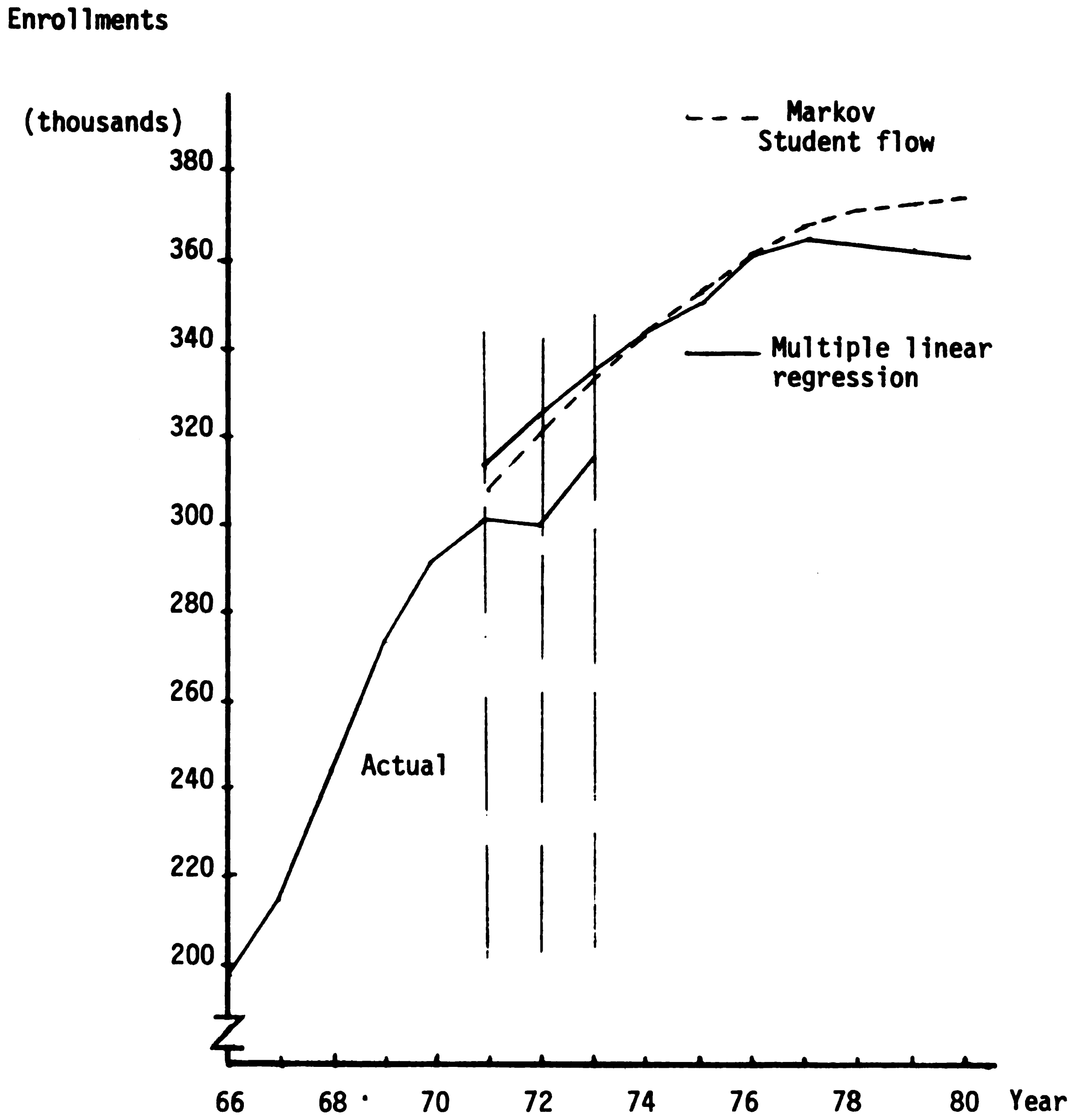

Figure 6. Multiple linear regression and Markov student flow model forecasts of total public undergraduate enrollments for the State of Michigan.

of greatest divergence in 1980, there is only a 3.6 percent difference between the two forecast totals. Considering the community college forecasts, Figure 4, and the baccalaureate undergraduate forecasts, Figure 5, the general shape of the enrollment forecast curves are the same. The linear regression method forecasts a larger number of community college enrollments for each year in the 10 year forecast period. The student flow method forecasts the larger number of enrollments at the baccalaureate institutions. The shape of the 10 year enrollment forecast curves and the relationship of the individual curves reinforce the conclusion that neither method is clearly superior to the other.

Statewide Student Flow Forecasts

Utilizing the three state student flow model described earlier in this chapter, the statewide forecast for the years 1971 to 1980 was generated. The three state model aggregates the community college enrollments into one state. The university enrollments are aggregated into two states, one for the upper division and one for the lower division students.

The data used in testing the higher education Markov student flow, enrollment forecasting model is from the 29 public community colleges and 13 public colleges and universities in Michigan. Table 24 contains a list of both types of institutions. Fall enrollment data from these institutions for the years 1966 to 1970 was used to estimate the Markov transition matrix, P.

Using Equation 3.1 and starting with the enrollments of a given year, the enrollments for each succeeding year was forecast.

Table 24. List of public higher education institutions in the State of Michigan

BACCALAUREATE INSTITUTIONS
Central Michigan University
Eastern Michigan University
Ferris State College
Grand Valley State College
Lake Superior State College
Michigan State University
Michigan Technological University
Northern Michigan University
Oakland University
Saginaw Valley College
University of Michigan
Wayne State University
Western Michigan University
COMMUNITY COLLEGES
Alpena
Bay de Noc
Delta
C.S. Mott
Glen Oaks
Gogebic
Grand Rapids
Henry Ford
Highland Park
Jackson
Kalamazoo
Kellogg
Kirtland
Lake Michigan
Lansing
Macomb County
Mid Michigan
Monroe County
Montcalm
Muskegon
North Central
Northwestern
Oakland
St. Clair County
Schoolcraft
Southwestern
Washtenaw
Wayne County
West Shore

The procedure followed was to initialize the model with the starting values, then iterate the model in steps of one year and use the first-time entrant values forecast in Chapter IV. The resulting forecast with a starting year of 1970 is displayed in Table 25. Although a portion of the forecasts has high error rates, the amount of error for each annual total forecast is not exceedingly large. This is indicated by the relatively small size of U, Theil's inequality coefficient, for each year where the actual and forecast values could be compared. In addition, the performance of a t-test on the prediction discrepancy and the enrollment error indicates that the difference between the predicted enrollment and the actual enrollment is zero at the five percent level of confidence for 1971 and 1973. The t-test rejects this null hypothesis at the five percent level of confidence for the predictions for 1972. Table 26 contains a statistical summary of this enrollment forecast for the Michigan system of higher education.

Returning to Figure 6, we see that there was no growth in enrollments between 1971 and 1972. Whatever the cause of this discontinuity, it is the reason for the rejection of the null hypothesis for 1972. Also looking at Figure 5, we see the enrollments for the baccalaureate institutions dropping during the years of 1971 to 1973. In this case, the forecasts for the higher education enrollments are being made at a time of change in the particular system being forecast. Using more recent data to estimate the transition probabilities and the starting point of the forecasts should reduce the error due to the change in the system.

Table 25. Three state Markov enrollment forecast, 1971 to 1973.

Year		Community Colleges	BACCALAUREATE Fresh Soph	Junior Senior	Total
1970	A	126,621	86,888	78,163	165,051
1971	P	137,453	87,826	82,961	170,787
	A	132,189	87,352	81,826	169,178
	R	.040	.005	.014	.010
1972	P	146,211	88,418	86,699	175,118
	A	136,121	85,112	79,447	164,559
	R	.074	.039	.091	.064
1973	P	154,704	88,792	89,570	178,362
	A	151,652	85,772	77,555	163,327
	R	.020	.035	.155	.092

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Table 26. Statistical summary for the three state enrollment forecast, 1971 to 1973.

YEAR	RELATIVE ERROR AVG	STD DEV	t VALUE	PREDICTION ERROR AVG	STD DEV	t VALUE	U
1971	.020	.018	1.904	2291	2596	1.529	.030
1972	.068	.027	4.411*	6883	3407	3.499*	.072
1973	.070	.074	1.644	6029	5184	2.014	.067

*: Accept the alternative hypothesis that the average forecast error is different from zero at the .05 level of significance, n=3.

Estimating P From More Recent Data

Applying the moving average concept, new transition probabilities can be estimated for P. The previous forecasts used a P estimated from enrollment data from the years 1966 to 1970. Now the transition probabilities will be estimated from enrollment data from the years 1967 to 1971. The five year historical data base for estimation of P has been advanced one year. The resulting enrollment forecasts starting from the base year of 1971 are displayed in Table 27. This forecast predicts the 1972 and 1973 enrollments more accurately than the previous forecast. The updated starting point and transition probabilities cause a more accurate forecast to be generated.

Statistically the forecasts for 1972 and 1973 have improved over the previous forecasts for these years. Table 28 contains the statistical summary for the forecasts in Table 27. The size of the error values have decreased. However, the forecast error for 1972 is still different from zero at the .05 level of significance. The portion of the forecast for the baccalaureate institutions is more accurate than the earlier forecast. However, the forecast for the community colleges results in a change of sign for the relative error. It appears from the data that there has been a disturbance in the general increasing enrollment pattern for the community colleges in 1972. If it is a one time occurrence, its impact will lessen over the years. However, if it is an indication of a permanent change in the enrollment pattern, then it will require a several year delay before

Table 27. Three state Markov enrollment forecast, 1972 and 1973.

Year		Community College	BACCALAUREATE Fresh Soph	Junior Senior	Total
1971	A	132,189	87,352	81,826	169,178
1972	P	139,538	88,556	85,229	173,784
	A	136,121	85,112	79,447	164,559
	R	.025	.040	.073	.056
1973	P	146,890	89,322	87,889	177,210
	A	151,652	85,772	77,555	163,327
	R	-.031	.041	.133	.085

P=Predicted Value R=Relative Error $\frac{P-A}{A}$
A=Actual Value

Table 28. Statistical summary for the three state enrollment forecast, 1972 and 1973.

YEAR	PREDICTION DISCREPANCY AVG	STD DEV	t VALUE	PREDICTION ERROR AVG	STD DEV	t VALUE	U
1972	.046	.024	3.283*	4214	1357	5.377*	.042
1973	.048	.083	1.002	3040	7561	.696	.062

*: Accept the alternative hypothesis that the average forecast error is different from zero at the .05 level of significance, n=3.

enough historical data can be collected upon which future forecasts can be made. This also applies to the college and university enrollments which have decreased from 1971 to 1973.

Applying the moving average concept once again and estimating P with the 1968 to 1972 data, the forecast displayed in Table 29 is obtained using 1972 as the base year. The forecast of the community college enrollment deviates further than it did in the two previous forecasts for 1973. The forecast for both the upper and lower divisions of the colleges and universities improved. The size of the error measurements for the enrollment prediction decreased, as indicated in Table 30. In addition, the null hypothesis that the prediction error is zero is not rejected at the .05 level of significance. The question arises again as to whether this is a one time fluctuation in the community college enrollments or whether it is the leading impulse of a new enrollment pattern.

Considering each of the separate forecasts, the student flow method generated some fairly accurate forecasts in total, but some of the individual forecasts were somewhat inaccurate. The amount of error was not too large in most instances and Theil's inequality coefficient remained reasonable small for all the forecasts. It would appear that for the highly aggregated, three state model, the Markov student flow method generates reasonable good forecasts.

Expanded Student Flow Model

One of the main advantages of the student flow model is that, conceptually, the number of states can be easily expanded. Expanding the number of states, i.e., having subgroups of institutions instead

Table 29. Three state Markov enrollment forecast for 1973.

Year		Community College	BACCALAUREATE Fresh Soph	Junior Senior	Total
1972	A	136,121	85,112	79,447	164,559
1973	P	143,113	86,195	80,697	166,892
	A	151,652	85,772	77,555	163,327
	R	-.056	.005	.041	.022

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Table 30. Statistical summary for the three state enrollment forecast for 1973.

YEAR	RELATIVE ERROR AVG	STD DEV	t VALUE	PREDICTION ERROR AVG	STD DEV	t VALUE	U
1973	-.004	.049	-.128	-1658	6112	-.470	.048

Note: The null hypothesis that the average forecast error is zero at the .05 level of significance is not rejected.

of a single group of each type, should improve the ability to analyze the statewide distribution of future enrollments. However, there are two problems involved with increasing the number of states of the model.

The first problem concerns computational capability. As the number of states increase, the amount of computer memory required for manipulation of the data increases by about a power of 3. This problem can be handled with access to a larger computer or with an additional expenditure of time to employ more sophisticated data handling techniques. An example of this is the comparison of the amount of memory needed to estimate P for the three state Markov model for Phase II--45,000 locations--and the memory needed to estimate Q for Phase I, which contains twelve states--70,000 locations. This is one of the reasons for separating the flow model into two phases, the avoidance of excessive core memory requirements.

The second problem concerns the availability of data. This is a problem with the amount of detail contained in the data. Additional detail would be required concerning the number of transfers between institutions in order to greatly expand the number of states. This is particularly true with respect to the number of transfers from the community colleges and the baccalaureate institutions. Problem two is evident when we separate the two groups of institutions into several classifications. This problem is also evident when we separate the community college group into the vocational-technical or non-degree oriented classification and the transfer or degree oriented classification. This will be explored in the following presentation of relevant forecasts from the Markov student flow model.

Separating Transfer and Terminal Enrollments

Expanding the model to include the dichotomy between transfer and non-transfer students at the community colleges can only be accomplished with historical data from 1968 and following years. With this constraint, the transition matrix, P, was estimated using data from 1968 to 1971. This transition probability matrix was used in Equation 3.1 to forecast the enrollments of the statewide system following the base year 1971. The results of this forecast are shown in Table 31.

Table 31. Student flow model expanded to include community college vocational enrollments with the transition matrix estimated from 1968 to 1971 data.

		COMMUNITY COLLEGE			BACCALAUREATE		
Year		Vocat	Trans	Total	Fresh Soph	Junior Senior	Total
1971	A	49,612	82,577	132,189	87,352	81,826	169,178
1972	P	54,125	84,750	138,875	87,595	84,590	172,185
	A	59,485	76,636	136,121	85,112	79,447	164,559
	R	-.090	.106	.020	.029	.065	.046
1973	P	62,339	92,724	155,063	87,981	86,454	174,435
	A	70,988	80,664	151,652	85,772	77,555	163,327
	R	-.122	.150	.022	.026	.155	.068

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

This and the following forecasts were generated in a slightly different manner than those in the preceding discussion. Instead of using the forecast values for the first-time entrants, the actual, known values were used in the student flow model forecasts. This was done in order to avoid the problem of partitioning the aggregare first-time entrant forecasts into individual group forecasts for any possible grouping of institutions.

The result of using the actual first-time entrants instead of forecast first-time entrants should be a more accurate forecast when compared to the actual enrollment figures. The error due to the forecast of first-time enrollments has been excluded from the forecast comparison of total enrollments. However, the amount of error in the community college enrollments is a -9.0 percent for the non-transfer enrollments and 10.6 percent for the transfer enrollment in the first year forecast. The amount of error increases even more in the second year of the forecast. The size of the forecast errors are also indicated by the increased size of Theil's inequality coefficient. The results of the t-tests, displayed in Table 32 indicate that the forecase errors are not different from zero at the five percent level of confidence. However, the t-test results do not imply accurate forecasts since the standard deviations of the error measures are large and the individual prediction descrepancies are large.

Adding one more year of enrollment data to the estimation process for P, i.e., using enrollment data from 1968 to 1972, a new transition matrix P is obtained. Using actual first-time enrollments, as in the above, the forecast in Table 33 is obtained. The accuracy of this one year forecast is an improvement over the previous first year forecast in Table 31. The accuracy of the community college forecast has improved by a factor of two. This is evident in the comparison of the statistical summaries in Tables 32 and 34. The

Table 32. Statistical summary for the expanded flow model with the transition matrix estimated from 1968 to 1971 data.

	RELATIVE ERROR			PREDICTION ERROR			
Year	Avg	Std Dev	t Value	Avg	Std Dev	t Value	U
1972	.027	.084	.650	2595	5781	.898	.074
1973	.042	.121	.695	3630	9158	.793	.111

Note: The null hypothesis that the average forecast error is zero at the .05 level of significance, n=4, is not rejected.

Table 33. Student flow model expanded to include community college vocational enrollments with the transition matrix estimated from 1968 to 1972 data.

	COMMUNITY COLLEGE			BACCALAUREATE		
Year	Vocat	Trans	Total	Fresh Soph	Junior Senior	Total
1972 A	59,485	76,636	136,121	85,112	79,447	164,559
1973 P	67,904	84,969	152,873	85,728	80,697	166,425
A	70,988	80,664	151,652	85,772	77,555	163,327
R	-.043	.053	.008	-.001	.041	.019

P=Predicted Value
A=Actual Value
R=Relative Error $\frac{P-A}{A}$

relative error, the error in units, and their standard deviations for the latter forecast are an improvement over those of the preceding forecast.

Table 34. Statistical summary for the expanded flow model with the transition matrix estimated from 1968 to 1972 data.

	PREDICTION DISCREPANCY			PREDICTION ERROR			
Year	Avg	Std Dev	t Value	Avg	Std Dev	t Value	U
1973	.012	.044	.057	1080	3329	.049	.039

Note: The null hypothesis that the forecast error is zero at the .05 level of significance, n = 4, is not rejected.

The improvement in accuracy may be the result of one of several factors. The first is that the start up period of any process will contain errors in the processing activity and the reporting of the activity. This is evident in the errors in the needed breakdown of data in every year the data has been collected. (See the Appendix for the community college enrollment data.) For example, in 1971 only 25 of the 29 community colleges reported the breadkown for the first-time enrollments. In addition, only 28 of the community colleges reported the total enrollment breakdown. The liberal arts vs. vocational-technical data breakdown was collected in 1968, however, it was not until 1973 that all schools reported the enrollment breakdown. The second reason for poor data was that community college administrators were playing with the enrollment numbers in order to obtain more state funds. The extent to which this activity was practiced is unknown. However, this practice was reported in two consecutive Audit Reports from the Office of the Auditor General for the State of Michigan.[4] A third factor may be the result of the short period of time over which the data has been collected. When using only a few observations as the basis of an estimate, the impact of a single deviation from the norm is significant. As the number of observations increases, this impact of a single deviation on the estimated value is reduced. This may be the reason that the accuracy of the second forecast was better. The number of observations had increased by one.

The process of modeling the community college enrollments in two groups causes no problems in the Markov student flow model. Conceptually, it has the appealing feature of segmenting the community college enrollment population into two separate and distinct groups. The two groups require different types of educational facilities and require different faculty structures. This additional knowledge can assist the various levels of administration in planning for the two different types of students.

Forecasting System Subgroups

The student flow model is not limited to the forecasting of enrollments for each institutional type as a whole. The institutional classifications could be divided into groups or subgroups within the institutional type. One possible subgrouping is illustrated in Table 35. The community colleges are divided on the basis of total enrollment growth. Group II community colleges individually experienced a significant growth in total enrollment and Group I did not experience

Table 35. First institutional subgrouping.

Community Colleges	Baccalaureate Institutions
Group I	Group I
Alpena	Central
Bay de Noc	Eastern
Delta	Ferris
C.S. Mott	Grand Valley
Gogebic	Lake Superior
Grand Rapids	Michigan Tech
Jackson	Northern
Kellogg	Oakland
Lansing	Saginaw Valley
Macomb County	Western
Monroe County	
Montcalm	
North Central	
Northwestern	
Oakland	
St. Clair County	
Schoolcraft	
Southwestern	
Washtenaw	
Group II	Group II
Glen Oaks	Michigan State University
Henry Ford	University of Michigan
Highland Park	Wayne State University
Kalamazoo	
Kirtland	
Lake Michigan	
Mid Michigan	
Muskegon	
Wayne County	
West Shore	

a significant amount of growth. The four year institutions are divided into the senior institutions, Group II, and the remainder of the four year schools in Group I. The transition matrix for this subgrouping and the two following were estimated with enrollment data from the years 1966 to 1970. The forecast for these institutional subgroups is displayed in Table 36. As in the preceding forecast for the vocational-technical breakdown, this forecast was generated using the actual first-time enrollments for each subgroup. The following forecasts discussed in this section are also generated using actual first-time enrollments. This should be remembered as the relative error between the forecasts and actual enrollments is discussed. These forecasts should be more accurate since no error enters the forecast results due to an error in the number of first-time entrants. The first year forecast for the baccalaureate institutions is quite good for both subgroups and the total forecast error for the baccalaureate institutions in only 1.1 percent. The community college forecasts were only partially good. The total enrollment group forecast resulted in a large forecast error. The first year forecast was off by 12.4 percent and in the third year the forecast error was 31.5 percent. The discrepancy of the total enrollment forecast, as measured by Theil's inequality coefficient was not exceedingly large for the first and second years of the forecast. The null hypothesis that the forecast error is zero is not rejected for the 1971 enrollment forecast. However, the null hypothesis is rejected for the 1972 and 1973 enrollment forecasts. The statistical summary of this forecast appears in Table 37.

Table 36. Enrollment forecast for the first institutional subgroups.

Year		COMMUNITY COLLEGE			BACCALAUREATE				
					Group 1		Group 2		
		Group 1	Group 2	Total	Fresh Soph	Junior Senior	Fresh Soph	Junior Senior	Total
1970	A	86,503	40,118	126,621	47,200	35,699	39,688	42,464	165,051
1971	P	96,747	44,276	141,024	48,028	39,295	40,116	43,564	171,004
	A	92,802	39,387	132,189	46,449	47,519	40,903	44,307	169,178
	R	.043	.124	.067	.034	.047	-.019	-.017	.011
1972	P	99,075	49,770	148,845	47,640	41,989	40,169	44,518	174,316
	A	94,395	41,726	136,121	44,897	38,393	40,215	41,054	164,559
	R	.050	.193	.093	.061	.094	-.001	.084	.059
1973	P	108,085	58,226	166,311	47,109	43,683	40,724	45,251	176,767
	A	107,377	44,275	151,652	43,973	35,897	41,799	41,658	163,329
	R	.007	.315	.097	.071	.217	-.026	.086	.082

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Table 37. Statistical summary for enrollment forecast of the first institutional subgroups.

	RELATIVE ERROR			PREDICTION ERROR			
Year	Avg	Std Dev	t Value	Avg	Std Dev	t Value	U
1971	.035	.052	1.649	1777	2339	1.861	.052
1972	.080	.064	3.041*	3747	2639	3.477*	.083
1973	.112	.130	2.104	4683	5442	2.108	.118

* Accept the alternative hypothesis that the average forecast error is not zero at the .05 level of significance, n = 6.

A second subgrouping of the institutions is indicated in Table 38. In this attempt to locate a relevant grouping of institutions, the community colleges were separated on the basis of geographical location and the universities were separated on the basis of increasing first-time enrollments. Group II of the community colleges are all located in the southeastern third of Michigan. Group II of the baccalaureate institutions consists of those colleges that individually experienced a significant growth in first-time enrollments. The student flow forecast for this grouping is displayed in Table 39. The statistical summary appears in Table 40.

The results of the t-tests for the second subgrouping indicate that the prediction error is not significantly different from zero. However, comparing the values of U in Table 40 with the values of U in Table 37, the conclusion is that the forecasts of the second subgrouping are less accurate than the first since the inequality coefficients are larger for each of the three years forecast. This

Table 38. Second institutional subgrouping.

Community Colleges	Baccalaureate Institutions
Group I	Group I
Alpena	Central
Bay de Noc	Eastern
Glen Oaks	Ferris
Gogebic	Lake Superior
Grand Rapids	Michigan State
Kalamazoo	Michigan Tech
Kellogg	Northern
Kirtland	Wayne State
Lake Michigan	Western
Mid Michigan	
Montcalm	
Muskegon	
North Central	
Northwestern	
Southwestern	
West Shore	
Group II	Group II
Delta	Grand Valley
C. S. Mott	Oakland
Henry Ford	Saginaw Valley
Highland Park	University of Michigan
Jackson	
Lansing	
Macomb County	
Monroe County	
Oakland	
St. Clair County	
Schoolcraft	
Washtenaw	
Wayne County	

Table 39. Enrollment forecast for the second institutional subgroups.

		COMMUNITY COLLEGE			BACCALAUREATE				
					Group 1		Group 2		
Year		Group 1	Group 2	Total	Fresh Soph	Junior Senior	Fresh Soph	Junior Senior	Total
1970	A	27,714	98,907	126,621	69,810	60,314	17,078	17,849	165,051
1971	P	29,135	111,945	141,079	70,315	63,776	17,793	19,517	171,402
	A	29,495	102,694	132,189	69,116	63,337	18,236	18,489	169,178
	R	-.012	.090	.067	.017	.007	-.024	.056	.013
1972	P	30,162	119,083	149,245	68,845	66,516	18,886	20,984	175,231
	A	30,038	106,083	136,121	65,756	61,205	19,356	18,242	164,559
	R	.004	.123	.096	.047	.087	-.024	.150	.065
1973	P	31,248	135,527	166,775	68,080	68,193	19,594	22,461	178,329
	A	32,035	119,617	151,652	64,364	58,399	21,408	19,156	163,327
	R	-.025	.133	.100	.058	.168	-.085	.173	.092

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Table 40. Statistical summary for enrollment forecast of the second institutional subgroups.

	RELATIVE ERROR			PREDICTION ERROR			
Year	Avg	Std Dev	t Value	Avg	Std Dev	t Value	U
1971	.022	.043	1.262	1852	3688	1.230	.065
1972	.064	.068	2.323*	3966	4901	1.982	.101
1973	.070	.107	1.612	5021	6730	1.828	.126

*: Accept the alternative hypothesis that the average forecast error is not zero at the .05 level of significance, n = 6.

conclusion is supported by the comparison of the relative error standard deviation; they are all larger in the second subgrouping forecasts.

A third subgrouping of the institutions is indicated in Table 41. This grouping of institutions divides the community colleges such that Group II contains only institutions that admitted their first students in 1966 or following years. The community colleges undergoing start up conditions have been segregated from the remainder of the community colleges. The baccalaureate institutions are divided such that Group II contains those institutions that experienced a significant decline in first-time enrollments. The resulting forecast for this subgrouping is displayed in Table 42. The community college forecasts are interesting. The forecasts for the start up group of community colleges have a very high degree of error. The prediction discrepancy is 45 percent in the third year of the forecast. However, the forecasts for the remaining community colleges are fairly accurate and the

Table 41. Third institutional subgrouping.

Community Colleges	Baccalaureate Institutions
Group I	**Group I**
Alpena	Central
Bay de Noc	Ferris
Delta	Grand Valley
C.S. Mott	Lake Superior
Gogebic	Michigan Tech
Grand Rapids	Oakland
Henry Ford	Saginaw Valley
Highland Park	University of Michigan
Jackson	Western
Kellogg	
Lake Michigan	
Lansing	
Macomb County	
Muskegon	
North Central	
Northwestern	
Oakland	
St. Clair Shores	
Schoolcraft	
Group II	**Group II**
Glen Oaks (1967)	Eastern
Kalamazoo (1968)	Michigan State
Kirtland (1968)	Northern
Mid Michigan (1968)	Wayne State
Monroe County (1966)	
Montcalm (1966)	
Southwestern (1966)	
Washtenaw (1966)	
Wayne County (1969)	
West Shore (1968)	

Table 42. Enrollment forecast for the third institutional subgroups.

Year		COMMUNITY COLLEGE			BACCALAUREATE				
					Group 1		Group 2		
		Group 1	Group 2	Total	Fresh Soph	Junior Senior	Fresh Soph	Junior Senior	Total
1970	A	101,492	25,129	126,621	44,772	38,417	42,116	39,746	165,051
1971	P	111,152	30,612	141,764	46,218	42,048	41,793	41,609	171,668
	A	106,776	25,413	132,189	45,355	39,829	41,997	41,997	169,178
	R	.041	.205	.072	.019	.056	-.005	-.009	.015
1972	P	114,569	36,012	150,581	46,764	44,919	40,786	43,102	175,570
	A	108,368	27,753	136,121	44,759	39,711	40,353	39,736	164,559
	R	.057	.298	.106	.045	.131	.011	.085	.067
1973	P	127,314	41,708	169,022	47,271	46,952	40,181	44,149	178,553
	A	122,889	28,763	151,652	45,877	39,015	39,895	38,540	163,327
	R	.036	.450	.115	.030	.203	.007	.146	.093

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

relative error does not exceed 5.7 percent for any of the three years. The forecast error of the start up group is probably the result of the instability of the group as individual institutions began to admit students. Part of the observed forecast error could also be the result of inaccurate data reporting by inexperienced school administrators. The null hypothesis of a prediction error of zero is rejected at the .05 level of significance for all three years. This is indicated in the statistical summary in Table 43.

Table 43. Statistical summary for enrollment forecast of the third institutional subgroups.

	RELATIVE ERROR			PREDICTION ERROR			
Year	Avg	Std Dev	t Value	Avg	Std Dev	t Value	U
1971	.051	.079	1.575	2011	2357	2.090*	.052
1972	.104	.103	2.483*	4245	2870	3.623*	.088
1973	.145	.168	2.125*	5433	4616	2.883	.112

*: Accept the alternative hypothesis that the average forecast error is not zero at the .05 level of significance, n = 6.

Expanding the second and third groupings to include the dichotomy between the transfer and non-transfer oriented students at the community colleges, the following results are obtained. Using enrollment data from the years 1968 to 1971, the transition matrix P was generated for the subgrouping of Table 38. The forecast generated with this transition matrix is displayed in Table 44. An interesting result is that the total enrollment for the community colleges is quite accurate, but individual forecasts for each part of the

Table 44. Enrollment forecast for the second institutional subgroups including community college vocational enrollments.

		COMMUNITY COLLEGE					BACCALAUREATE				
		Group 1		Group 2			Group 1		Group 2		
Year		Vocat	Trans	Vocat	Trans	Total	Fresh Soph	Junior Senior	Fresh Soph	Junior Senior	Total
1971	A	9,501	19,994	40,432	62,262	132,189	69,116	63,337	18,236	18,489	169,178
1972	P	9,841	20,516	44,677	63,747	138,781	68,149	64,790	19,441	19,528	171,908
	A	10,430	19,608	49,607	56,476	136,121	65,756	61,205	19,356	18,242	164,559
	R	-.056	.046	-.099	.129	.020	.036	.059	.004	.071	.045
1973	P	10,670	20,673	52,050	71,604	154,997	64,704	65,355	20,235	20,637	173,931
	A	12,373	19,662	58,615	61,002	151,652	64,364	58,399	21,408	19,156	163,327
	R	-.138	.051	-.112	.174	.022	.052	.119	-.055	.077	.065

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Theil's Inequality Coefficient

1971	.081
1972	.118

enrollment contains descrepancies that are too large in even the first year of the forecast. The inequality coefficients for each of the forecast years are larger than those of the preceding first and second year.

The forecast results for the third subgrouping is displayed in Table 45. The accuracy of the total community college enrollment is better than the previous forecast. However, the accuracy of some of the individual forecasts are not as good. For example, the forecast error for Group II in 1973 is 35.9 percent. The reasons for the large amount of error between the forecast and the observed enrollments are the same as those discussed earlier.

The results of the student flow model when expanded to include additional states were not as good as those of the restricted model. The problems of using the expanded model and possible means of reducing the size of the forecast error will be discussed in the next chapter.

Table 45. Enrollment forecast for the third institutional subgroups including community college vocational enrollments.

		COMMUNITY COLLEGE					BACCALAUREATE				
		Group 1		Group 2			Group 1		Group 2		
Year		Vocat	Trans	Vocat	Trans	Total	Fresh Soph	Junior Senior	Fresh Soph	Junior Senior	Total
1971	A	42,113	64,663	7,837	17,576	132,189	45,355	39,829	41,997	41,997	169,178
1972	P	44,610	64,872	9,468	19,269	138,219	46,197	41,834	41,255	43,971	173,257
	A	48,782	59,586	10,922	16,831	136,121	44,759	39,711	40,353	39,736	164,559
	R	-.086	.089	-.133	.145	.015	.032	.053	.022	.107	.053
1973	P	50,811	70,943	11,190	21,050	153,993	46,883	43,845	40,818	45,693	177,238
	A	57,712	65,177	13,276	15,487	151,652	45,877	39,015	39,895	38,540	163,327
	R	-.120	.088	-.157	.359	.015	.022	.124	.023	.186	.085

P=Predicted Value
A=Actual Value

R=Relative Error $\frac{P-A}{A}$

Theil's Inequality Coefficient

1971	.077
1972	.114

Chapter V Notes

[1]Michigan State University, Graduate School of Business Administration, Michigan Statistical Abstract (10th ed.; East Lansing, Michigan: 1974), p. 155.

[2]Michigan, Department of Education, Financial Requirements of Public Baccalaureate Institutions and Public Community Colleges (Lansing, Michigan: 1971).

[3]Ibid., p. 3.

[4]Office of the Auditor General, State of Michigan, Michigan Community Colleges Enrollment Audit (Lansing, Michigan: 1971).
Office of the Auditor General, State of Michigan, Michigan Community Colleges Enrollment Audit (Lansing, Michigan: 1972).

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Many models have been developed and used to forecast enrollments in higher education. Examples of several of these models were described in Chapter II. Some simplistic models have been developed for forecasting models for statewide systems, although, their use has been quite limited. Most of the developmental effort has been in the area of institutional flow models due to the accessibility and availability of good, accurate historical data at the institutions where the research was being performed. The development of a Markov student flow model requires the estimation of the Markov transition matrix. This is most easily accomplished by using the individual student data with the maximum likelihood technique. However, this is not practical for a statewide system of higher education.

The purpose of this research was to develop a Markov student flow model that did not rely on the maximum likelihood technique and individual student data and which could be used to forecast statewide student enrollments from aggregate data. The Markovian flow model is the most recent type of enrollment model, with research and development continuing on its applicability to the forecasting of student enrollments. The application of the Markov student flow model to the forecasting of statewide student enrollments was described. The estimation of the Markov transition probability matrix for this application does not utilize the maximum likelihood technique. Instead, a quadratic programming, least squares technique was used on aggregate enrollment data in order to estimate the substochastic Markov transition matrix. In addition, the model utilized forecast first-time enrollments generated by applying the same procedure to elementary and secondary enrollments. The other Markov models reviewed did not discuss the problem of first-time entrants or obtained the first-time entrants from outside the Markov student flow model.

The application of the Markov student flow model to the forecasting of first-time entrants into the system of higher education begins with the forecasting of statewide elementary and secondary enrollments. The flow concept is used to forecast the first-time entrants by tracing the flow of the elementary and secondary students into the condition or state of entering an institution of higher education. The model was compared to the progression ratio method currently used to forecast statewide elementary and secondary enrollments. The model generated slightly more accurate forecasts of the student enrollments than did the progression ratio method for statewide elementary and secondary enrollments for the states of Michigan and New York. However, the difference was statistically significant only in the fourth year forecast for New York.

The statewide Markov student flow model enrollment forecasts were compared with forecasts generated by two other methods. The model provides a much more accurate enrollment forecast than does the population ratio method. The comparison with a multiple linear regression model did not establish the clear superiority which the comparison to the population ratio method established. The model generated slightly better results than did the multiple linear regression model when evaluated with Theil's inequality coefficient. The Markov student flow model was then used to generate forecasts of several different groups of institutions and forecasts of community college enrollments separated into liberal arts and vocational-technical enrollments.

Conclusions

This student flow model appears to represent the characteristics of the student flow mechanism accurately. Comparing the results generated by the model for elementary and secondary enrollments with the results for higher education, the forecasts of the elementary and secondary were more accurate. The average prediction discrepancy for the elementary and secondary statewide school systems was in the range of one to two percent, even into the third year of the enrollment forecast. In contrast to this, the higher education average prediction discrepancy ranged from two to seven percent for the three state model. Increasing the number of states in the model resulted in increasing the size of individual discrepancies to over 30 percent in some cases.

The statewide elementary and secondary education system is basically a closed system. The students enter the system and remain in the system until their late teens. Dropouts are permitted only after a legally determined age. This results in a relatively smooth flow pattern of students from the entry level to a dropout point or graduation. In addition, the educational influence for the continuing of education into the higher level appears to be relatively constant.

The system of higher education is much more open, and as a result, much less stable. The environment of higher education--the economy and societal attitudes--has a significant impact on enrollments. This is evident in the fluctuations in the enrollments when compared with the smoothed model forecasts and in the regression parameters identified by the regression model. The more open the system, the more difficult it becomes to accurately forecast the future of the system.

Expanding the model to segment the community colleges and universities into several different groups did not increase the accuracy of the forecasts. The overall accuracy, as measured by Theil's inequality coefficient, did not significantly improve. However, some of the segment forecasts developed a significant error over several time periods. There are several possible reasons for the errors in some of the forecast segments.

The first reason is that the change in economic conditions and the elimination of the draft caused a significant change in enrollment patterns. Some institutions were able to adapt quickly to these changes and maintain their historical enrollment patterns. Other schools were not able to adapt and lost enrollments as the total number of students declined.

The second reason is that the data collection process may not have been very accurate during its initial years. This would result in larger errors for forecasts using parameter values estimated from data collected in the early years. This is indicated by the decrease in the error for the first year forecasts as the base years for estimating P were moved toward the present, dropping the initial years from the estimation base. In addition, there are start up errors inherent in the data for the community colleges. One type of start up error is in the collection of the data, just as for the baccalaureate institutions. The shift to the collection of data on the numbers of students enrolled in academic and vocational-technical programs resulted in inaccurate counts of the two classifications of students. A second type of start up error exists as a result of some of the community colleges not being open before the base years used in the estimation process. Of the twenty-nine community colleges currently operating, ten have opened their doors during the base years; four opened in 1966, one in 1967, four in 1968, and one in 1969. The enrollments of these schools would cause the enrollment trend to appear to be increasing at a relatively faster rate than it actually was. When sufficient data is available following the start up conditions and when the unstable economic and societal conditions can be incorporated into the model, expansions of the model to the point of forecasting the enrollments of individual institutions should become possible.

The validity of the Markov student flow model is still somewhat open to question. The model generated good results when used to generate first-time enrollments. Reasonably good results were obtained with the three state application to the Michigan system of public higher education. If we apply the generally accepted definition of model validation that "Validation tests whether a simulation model reasonably approximates a real system."[1] or that validation ". . . consists of testing the model's ability to predict the behavior of the system under study."[2] then we do not have enough historical data with which to statistically validate the predicted values against actual values. The comparison of predicted enrollments with actual enrollments was attempted over a short time interval with some encouraging results from the limited comparison.

Uses of the Model

If we have a usable method of forecasting statewide, higher education enrollments, the question arises as to the possible uses of these forecasts in statewide educational decision making. Enrollment forecasts are the primary basis for all educational planning. Some estimate as to the nature and size of the future student population of the higher educational system is fundamental to all areas of educational planning. The major areas of concern for higher education planning can be grouped under the five interdependent areas of faculty, facilities, curricula, finances, and student enrollment. Some aspect of these five areas will be affected by decisions made at the state level by coordinating or governing agencies. Only three states, Delaware, Vermont, and Nebraska, have no legal state agency for coordinating or controlling higher education. Of the other states, 25 have coordinating boards, and 22 have governing boards for administration at the state level.[3] Each of these agencies is involved in the coordination or control of the institutions of higher education. For example, the Constitution of the State of Michigan provides that the State Board of Education "shall serve as the general planning and coordinating body for all public education, including higher education, and shall advise the legislature as to the financial requirements in connection therewith."[4] With regard to the fulfillment of this constitutional obligation, the Michigan State Board of Education established a set of thirty-eight goals as guides in the direction of developing an effective and efficient system of public higher education. Three goals are particularly relevant to enrollment forecasts.

Goal 6. Since revisions of long-range enrollment projections are necessary in determining the need for educational programs, space, and faculty, and because of the important variables affecting the college-going rate, it is the responsibility of the State Board of Education to maintain updated long-range projections of potential and probable student enrollments.[5]

Goal 36. The importance of annual revision of projections for operations cannot be stressed too strongly because conditions constantly change. Therefore, in keeping with its constitutional mandate to advise the Legislature, the State Board of Education will carry on a continuous study of the operating needs of both the baccalaureate institutions and community and junior colleges.[6]

Goal 38. The projected costs of facilities in terms of future enrollments and programs is an important undertaking if sufficient student spaces are to be available. The State Board of Education will submit updated annual capital outlay projections to the Legislature, consistent with the constitutional mandate to advise concerning the financial requirements of higher education.[7]

The activities implied by each of these three goals are involved with each of the planning areas identified earlier. With regard to faculty, planning must consider such factors as desired levels of teaching and research expertise, sources of teaching manpower, and the pattern of distribution of the manpower among the academic fields of study for the coming years. The level of expertise, source, and distribution of the faculty should be considered with some knowledge of the future student population. If more students take the vocational or technical programs at the community college level, faculty with the desired background will need to be trained. The distribution of faculty among the various institutions and areas of study depends on the pattern of choices made by students regarding their fields of study.

In facilities planning, total space requirements--the allocation of that space among classrooms, laboratories, faculty office space, and administration--are a function of the distribution and characteristics of the future student population. The duration of the expected population trend can influence capital budget considerations with regard to new academic facilities. Consideration of the major geographic distributions of future students would aid in decisions regarding the possible location of new facilities or new institutions.

The decision to expand or offer new curricula is partially dependent on future student enrollments and on the societal needs for trained manpower. Although the decision to offer a new field of study at an institution depends more on the goals of the system of higher education, the size of the new program and the amount of resources assigned to the program depend on expected future enrollments.

A large portion of the finances for the higher education institutions is from student tuition. Future financial planning is directly dependent on the forecasts of future enrollments. In addition, the appropriations for the individual schools is based on two items, the predetermined activity of the institution, eg. teaching vs. research, and the number of students expected to attend the institution during the next fiscal year.

The forecasting of enrollments is a basic tool for the statewide planning and coordination of the previously mentioned four academic areas. However, the enrollment forecasts can also be used in the planning for future enrollments. If the enrollment forecast does not meet with the goals established for the system, then action can be planned and implemented in order to achieve specific enrollment goals.

Planning Applications

In the context of the planning of future enrollment goals or enrollment policies, the Markov student flow enrollment forecasting model can be used in a planning mode. In this mode the model can be used to answer or explore various kinds of "what if?" questions that might be asked by educational planners. One such question could be "what will be the impact of an effective, nondiscriminatory enrollment policy?" A nondiscriminatory enrollment policy is one that would result in a racial mix in the higher education enrollment equal to the racial mix of the state population. If it can be assumed that the retention rate among the different racial groups is the same, than a plan that results in a balanced racial mix for the first-time entrants would result in racially balanced enrollments after four years of effective operation of the new policy. However, the question to be answered is "what will be the enrollments after the implementation of this policy?" This can be determined with the use of the forecasting model. The only aspect of the enrollment forecast that will change will be the number of first-time entrants. The number of entrants may increase to reflect the increased accessibility of higher education for minority groups. The effect of this policy change can be traced with the forecasting model to determine the effects on future enrollments.

If the policy was fully effective the first year of operation, then in four years the system of higher education would be racially balanced. However, this situation is unlikely. It would probably take several years of operation in order to become fully effective. This process of a gradual increase in minority students could also be easily simulated with the enrollment forecasting model to determine the effects of the policy on total enrollments.

A second planning use of the model could be the prediction of enrollment changes due to increased financial aid. Financial aid programs have been developed to help economically disadvantaged student stay in school. Increased financial aid should cause more students to remain in school to complete their education. The effect of the aid is to increase the retention rate, which would be reflected in an increase in the transition probabilities within the student flow model. The effect of changes in future aid programs on future enrollments can be simulated with the forecasting model. This and the preceding example illustrates the use of the Markov student flow model as an aid in the planning and decision making process for statewide higher education.

The use of enrollment forecasts are essential in the decision process necessary to control or coordinate the activities of the institutions of higher education. Considering the underlying importance of the enrollment forecast in the decision processes for higher education several suggestions are made in order to improve or expand the statewide enrollment forecasts.

Recommendations

At this stage of development, the student flow model still needs improvement. Additional time series data could be used for additional testing and validation of the model. The additional historical data would allow a longer time period over which to validate the forecasts of the model. The additional data would also allow two additional tests to be made on the transition matrix. The first test would be to determine if the transition probabilities are stationary. Any time dependency of the transition probabilities would have a significant impact on the model's forecast accuracy in the more distant future. A second test that could be carried out with the additional data would be to determine if there is any correlation between the state of the economy and any changes in the transition probabilities. The identification of a dependency relationship between the transition probabilities and the economy could lead to improved short term forecasts.

The above suggestions only require the passage of time, and the research can wait until the additional data is available. The final recommendation pertains to the types of data collected by the educational agencies. The availability of transfer data for the baccalaureate institutions would probably improve the accuracy of the model's forecasts. This data may already be available at some of these institutions, but it would be very difficult for an individual research to collect. The Markov student flow model can easily be modified to incorporate the counts of the number of transfers from the community colleges to the colleges and universities. This improvement could result in more accurate forecasts for the baccalaureate institutions.

Chapter VI Notes

[1]George S. Fishman and Philip J. Kiviat, "The Statistics of Discrete-Event Simulation," in John M. Dutton and William H. Starbuck, Computer Simulation of Human Behavior (New York, N. Y. John Wiley & Sons, 1971), p. 596.

[2]Thomas Naylor, Computer Simulation Experiments With Models of Economic Systems (New York: John Wiley & Sons, 1971), p. 158.

[3]Paul Wing, Statewide Planning for Postsecondary Education: Conceptualization and Analysis of Relevant Information (Boulder, Colorado: National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, March, 1972), p. 1.

[4]Mich. Const. art. i, sec. 3 (1963).

[5]Michigan Department of Education, State Plan for Higher Education in Michigan (Lansing, Michigan: February, 1970), p. 20.

[6]Ibid., p. 58.

[7]Ibid., p. 64.

APPENDIX

AGGREGATE PUBLIC COMMUNITY COLLEGE AND BACCALAUREATE UNDERGRADUATE ENROLLMENTS FOR THE STATE OF MICHIGAN

APPENDIX

COMMUNITY COLLEGE ENROLLMENTS

					LIBERAL ARTS -VS- VOCATIONAL-TECHNICAL					
					First-Time			Total Enrollment		
Year	First-Time	Fresh	Soph	Total	Lib	Voc	Total	Lib	Voc	Total
1966	28,315 (23)			69,504 (23)						
1967	31,864 (24)	49,082 (22)	28,914 (22)	79,250 (24)				32,209	22,226	54,435* (19)
1968	40,653 (28)			95,065 (28)	24,381	13,563	37,944* (27)	59,877	35,188	95,065 (28)
1969	49,145 (29)	71,605 (19)	26,711 (19)	115,791 (29)	31,006	14,652	45,658* (26)	74,221	41,798	116,019* (29)
1970	47,056 (29)	65,650 (25)	31,592 (25)	126,621 (29)	29,492	16,817	46,309* (27)	77,565	48,148	125,713* (29)
1971	53,436 (29)	27,146 (4)	6,585 (4)	132,189 (29)	25,645	19,011	44,656* (25)	71,522	42,970	114,492* (28)
1972	51,333 (29)	28,768 (5)	9,028 (5)	136,121 (29)	25,806	18,349	44,155* (28)	67,258	52,206	119,464* (28)
1973	63,097 (29)			151,652 (29)	36,535	26,562	63,097 (29)	80,662	70,986	151,648* (29)

*: Indicates discrepancy between two enrollment counts.

(): Number of respondents.

Source: Higher Education General Information Survey, Michigan Association of Collegiate Registrars and Admissions Officers, and the Michigan Bureau of Programs and Budget as described in Chapter I.

BACCALAUREATE ENROLLMENTS

Year	First-Time	Fresh	Soph	Fr + So	Junior	Senior	Jr + Sr	Total
1966	30,084			74,741			53,467	128,208
1967	31,463	41,598	33,960	75,558			60,015	135,573
1968	31,358			80,314			69,033	149,347
1969	33,022	41,236	32,296	83,763	30,724	30,434	74,367	158,130
1970	32,633	46,978	39,910	86,888	38,056	40,107	78,163	165,051
1971	33,425	47,752	39,600	87,352	41,047	40,779	81,826	169,178
1972	32,301	47,701	37,411	85,112	40,064	39,383	79,447	164,559
1973	32,533			85,772			77,555	163,327

Source: Higher Education General Information Survey, Michigan Association of Collegiate Registrars and Admissions Officers, and the Michigan Bureau of Programs and Budget as described in Chapter I.

BIBLIOGRAPHY

BIBLIOGRAPHY

Armitage, Peter; Smith, Cyril; and Alper, Paul. Decision Models for Education Planning. London: Allen Lane the Penguin Press, 1969.

Campbell, Robert, and Siegal, Barry N. "The Demand for Higher Education in the United States." American Economic Review. Vol. 57 (June, 1967), pp. 482-494.

Clark, David G.; Huff, Robert A.; Haight, Michael J.; and Collard, William J. Introduction to the Resource Requirements Prediction Model 1.6. Technical Report No. 34A, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education, Boulder, Colorado: 1973.

Corazzini, A. J.; Dangan, D.; and Grabowski, H. G. "Determinants and Distributional Aspects of Enrollment in U.S. Higher Education." Journal of Human Resources. Vol. 7, No. 1 (Winter, 1972), pp. 39-49.

Feller, William. An Introduction to Probability Theory and Its Applications. Vol. I, 3rd ed.; New York: John Wiley & Sons, Inc., 1968.

Galper, Harvey, and Dunn, Robert M., Jr. "A Short-Run Demand Function for Higher Education in the United States." Journal of Political Economy. Vol. 77 (September-October, 1969), pp. 765-777.

Gani, J. "Formulae for Projecting Enrolments and Degrees Awarded in Universities." Journal of the Royal Statistical Society. A126 (1963).

Hadley, George. Nonlinear and Dynamic Programming. Reading, Massachusetts: Addison-Wesley Publishing Co., Inc., 1964.

Heenach, Stephen A. and Feldman, Paul. "Private Demand for Higher Education in the United States." The Economics and Financing of Higher Education in the United States, a Compendum of Papers submitted to the Joint Economic Committee, Congress of the United States, Washington, D.C.: Government Printing Office, 1969, pp. 375-395.

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

Lee, T. C.; Judge, G. G.; and Zellner, A. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. Amsterdam: North-Holland Publishing Co., 1970.

Lovell, C. C. Student Flow Models: A Review and Conceptualization. Technical Report No. 25, National Center for Higher Education Management Systems at Western Interstate Commission for Higher Education, Boulder, Colorado: August, 1971.

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

Michigan Department of Education. Financial Requirements of Public Baccalaureate Institutions and Public Community Colleges. Lansing, Michigan: 1971.

Michigan Department of Education. State Plan for Higher Education in Michigan. Lansing, Michigan: 1969.

Michigan State University, Graduate School of Business Administration. Michigan Statistical Abstract. 10th ed.; East Lansing, Michigan: 1971.

New York State Division of the Budget, Office of Statistical Coordination. New York State Statistical Yearbook. Albany, New York: 1973.

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

Office of the Auditor General, State of Michigan. Michigan Community Colleges Enrollment Audit. Lansing, Michigan: 1971.

Office of the Auditor General, State of Michigan. Michigan Community Colleges Enrollment Audit. Lansing, Michigan: 1972.

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

Oliver, R. M. and Hopkins, D. S. P. "An Equilibrium Flow Model of a University Campus." Operations Research, Vol. 20, No. 2 (March-April, 1972), pp. 249-264.

Oliver, R. M.; Hopkins, D. S. P.; and Armacost, R. An Academic Productivity and Planning Model for a University Campus. Report No. 3, University of California Administrative Studies Project in Higher Education, Berkeley, California: 1970.

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

Perkins, William C. and Paschke, Paul E. "A Simulation Model of the Higher Education System of a State." Decision Sciences, Vol. 4, No. 2 (April, 1973), pp. 194-215.

Radner, R. and Miller, L. S. "Demand and Supply in U.S. Higher Education: A Progress Report." American Economic Review, Vol. 60 (May, 1970), pp. 326-339.

Theil, Henri. Applied Economic Forecasting. Amsterdam: North-Holland Publishing Co., 1966.

Theil, Henri and Rey, Guido. "Quadratic Programming Approach to the Estimation of Transition Probability." Management Science, Vol. 12, No. 9 (May, 1966), pp. 741-721.

U.S. Department of Health, Education, and Welfate, Office of Education, National Center for Educational Statistics. Projections of Educational Statistics to 1979-1980. Washington, D.C.: Government Printing Office, 1971.

Wolfe, P. "The Simplex Method for Quadratic Programming." Econometrica, Vol. 27, No. 3 (July, 1959), pp. 382-398.

Typed and Printed in the U.S.A.
Professional Thesis Preparation
Cliff and Paula Haughey
144 Maplewood Drive
East Lansing, Michigan 48823
Telephone (517) 337-1527