A method of projecting an institution's demand for higher education programs is presented. Demand was measured by the dependent variable of number of applications received by the institution. A model of independent variables that determine the demand was presented, including both exogenous variables, such as demographic patterns, national and regional economic conditions, the cost of competing institutions, and the economic benefit of a college education; and endogenous variables, such as costs of the institution, financial aid available, alumni assistance, students enrolled, recruiting effort, and public relations effort. Two anonymous case studies of institutions that were analyzed by this method are presented. A regression analysis was applied, and those measures of the variables that explain the most variation in the dependent variable were determined. It was concluded that this approach is one with much potential for specific institutions, but that the results obtained are not yet generalizable. It is suggested that further research with several institutions be undertaken to determine those factors which seem to be most determinant of the demand for college education. (Author/PHR)
Projecting Applications for Admissions in Higher Education

By

Eliot S. Elfner, Ph.D.
Associate Professor of Business Administration
St. Norbert College
De Pere, WI 54115

Paper presented to the fourteenth Annual International Conference of the Society for College and University Planning
July, 1979, Kansas City, Missouri.
ABSTRACT

This study presents a method of projecting an institution's demand for higher education programs. Demand was measured by the dependent variable of number of applications received by the institution. A model of independent variables which determine the demand was presented, including both: exogenous variables - such as demographic patterns, national and regional economic conditions, the cost of competing institutions, and the economic benefit of a college education, and endogenous variables; such as costs of the institution, financial aid available, alumni assistance, students enrolled, recruiting effort, and public relations effort. Two case studies of institutions which were analyzed by this model were then presented. A regression analysis was applied, and those measures of the variables which explained the most variation in the dependent variable were determined. The resulting models had $R^2$ for one institution in the range of 85% - 90%. For the other case institution, the $R^2$ range was 97% - 99%. It was concluded that this approach was one with much potential for specific institutions, but that results obtained were not yet generalizable. It is suggested that further research with several other institutions be undertaken to determine those factors which seem to be most determinant of the demand for college education.
The demographic patterns of the eighties indicate a clear decline in the number of highschool graduates. Those colleges and universities which draw heavily on the traditional student for the majority of their enrollments are undeniably facing a period of decreasing demand and potential enrollment declines. The long range planning efforts of these institutions require accurate assessments of potential demand in the face of expected declines.

Several current authors concern themselves with the planning and forecasting problem. "Recent higher education enrollment trends suggest a need for new strategies and techniques for planning, especially in the area of quantitative systems and forecasting methods" [Healey and Brown, 1978, p. 417]. Salley [1979, p. 323] notes that budget forecasting has become more difficult recently, usually because planners assert that "... enrollment variations since the early 1970's have been erratic." Healey and Brown emphasize "ratio-smoothing", while Salley presents a time-series analysis. Other studies have determined that most forecasting of demand is done by long term trend projection, often through extrapolation [Wing, 1974; and Rudd, 1979].

Healey and Brown [1978] list several alternative forecasting methodologies: first are those made strictly on the qualitative judgement of planners; next is the use of Markov processes to forecast future enrollments within educational systems; a third is the use of trend lines and time series analyses; a specific simulation approach is then described; they present and explore
in detail the ratio-smoothing technique; and finally, they suggest the use of indicator methods employing multiple regression analysis.

The purpose of this paper is to present the results of a demand analysis using the indicator method with multiple regression techniques. First, the model is presented, the measures of the independent variables specified, and the regression methodology is described. Finally, two case studies are presented, and a discussion of the results follows.

Model:

The general model developed for this study posits that demand is a function of several independent "indicator" variables. These independent variables are both exogenous and endogenous in nature. Included among the exogenous variables are: demographic patterns of college attendees (DP), the general health of the nation/region's economy (N/REC), the cost of attending competing institutions (CCI), and the general economic benefits of a college education (ECB). The endogenous variables used in this model include: the costs of attending the institution (CI), the amount of financial assistance available (FA), the amount of assistance by alumni in the recruiting effort (ALA), the number of students already enrolled in the institution (NS), the effort exerted in recruiting students (RE), and the effort exerted in public relations (PRE). In summary, the model used in this study can be presented as:

\[(\text{DEMAND}) = f(\text{DP}, \text{N/REC}, \text{CCI}, \text{ECB}, \text{CI}, \text{FA}, \text{ALA}, \text{NS}, \text{RE}, \text{and PRE})\]
The justification of these endogenous and exogenous variable comes from both the literature, and from an analysis of the factors felt to be important by several admissions officers consulted by the author.

Measures of Variables:

First, it is argued that the dependent measure of demand for many institutions, especially small, private colleges, is the number of applications received, rather than enrollments per se. This is because the enrollment is often limited in such institutions. To the extent that applications exceed such limits, the institution can be selective in accepting applications. In cases where enrollment limits are not yet achieved, the ratio of accepted applications to enrollments can be established, and projections of enrollments made as a function of applications received and accepted.

Another reason for using applications received as the dependent variable in such forecasting methods is that institutions do have methods of generating applications over which they have control. Knowledge about what indicator activities seem to be most effective in securing applications provides a benefit to the admissions/student recruiting effort in planning its activities. Healey and Brown 1978, p. 419 state that "...once key indicators

1The literature related enrollments to exogenous variables such as N/REC [Spies, 1973; and Stewart and Vest, 1978], CCI [McNally, 1977; and Litten, 1978], and ECB [Lentz, 1977]. The endogenous variables related to enrollments in the literature were CI [Spies, 1973; and Hopkins, 1974], FA [Virginia University, 1977; and Leslie, 1978; and Corwin, 1978], and PRE [Keefe, 1970].
and their lead times are determined, they may be used to predict enrollment changes in an explanatory, rather than projective, manner. The task of the researcher then becomes determining appropriate measures of the independent variables, and the most predictive lead time associated with these measures. Therefore, the dependent demand measure in this study is the number of applications received (APR), which will be determined as a function of the independent, "indicator" variables. More specifically, the measure of the dependent variable was the applications received by an institution for new students for each academic year. Those applications which are received during one academic year from students desiring first-time entrance to the institution in the fall of the following year were used as the measure of the dependent variable in this model. Applications from both readmitting and transferring students were omitted from this analysis.

The exogenous variables in the model include demographic patterns (DP), the state of the national/regional economy (N/REC), the costs of competing institutions (CCI), and the economic benefits of attending college (ECB). Each of these classes of variables can be operationalized for input to the statistical analysis through the use of any number of proxy measures. The collection of these data is most appropriately customized by institution, rather than from macroeconomic and social measures. This requires some refinement of generally available data, such as census statistics, economic data, and other specialized collections of relevant information. To the extent that these data can be determined with
respect to the analyzed institution's geographic and demographic locality, the likelihood that they will explain some of the variance in the dependent variable will be greater. Therefore, it is important to localize the gathering of the data collected for the institution being analyzed.

The major demographic information of interest in this study is that related to the number of highschool graduates. However, to the extent that an institution's entering classes possess certain characteristics (such as predominantly one religion, or from specific geographic regions, or of one sex), the more general number of highschool graduates can be refined to reflect these characteristics.

The measures of the national and regional economy can be any reliable, available data. The more localized the data, the better it is likely to relate to the dependent variable. However, even gross measures can be used, such as a dummy variable which reflects the condition of an economic recession.

Competing institutions can be determined by a particular institution, and the costs of attending that institution used as an indicator. A net difference between the case institution and the competing institution(s) might be a better way of characterizing this indicator.

The final exogenous variable to be operationalized is that concerned with the economic benefits of going to college. One way of deriving a measure for this is by securing the average starting salaries of college graduates. The difference between this and the average starting wage for non-college graduates can become an
appropriate indicator of this variable. Or the ratio of college starting salaries to non-college starting salaries could be used. In any case, some indicator of these economic benefits may well be relevant in this model.

The endogenous variables used in this model include the costs of attending the institution (CI), the amount of financial assistance available (FA), the amount of assistance received from alumni in recruiting new students (ALA), the number of students already enrolled in the institution (NS), the effort exerted in recruiting students (RE), and the effort exerted in public relations on behalf of the college (PRE). Most of these data are available from institutional records. The gathering of these data requires effort on the part of the researcher in assuring their reliability. Some data can be gathered by various offices, such as the admissions office, the registrar, the alumni director's office, or an institutional research office, if available. The important issue in collecting such data is to make sure it is reliable over as many past years as possible. The records of applications received determine how many past years of data on the other measures is desired. If ten years of data on applications received are available, there is no need to gather data on the other variables before that.

Costs of attending the institution include tuition, room and board, books, incidentals, and transportation. The sum of these components can be used as the total cost, but any single component can also be entered into the analysis rather than the total cost measure.
Another variable which is of interest is the amount of financial aid available to incoming students. Components of this variable include such items as grants, scholarships, loans, and student employment. Each of these components may be as important as the total. Only an analysis can determine which is the best measure to use.

An interesting combination of the above two measures can be generated. The net cost of attending the institution might prove to be relevant. This could be determined by dividing the total financial aid available by total enrollment to estimate a per student amount of financial aid, and then subtracting this quotient from the total cost of the institution. This becomes a single measure which combines the input characteristics of two other independent measures. This sort of combination is representative of the judgement and expertise that must be exercised by an analyst when gathering data for a particular problem at hand.

Another variable over which the institution has control is the amount of effort expended by the alumni organization in recruiting new students. Possible measures of this variable include the number of members of the alumni association, the budget for alumni activities, or perhaps that portion of the alumni budget which is devoted to recruiting of new students. Once again, the most relevant measure of this variable must be determined by the analyst.

Relevant measures of student enrollment are rather obvious. The total number of enrollees in each class provides an estimate for this variable. As well, the total number of graduates might be used.

The effort expended in recruiting new students can be measured in several ways. The number of full time equivalent employees in
the admissions office who actually are involved in recruiting is one measure. The total admissions office budget, or the portion devoted to the recruitment effort, can also be used. Other possible measures of this variable include the number of different high schools visited, and/or the number of total visits to high schools, including repeat visits.

Finally, the institution's public relations effort can be measured by the full time equivalent employees in the office, the budget of the office, or some other measure of output from that office, such as number of features presented by the various media, or column inches of published material.

The challenge to the analyst is to secure reliable data for each of the potential variables of the model, and then determine which combinations of the specific measures become the best indicator of the dependent variable. In addition, the lead times associated with each particular measure must be determined.

The following case analyses employ a forward stepping procedure applied to a multiple-regression technique [Hays, 1973, Chapter 16]. The Statistical Package for the Social Sciences (SPSS) was the analysis tool. This is an empirical approach, in which likely combinations of the measures are used to develop models, and subsequent analyses are developed on the basis of results obtained in prior analyses.

The two case studies which follow represent the experience of the author with two institutions that underwent an analysis of the type described. Based on the model presented above, and the collection described, an empirical analysis was undertaken for each
institution separately. Because it was felt that each institution was unique, no attempt was made to aggregate data for the purposes of generalization to other institutions. Results presented are, therefore, only suggested as relevant to the particular case institutions, respectively.

Case Studies:

The first institution studied was a private, Catholic college located in a large metropolitan area. It has a total enrollment of about 4000-5000 students. It is primarily a liberal arts school, but offers several professional training programs, including an M&A. Fourteen years of data were gathered from this institution for analysis.

Results of the analyses led to four models which predicted quite reliably the historical variation in applications received. Table One summarizes these models. The multiple $R^2$ of these models ranged from 0.85 to 0.90. The F statistic for all was significant at the 0.01 level or better. Using each of these models to predict the dependent variable, the absolute value of residuals ranged from 0.82 to 90.67. Stated as percentages of actual applications received, the residuals ranged from less than one tenth of 1% to 10.4%.

Because each of the models used a different set of measures for the variables, the predictions of all four models were aggregated to gain the benefit of all the input measures. Figure One presents a graph of the actual number of applications received versus the aggregated predicted number of applications received. The absolute value of the residuals of these aggregated predictions ranged from 3.44 to 83.2. Stated as a percentage, these residuals ranged from 0.3% to 9.7%, averaging 3.8%.
<table>
<thead>
<tr>
<th>Measure Label</th>
<th>DMY</th>
<th>TOTCST</th>
<th>TFA1</th>
<th>NFRSH</th>
<th>HSG</th>
<th>TGRAD2</th>
<th>TUITION</th>
<th>Constant</th>
<th>R²</th>
<th>F</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-204.76</td>
<td>-0.2422</td>
<td>+0.0152</td>
<td>+0.2462</td>
<td></td>
<td></td>
<td>1515.54</td>
<td>0.85</td>
<td>13.7</td>
<td>13.7</td>
<td>4,9</td>
</tr>
<tr>
<td>2</td>
<td>-142.39</td>
<td>-0.3273</td>
<td>+0.002432</td>
<td>+0.0178</td>
<td>-1.09</td>
<td>1538.25</td>
<td>0.90</td>
<td>14.1</td>
<td>14.1</td>
<td>5,8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-143.35</td>
<td>+0.00167</td>
<td>-1.066</td>
<td>-0.2101</td>
<td>1668.18</td>
<td>0.87</td>
<td>15.6</td>
<td>df=4,9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-147.13</td>
<td>-0.1908</td>
<td>+0.00185</td>
<td>-1.031</td>
<td>1802.27</td>
<td>0.88</td>
<td>16.5</td>
<td>df=4,9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE I**

Beta Estimates, $R^2$ and $F$ for Four Models

**Institution 1**

*The key for these measure labels is as follows: Explanations of the variables appear in the text. DMY = dummy; TOTCST = total cost of attending the institution; TFA1 = the total financial aid lagged one year; NFRSH = number of freshmen; HSG = highschool graduates; TGRAD2 = total number of graduates lagged 2 years. TUITION = annual tuition charged by the institution; Constant = the estimate of the value of $a$ in the multiple regression model.*
Actual vs. Aggregated Projected Applications Received (APR)

Institution 1

FIGURE 1
Among the five variables which appeared in these models, two were exogenous in nature. The dummy variable (DMY) was inserted to reflect the years in which the economy was judged to be in a recession. This measure appeared in each of the four models. Its negative beta estimates are consistent with an expectation that a recessionary economy would lead to fewer applications.

The other exogenous variable to demonstrate importance was the demographic one, reflected in the number of highschool graduates (HSG). Its positive beta estimate is also consistent with expectations, but it only appeared in one model.

The endogenous variables included measures of the costs of attending the institution (CI), the financial aid available (FA), and the institution's enrollment (NS). Costs of attending the institution appeared through two measures, the total cost (TOTCST) and tuition (TUITION). In fact models Three and Four are identical, except that Three uses tuition and Four uses total costs. These measures of institutional cost appeared in each of the four models. The negative beta estimators are consistent with expectations.

Another endogenous variable that appeared in all four models was the financial aid available. This variable was measured by the total financial aid distributed each year, but proved to be a more relevant measure when lagged one year (TFA1). It's positive beta estimator is consistent with expectations.

Finally, the institution's enrollment was measured by the number of freshmen (NFRSH) and by the total number of graduates, which was then lagged two years (TGRAD2). These measures appeared
in all four models. TGRAD2 was in three, while NFRSH was only in one. The positive value of the beta estimate for the NFRSH measure is consistent with expectations. However, the negative estimate for TGRAD2 is unexpected.

The second institution studied was also a private, Catholic college, but it was located in a small city in a rural community, and has only a four-year undergraduate program. It has a total enrollment of less than 500, and is primarily a liberal arts school. There were fifteen years of data available for applications received, but the other data provided resulted in the use of only the most recent eight or ten years of data.

Again, four models were derived from the analysis which predicted the variation in applications received well. Table Two summarizes these models. The multiple $R^2$ of these models were surprisingly high, ranging from 0.97 to 0.99. The F statistic was also significant for each of these models at the 0.001 level or better. Using these models to predict the dependent variable, the absolute value of the residuals ranged from 0.07 to 6.42. Stated as percentages, the residuals ranged from less than 0.1% to 3.4%.

These four models were also aggregated in the same fashion as described for the first institution. Figure Two presents a graph of the actual number of applications received versus the aggregated number of applications received. The absolute value of the residuals of these aggregated predictions ranged from less than 0.07 to 3.00. Stated as a percentage, the aggregated residuals ranged from less than 0.1% to 1.4%.
<table>
<thead>
<tr>
<th>Measure label</th>
<th>TOTCST</th>
<th>HSG</th>
<th>DMY</th>
<th>TFA1</th>
<th>TGRAD²</th>
<th>NFRSH</th>
<th>AMEM</th>
<th>ECBEN</th>
<th>TGNTS</th>
<th>Constant</th>
<th>F</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model No. 1</td>
<td>0.0361</td>
<td>0.0016</td>
<td>1.485</td>
<td>0.0149</td>
<td>1.435</td>
<td>-1.728</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>17.93</td>
<td>57.6</td>
<td>0.95</td>
</tr>
<tr>
<td>Model No. 2</td>
<td>-0.0618</td>
<td>-0.0024</td>
<td>-2.730</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0.2212</td>
<td>X</td>
<td>X</td>
<td>304.9</td>
<td>46.4</td>
<td>0.98</td>
</tr>
<tr>
<td>Model No. 3</td>
<td>-0.0082</td>
<td>0.0066</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1.336</td>
<td>X</td>
<td>-462.0</td>
<td>46.3</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Model No. 4</td>
<td>0.0146</td>
<td>0.0053</td>
<td>-0.208</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0.0095</td>
<td>X</td>
<td>-341.5</td>
<td>43.8</td>
<td>0.98</td>
</tr>
</tbody>
</table>

**TABLE II**

Beta Estimates, $R^2$, and $F$ for Four Models

*The key for these measure labels is as follows: TOTCST = total costs of attending; HSG = highschool graduates; DMY = dummy; TFA1 = total financial aid lagged one year; TGRAD² = total number of graduates lagged two years; NFRSH = number of freshmen; AMEN = alumni association membership; ECBEN = economic benefit of a college education; TGNTS = total grants; Constant = the estimate of the alpha value in the regression model.*
Actual Vs. Aggregated Projected Applications Received (APR)
Among the seven variables which appeared in the models for the second institution, three were exogenous in nature. These included demographic patterns, economic conditions, and the economic benefits of attending college. Again, the dummy variable (DMY) was inserted to reflect the years in which the economy was judged to be in a recession. Two of the three models in which DMY appeared had negative beta estimators which is consistent with expectations.

The demographic pattern variable was again measured by the number of highschool graduates (HSG). The beta estimates for three of the four models in which this measure appeared were positive, which was again consistent with expectations.

The last exogenous variable which appeared in this analysis was the economic benefit of attending college, measured by combining the starting salaries of college graduates with the starting salaries of non-college graduates by subtracting the latter from the first. This measure is labelled ECBEN. Only one of the models found this to be a significant measure. Its positive beta estimate is consistent with expectations.

Four endogenous variables appeared in the models. The first was costs of attending the institution as measured by the total cost (TOTCST). This measure appeared in all four models, but two of the beta estimators are positive, while two are negative.

Again, the financial aid available appeared, but this time only in two of the analyses. It was measured in one model by total financial aid, again lagged by one year (TFAI), and in the other model by the total amount of grants available (TGRNTS). Both beta estimators are positive, as would be expected.
Institutional enrollment appeared again, measured by both the number of freshmen (NFRSH) and the total number of graduates lagged two years (TGRAD2). These both appear in the same model. The positive beta estimate for TGRAD2 is as would be expected, but the negative estimate for the beta of NFRSH is not to be expected.

The final endogenous variable which appeared in the second institution's analysis was the alumni effort as measured by the number of members in the alumni association (AMEM). This measure appeared in one model, and its positive beta estimator is consistent with expectations.

These two case analyses, each with four models, provide the analyst with some insight to the generation of applications. They also give the planner a tool to use in projecting future applications. Both of these functions must be done with care and caution. The next section discusses these issues.

Discussion:
To the extent that the values of the independent variables for future periods can be determined, the expected number of applications received can be projected. Because several of the variables which have appeared in these analyses are lagged one or two years, their values for future estimates are known with certainty. Estimates of other variables can often be secured from outside sources, which provide objective estimates, usually with defined levels of certainty. The projections made from these estimates can then be associated with certain levels of certainty, and the decisionmaker can operate within a defined confidence level.
In addition, because the independent variables are assumed to be determinant of the dependent variable, it is possible for those responsible for securing applications to emphasize those activities which seem to most important in such analyses. By focusing on such activities, it is clear that the efforts of the admissions people can be more efficiently planned and implemented. As these analyses continue annually, the effect of new recruiting programs can be judged to determine their effectiveness. This approach to projecting dependent variables of this nature provides this explanatory benefit which others do not. It is this benefit which becomes most advantageous to the indicator method of forecasting.

With regard to the results of the two case analyses presented above, several questions are raised. First, the measure TGRAD2, which was presented as a measure of the institutions enrollment, shows a negative beta estimator for the first institution, which is contrary to expectations. If the institution is in fact providing a quality education, you would expect measures of enrollment to be positively correlated with the number of applications received. Since this is not the situation for the first case institution, other phenomena must exist. Perhaps the students at the college are not pleased with the education they are receiving, and therefore are discouraging potential applicants. This argument also applies to the negative estimator for NFRSH in the second case institution. Or perhaps the measure TGRAD2 is more appropriately measuring some variable other than enrollment. To the extent that that variable is negatively related to applications received, then that is an explanation for the negative estimators.
The confusion surrounding the sign of the beta estimators of the TOTCST measure in the second analysis must also be addressed. It is possible that the more relevant variable for this institution is relative costs compared to other institutions, rather than the absolute cost of attending this particular institution. If that is true, then it is appropriate for the analysis to be continued to determine if either differences, or ratios between the case institution's costs and those of the potential competing institutions appear as relevant indicators. This has not yet been done in this analysis.

The high $P^2$ obtained for both institutions indicates that this approach to forecasting demand for institutions of higher education might have some applicability. Because of the low number of total cases for each institution, it was not possible to cross validate any of these models. Also, because only two institutions were studied, there is not enough data to generalize beyond the particular institutions themselves. As a technique for forecasting demand at a specific institution, the process described in this paper seems to be promising. However, as a means of defining those general factors which seem to be most determinant of applications received, the approach must await further analyses.

It is also possible for factors which are nonquantifiable to have an effect on demand. Variation in demand would then occur, but not be explained by the quantitative model. Examples of such factors could include the end of military conscription, turnover of people who are responsible for recruiting new students, other changes in academic and administrative personnel, and the change
of, or addition to, the academic program. While many of these factors can be understood, they cannot be incorporated into the quantitative models. These factors must be judgementally integrated with the analysis to obtain appropriate results.

Finally, a unique aspect of this approach was the aggregating process, in which the four different models for each institution were integrated into a single prediction for each year. It is felt that doing so provides a means for integrating a number of statistically significant variables into an analysis which otherwise would violate some statistical assumptions. Results can then be used to project future demand on the basis of as much information as possible.

Summary and Conclusions:

This study presents a method of projecting an institution's demand for higher education programs. Demand was measured by the dependent variable of number of applications received by the institution. A model of independent variables which determine the demand was presented, including both: exogenous variables - such as demographic patterns, national and regional economic conditions, the cost of competing institutions, and the economic benefit of a college education, and endogenous variables; such as costs of the institution, financial aid available, alumni assistance, students enrolled, recruiting effort, and public relations effort.
Two case studies of institutions which were analyzed by this model were then presented. A regression analysis was applied, and those measures of the variables which explained the most variation in the dependent variable were determined. The resulting models had \( R^2 \) for one institution in the range of 85% - 90%. For the other case institution, the \( R^2 \) range was 97% - 99%. It was concluded that this approach was one with much potential for specific institutions, but that results obtained were not yet generalizable. It is suggested that further research with several other institutions be undertaken to determine those factors which seem to be most determinant of the demand for college education.
Bibliography


McNally, James L., and Richard W. Story, "Why Accepted Students Decline to Attend the College at Geneseos," Paper Presented at the Annual Meeting of the Association for Institutional Research, State University of New York, Geneseos, September, 1977.


---

