DOCUMENT RESUME

ED 046 065 EA 003 184

AUTHOR Banks, Jerry; Hohenstein, C. L.
TITLE Conceptual Models and Procedures for Predicting Higher Education Enrollment.
PUB DATE Oct 70
NOTE 34p.; Paper presented at Operations Research Society of America National Meeting (38th, Detroit, Michigan, October 28-30, 1970)

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Data Analysis, Enrollment, Enrollment Influences, *Enrollment Projections, *Enrollment Trends, *Mathematical Models, *Multiple Regression Analysis, Tables (Data)

ABSTRACT

Three mathematical models and their enrollment projections for higher education in Georgia are presented. The first model is an extrapolation of current trends. The second model is a simple linear regression model based on steady growth. The third is a multiple regression model attempting to account for the influences of several variables simultaneously. Projected enrollments for models I and II are similar to each other, but projections using model III are markedly higher. (Figure IV on page 33 may reproduce poorly in hard copy because of marginal legibility.) (RA)

Conceptual Models and Procedures for Predicting Higher Education Enrollment*

Jerry Banks
Georgia Institute of Technology

and

C. L. Hohenstein
C. L. Hohenstein and Associates

1.0 General.

Section Contents

1.1 Introduction. This paper represents a portion of the efforts instituted by the Georgia Higher Education Facilities Commission (GHEFC) to determine future higher education needs in the State of Georgia. The overall program originated from a 1966 request of the U. S. Congress for a careful study within each state to determine the amount and kind of buildings needed, the indicated cost, and probable sources of funds for ten years in the future.

This paper is designed to present several of the basic models that have been used by the authors to examine enrollment trends and to provide projections of future higher education enrollment

*Presented at the 38th National ORSA Meeting, October 28-30, 1970.

in Georgia. The ultimate hope is that these projections can be related to physical facilities needs, which will then be translated into a cost projection necessary to support that portion of the higher education system.

The authors realize that public education policy is a controlling factor in the allocation of resources to public and private institutions. This factor and others are illustrated in the General Model of Georgia Higher Education Enrollment.

1.2 General Index of Symbols and Abbreviations Used.

Symbol	Description
GHEFC	The Georgia Higher Education Facilities Commission
Y	Total State Enrollment in Higher Education
Y_i	Total State Enrollment in Higher Education for year i
i	A specified year
O_{FFE}	Out-of-State First-Time Freshmen Enrollments
O_{NFT}	Out-of-State Non-Freshman Transfers
G_{FFE}	First-Time Freshmen Enrollments by Georgians
R_{UO}	Returning Undergraduates and Others
T_{FFE}	Total First-Time Freshmen Enrollments
G_{HSG}	Georgia High School Graduates
P_1	Ratio of Total First-Time Freshmen Enrollment (T_{FFE}) divided by the Total State Enrollment in Higher Education Institutions (Y)
P_2	Ratio of Opening Fall Freshmen Enrollment (O_{FFE}) divided by the Total Fall Freshmen Enrollment (T_{FFE})
P_3	Ratio of First-Time Enrollment by Georgians divided by Georgia High School Graduates
X	Symbol used in the regression equation to denote a current year minus 1960
ρ	A coefficient of correlation of a regression line

2.0 Model I - Trends in College Going.

Section Contents

2.1 Definition and Assumption. In this section, a historical enrollment model is developed and applied to data from the State of Georgia. This model is based on the assumption of a continuation of the previous general pattern of state and national policy.

2.2 Total First-Time Freshmen Enrollment. An attempt is made to predict Y for every year from 1970 through 1980. The value can be determined from the following equation:

$$Y = O_{FFE} + O_{NFT} + G_{FFE} + R_{UO} \qquad (1)$$

A further distinction is made that:

$$T_{FFE} = G_{FFE} + O_{FFE} \qquad (2)$$

An examination of the past history of both Federal and State enrollment levels indicates that there is a constant relationship (P_1) between T_{FFE} and Y. That is:

$$T_{FFE} = P_1 \cdot Y \qquad (3)$$

and

$$Y = T_{FFE}/P_1 \qquad (4)$$

Evidence of the rather constant nature of P_1 for the State of Georgia and for the United States is shown in *Figure 1, "Tracking Characteristics of P_1 Values for Georgia." It can be observed that there is a close correspondence between historical values for Georgia and the United States. Hence, the values projected for the United States are used as projections for the State of Georgia. The investigators assume the value 0.220 for 1978 through 1980.

2.3 Out of State First Time Freshmen Enrollments. Rather than predict the actual enrollment created by the O_{FFE}, the proportion of O_{FFE} to T_{FFE} is determined, and the proportion is denoted by P_2. This value has varied only slightly over the historical data period. The value ranges from a low of 0.0646 to a high of 0.0784. The average value of P_2 is 0.0719 and this value is assumed as a constant for the proportion.

As discussed above,

$$O_{FFE} = P_2 \cdot T_{FFE} \quad (5)$$

and using the computed value

$$O_{FFE} = 0.0719\ T_{FFE} \quad (6)$$

This value is substituted into equation (2) to get:

$$G_{FFE} + 0.0719\ T_{FFE} = T_{FFE}$$

$$G_{FFE} = T_{FFE} - 0.0719\ T_{FFE}$$

$$G_{FFE} = 0.9281\ T_{FFE}$$

and finally,

$$T_{FFE} = 1.078\ G_{FFE} \quad (7)$$

---

*Values in Figure 1 are derived from "Projections of Educational Statistics," Office of Education, U.S. Department of Health, Education, and Welfare and from "Opening Fall Enrollment in Higher Education," Part A, Summary Data, National Center for Educational Statistics, United States Office of Education.

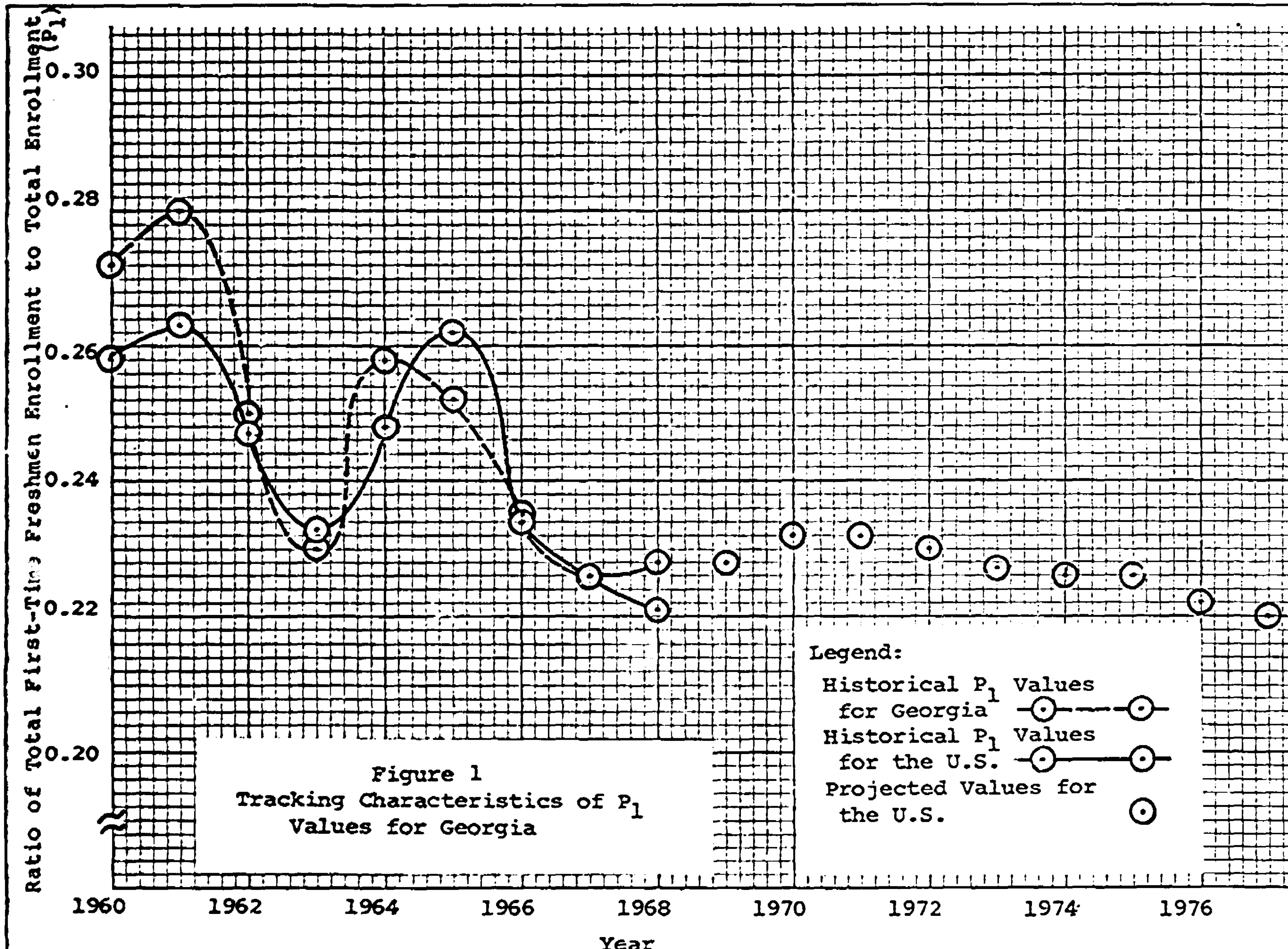

Figure 1
Tracking Characteristics of P_1 Values for Georgia

Equation (7) states the assumption that if G_{FFE} can be predicted for any year, then T_{FFE} can also be predicted for that year without predicting O_{FFE}. This assumption is used in the following procedures.

2.4 <u>First-Time Freshmen Enrollment by Georgians</u>. After computing T_{FFE} as in Equation (7), then Y can be computed from Equation (4). That is,

$$\hat{Y}_i = 1.078\ \hat{G}_{FFE}\ (n)/\hat{P}_1\ (n) \qquad (8)$$

where

i = year (1970, 1971, ..., 1980)

It is hypothesized that a direct relationship exists between G_{HSG} and G_{FFE}. This relationship is denoted as P_3. That is

$$P_3 = G_{FFE}/G_{HSG}$$

and

$$G_{FFE} = P_3 \cdot G_{HSG} \qquad (9)$$

Substituting Equation (9) into Equation (8) gives

$$\hat{Y}_i = 1.078\ \hat{G}_{HSG(i)}\ (\hat{P}_{3(i)}/\hat{P}_{1(i)}) \qquad (10)$$

where

i = year (1970, ..., 1980)

Historical values of P_3 for Georgia* are shown below:

Year	1960	1961	1962	1963	1964	1965	1966	1967	1968
P_3	0.400	0.408	0.389	0.396	0.418	0.401	0.410	0.418	0.466

*These values were derived using data from the "Digest of Educational Statistics," Office of Education, Department of Health, Education, and Welfare and from "Opening Fall Enrollments," Office of Education, Department of Health, Education, and Welfare.

Using the historical data, the simple linear regression equation is

$$\hat{P}_3 = 0.389 + 0.0057X \qquad (11)$$

where

$$X = (\text{current year} - 1960)$$

with

$$\rho = 0.694$$

The variance of $\hat{b}$ ($S_{\hat{b}}^2$) is 0.00005 which yields a t-value of 114. This is compared with the tabulated $t_{(7,.95)}$ of 1.895 from which we conclude that the relationship is significant.

Equation (10) can now be used to predict Total Enrollment in Higher Education, 1970 through 1980 as shown in Table 1.

Table 1

TOTAL ENROLLMENT IN HIGHER EDUCATION, 1970 THROUGH 1980

Year	($\hat{G}_{HSG}$)	1.078 X Column 2	($\hat{P}_3$)	Column 3 X Column 4	($\hat{P}_1$)	$\hat{Y}$
-1-	-2-	-3-	-4-	-5-	-6-	-7-
1970	55,799	60,000	0.446	26,800	0.232	115,500
1971	60,664	65,300	0.452	29,500	0.232	127,100
1972	63,823	68,600	0.457	31,400	0.230	136,600
1973	66,270	71,400	0.463	33,000	0.227	145,500
1974	69,269	74,600	0.469	35,000	0.226	155,000
1975	71,627	77,100	0.474	36,500	0.226	161,500
1976	73,571	79,200	0.480	38,000	0.222	171,000
1977	76,102	81,900	0.486	39,800	0.220	181,000
1978	78,412	84,400	0.492	41,500	0.220	188,500
1979	80,435	86,500	0.497	43,000	0.220	193,300
1980	82,698	89,100	0.503	44,800	0.220	203,500

Source: Column 2: "Projection of Georgia Public High School Graduates", October 1, 1969, Document Reference No. 10, Prepared for the Georgia Higher Education Facilities Commission by the Educational Research and Planning Group, Atlanta, Georgia; All other data derived.

3.0 Data for Sections 4.0 and 5.0

The input data for Sections 4.0 and 5.0 is summarized in Table 2.

The column headings are explained as follows:

X_1 - Deflated Family Income

X_2 - Consumer Price Index

X_3 - Total High School Graduates

X_4 - First-Time Freshmen Enrollment, Previous Year

X_5 - Current Year

Y - Total Higher Education Enrollment

Correlation between the input variables appeared to be extremely high as indicated in Table 3 below. High values of correlation, i.e., those close to the absolute value of one, indicate a high degree of dependence between the observed variables.

Table 3

Correlation Matrix
(Input Variables)

Y	1.000	.988	.981	.963	.974	.985
X_1	.988	1.000	.954	.960	.953	.974
X_2	.981	.954	1.000	.897	.954	.963
X_3	.963	.960	.897	1.000	.939	.957
X_4	.974	.953	.954	.939	1.000	.947
X_5	.985	.974	.963	.957	.947	1.000

This high correlation directed the investigators to build the simple linear regression model presented in the nexc section.

Table 2

INPUT DATA FOR REGRESSION ANALYSIS

X_5	X_1*	X_2**	X_3***	X_4****	Y*****
1960	5,310	103.1	35,527	12,461	50,220
1961	5,090	104.2	37,700	13,631	51,955
1962	5,300	105.4	37,614	14,579	56,228
1963	5,680	106.7	39,959	14,016	62,236
1964	5,950	108.1	45,093	14,347	69,527
1965	6,270	109.9	53,548	17,903	82,347
1966	6,690	113.1	53,652	20,711	91,280
1967	6,870	116.3	55,140	21,318	98,476
1968	7,050	121.2	55,470	22,259	108,816
1969	7,170	127.7			
1970		135.2			

Source: *Sales Management, the Marketing Magazine, Survey of Buying Power, June, 1961-1970.

**Consumer Price Index for Urban Wage Earners and Clerical Workers, U. S. Department of Labor, Bureau of Labor Statistics, Washington, D.C.

***Digest of Educational Statistics, Office of Education, Department of Health, Education, and Welfare.

****Opening Fall Enrollment in Higher Education; Part A, Summary Data, National Center for Educational Statistics, United States Office of Education.

*****Opening Fall Enrollment in Higher Education; Part A, Summary Data, National Center for Educational Statistics, United States Office of Education.

4.0 Model II - Simple Linear Regression

Section Contents

4.1 Mathematical Statement of the Model. Consider paired observations on variables X_5 and Y, where X_5 is the year and Y is Total Higher Education Enrollment. It is hypothesized that the equation

$$Y = a + bX_5$$

describes the relationship between the two variables. Using the data described in Table 2 and the least squares technique it is possible to obtain estimates $\hat{a}$ and $\hat{b}$ to determine estimates Y as shown below:

$$\hat{Y}_i = \hat{a} + \hat{b}X_{5_i}$$

For east of computation, the data were coded as shown below:

year	1960	1961	1962	1963	1964	1965	1966	1967	1968
X_5	-4	-3	-2	-1	0	1	2	3	4

4.2 Analysis of the Output.

The estimated values of $\hat{a}$ and $\hat{b}$ are indicated in the computed regression equation as:

$$\hat{Y}_i = 74565 + 7736X_{5_i}$$

The general form of the analysis of variance (ANOVA) table for linear regression is shown in Table 4 below.

Table 4

ANOVA Table for Linear Regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	n - 1	$SST = (Y_i - \bar{Y})^2$	$MST = \frac{SST}{n-1}$
Regression	1	$SSR = (\hat{Y}_i - \bar{Y})^2$	$MSR = SSR$
Error	n - 2	$SSE = (Y_i - \hat{Y}_i)^2$	$MSE = \frac{SSE}{n-2}$

In Table 4, n indicates the number of observations. The ANOVA table for the regression under study is shown in Table 5.

Table 5

ANOVA Table for Computed Regression Equation

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Total	8	3.702×10^9	0.463×10^9
Regression	1	3.591×10^9	3.591×10^9
Error	7	0.111×10^9	1.594×10^7

The Index of Determination is that proportion of the sum of squares which is explained and given by

$$\rho^2 = \frac{SSR}{SST}$$

A related quantity is the Simple Correlation Coefficient which is defined by

$$\rho = SGN(\hat{b})\sqrt{\frac{SSR}{SST}}$$

where

$$SGN(\hat{b}) = -1 \quad \text{if } \hat{b} < 0$$
$$SGN(\hat{b}) = 0 \quad \text{if } \hat{b} = 0$$
$$SGN(\hat{b}) = 1 \quad \text{if } \hat{b} > 0$$

The case at hand yields

$$\rho^2 = 0.970$$
$$\rho = 0.985$$

which indicates that a very good fit exists. To further ascertain the validity of the computed regression equation the F-test is applied. The test statistic is given by

$$F = \frac{MSR}{MSE}$$

with 1 degree of freedom in the numerator and n - 2 degress of freedom in the denominator. The test statistic is computed as 225.294. The tabulated statistic $F_{(.001,1,7)} = 29.2$ indicating that $b \neq 0$.

4.3 Enrollment Projection Using the Computed Regression Equation. Table 6 is a projection of enrollment for the next decade, as well as the lower and upper value of the 95% prediction limit in each case. The lower prediction limit is given by

$$Y_L = \hat{Y} - S_{\hat{Y}} t_{k,p}$$

and the upper prediction limit by

$$Y_U = \hat{Y} + S_{\hat{Y}} t_{k,p}$$

where

$$S_{\hat{Y}}^2 = MSE \left[1 + \frac{1}{n} + \frac{(X - \overline{X})^2}{\Sigma(X_i - X)^2} \right]$$

and $t_{k,p}$ denotes the 100P percentile point of a t-distribution with K degrees of freedom.

4.4 Comparison with Previous Projections. The projection indicated herein is compared with that of Model I as shown in Table 7 below.

The largest percentage difference is in 1970 where the absolute difference in predictions is 5,481 students and the percentage difference is approximately 4.5%. The smallest absolute difference is in 1972. In every case, the projection indicated by Model I is well within the 95% prediction limits of Model II.

Table 6

ENROLLMENT PROJECTION AND PREDICTION LIMITS
USING SIMPLE LINEAR REGRESSION

Year	Y	Y_L	Y_U
1970	120,981	108,608	133,354
1971	128,717	115,584	141,850
1972	136,453	122,495	150,412
1973	144,189	129,350	159,028
1974	151,925	136,161	167,690
1975	159,661	142,933	176,390
1976	167,397	149,673	185,121
1977	175,133	156,387	193,880
1978	182,869	163,079	202,660
1979	190,605	169,751	211,460
1980	198,342	176,406	220,277

Table 7

COMPARISON OF PROJECTION FROM SIMPLE LINEAR REGRESSION MODEL AND THAT OF MODEL I

Year	Simple Linear Regression Model	Model I
1970	120,981	115,500
1971	182,717	127,100
1972	136,453	136,600
1973	144,189	145,400
1974	151,925	155,000
1975	159,661	161,500
1976	167,397	171,000
1977	175,133	181,000
1978	182,869	188,000
1979	190,605	193,000
1980	198,342	203,500

5.0 Linear Multiple Regression Model.

Section Contents

5.1 Mathematical Statement of the Model. The linear multiple equation may be written

$$Y = B_o + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5$$

where

Y = the dependent variable
$X_1, \ldots\ldots X_5$ = the independent variables
$B_o, B_1, \ldots\ldots B_5$ = the unknown regression coefficients

a least squares technique is used to calculate estimates $\hat{B}_o, \hat{B}_1, \ldots\ldots \hat{B}_5$ such that estimates of the dependent variable Y_i can be obtained as:

$$Y_i = \hat{B}_o + \hat{B}_1X_{1_i} + \ldots\ldots + \hat{B}_5X_{5_i}$$

The justification for attempting this model building exercise after the apparent success of the Simple Linear Regression Model is the determination of the need for this class of model, an observation of the most significant and least significant variables, and a comparison of the projections made.

5.2 Analysis of the Computed Regression Equation. A regression equation was constructed with the five independent and one dependent variable indicated in Table 2. A portion of the information concerning the developed equation is given in Table 8 below:

Table 8

REGRESSION EQUATION FOR ALL FIVE INDEPENDENT VARIABLES

Variable	Coefficient	Standard Error	t-Statistic
0	-183,338		
1	6.136	1.805	3.400
2	1658.6	227.4	7.295
3	0.713	0.149	4.792
4	0.356	0.287	1.241
5	4.498	560.5	0.008

In addition $\rho^2 = 0.9996$, thus $\rho \doteq 1$ and the F-Ratio Statistic was computed as 1556.2.

5.3 Selection of a Regression Equation. The tabulated value of $t_{(.95,3)}$ is given as 2.35. Hence, Variables 4 and 5 are not significantly different from zero at the 95% confidence level. On this basis, Variables 4 and 5 were eliminated from the regression and a new equation was developed with Independent Variables 1, 2, and 3 and the Dependent Variable. The coefficients and computed t-statistic values are given in Table 9 below.

Table 9

REGRESSION EQUATION FOR THREE INDEPENDENT VARIABLES

Variable	Coefficient	Standard Error	t-Statistic
0	-194,867		
1	6.084	1.718	3.513
2	1793.3	133.9	13.393
3	0.787	0.103	7.638

The tabulated value of $t_{(.95,5)}$ is given as 2.015. Hence, the coefficients $\hat{B}_1$, $\hat{B}_2$, and $\hat{B}_3$ are significantly different from zero at the 95% Confidence Level.

The covariance matrix on the estimated coefficient is given in Table 10 below:

Table 10

COVARIANCE MATRIX (ESTIMATED COEFFICIENTS)

	B_1	B_2	B_3
B_1	2.950	-172.6	-0.140
B_2	-172.6	17928.8	3.190
B_3	-0.140	3.190	0.0106

These coefficients were used to determine the level of correlation between the independent variables. In general, the correlation we seek is given by

$$\rho_{\hat{B}_i \hat{B}_j} = \frac{C_{\hat{B}_i \hat{B}_j}}{S_{\hat{B}_i} S_{\hat{B}_j}}$$

where

$$S_{\hat{B}_1} = \sqrt{2.950} = 1.718$$

$$S_{\hat{B}_2} = \sqrt{17928.8} = 133.9$$

$$S_{\hat{B}_3} = \sqrt{0.0106} = .103$$

then

$$\rho_{\hat{B}_1, \hat{B}_2} = -0.753; \quad \rho_{\hat{B}_1, \hat{B}_3} = -0.462; \quad \rho_{\hat{B}_2, \hat{B}_3} = 0.232$$

The absolute value of each of these correlations is sufficiently less than one. Hence, it is important to include all of the independent variables in the equation.

The Index of Determination is

$$\rho^2 = 0.9994$$

Hence, the Simple Correlation Coefficient is

$$\rho \doteq 1$$

Finally, the computed F-ratio test statistic is 2777.3. The tabulated value of $F_{(0.0001,3,5)}$ is 86.292. Therefore, we may safely conclude that the developed relationship is significant.

5.4 Estimate of Deflated Family Income. The independent variables used in the selected equation are Deflated Family Income, Consumer Price Index, and High School Graduates. Predictions are being made for the future, based on future values of these variables. Hence, estimates must be made of the future value of the independent variables. In this section the procedure for predicting deflated family income will be made. The data in Table 2 were fitted to three curves and ρ^2 for each was determined.

$$\text{For } Y = a + bX;\ \rho^2 = 0.95478$$
$$Y = ae^{bX};\ \rho^2 = 0.950566$$
$$Y = 1/(a + bX);\ \rho^2 = 0.943181$$

where Y is deflated family income and X is the current year - 1960. Since ρ^2 is the largest for the simple linear regression, that curve is selected to predict deflated family income.

The actual value of deflated family income in 1969 is $7,170 and the estimated value is $7,298 for that same year. With this added information, the estimated values for 1970 through 1980, as well as the prediction limits, were adjusted by a reduction of $128 as indicated in Table 11 below.

5.5 Estimate of Consumer Price Index. It was the original intent of the investigators to use some index of tuition rather than Consumer Price Index in this position. However, this data was not currently available, and the Consumer Price Index was selected as an alternate. Prior to the regression analysis, it was philosophized that the sign of the coefficient for Consumer Price Index would be negative. That is, it was believed that the higher costs of living would serve as a restraining force on college going. However, this subjective reasoning was in opposition to the analytical result. In fact, the computed t-statistic has the largest value of the three discussed previously. The explanations for this inconsistency are varied. The investigators

Table 11

ESTIMATES OF DEFLATED FAMILY INCOME
1970 THROUGH 1980

Year	Unadjusted			Adjusted		
	$\hat{X}_1$*	X_{1L}	X_{1U}	$\hat{X}_1$	X_{1L}	X_{1U}
1970	7,553	7,054	8,053	7,425	6,926	7,925
1971	7,809	7,285	8,333	7,681	7,157	8,205
1972	8,064	7,514	8,615	7,936	7,386	8,487
1973	8,320	7,740	8,900	8,192	7,610	8,772
1974	8,575	7,964	9,187	8,447	7,836	9,054
1975	8,831	8,187	9,475	8,703	8,059	9,347
1976	9,086	8,408	9,765	8,958	8,280	9,637
1977	9,342	8,628	10,056	9,214	8.500	9,928
1978	9,597	8,847	10,348	9,469	8,719	10,220
1979	9,853	9,065	10,641	9,725	8,937	10,513
1980	10,108	9,282	10,935	9,980	9,154	10,807

*All Values Rounded

have considered the following as a possible explanation. When the Consumer Price Index (which is a measure of the costs of goods and services for urban wage earners and clerical workers) rises, children of rural wage earners and non-clerical workers are encouraged to obtain a higher education so that they can receive a higher income throughout life, thereby coping with the rise in the Consumer Price Index.

Numerous methods were used to describe the Consumer Price Index.

For $Y = a + bX$; $\rho^2 = 0.890$

$Y = ae^{bX}$; $\rho^2 = 0.909$

$Y = 1/(a+bX)$; $\rho^2 = 0.939$

Least squares prediction with trend based on previous three years data; $\rho^2 = 0.985$

The last mentioned technique yields the highest value of the Index of Determination. It will be used to predict the Consumer Price Index as shown in Table 12 below.

Table 12

ESTIMATES OF CONSUMER PRICE INDEX, 1970 THROUGH 1980

Year	$\hat{X}_2$	X_{2L}	X_{2U}
1970	135.2[a]	135.2	135.2
1971	142.0	137.56	146.44
1972	149.0	144.25	153.75
1973	156.0	150.90	161.10
1974	163.0	157.51	168.49
1975	170.0	164.08	175.92
1976	177.0	170.67	183.33
1977	184.0	177.24	190.76
1978	191.0	183.77	198.23
1979	198.0	190.27	205.73
1980	205.0	196.84	213.16

a - actual

5.6 Estimate of High School Graduates. A prediction of High School Graduates is given in Table 12 of "Projection of High School Graduates," October 1, 1969, Document Reference 10, prepared for the Georgia Higher Education Facilities Commission by the Educational Research and Planning Group. In the referenced document three projections are given and called low, expected, and high. With slight modification, these values will be used as X_{3L}, $\hat{X}_3$, and X_{3U} as shown in Table 13 below.

Table 13

ESTIMATE OF HIGH SCHOOL GRADUATES
1970 THROUGH 1980

Year	$\hat{X}_3$	X_{3L}	X_{3U}
1970	58,799	57,179	58,799
1971	60,664	59,393	60,664
1972	63,823	61,607	63,823
1973	66,270	63,821	66,468
1974	69,269	66,035	70,898
1975	71,627	68,249	75,109
1976	73,571	70,463	76,456
1977	76,102	72,677	79,813
1978	78,412	74,891	83,007
1979	80,435	77,105	85,770
1980	82,698	79,319	88,999

5.7 Enrollment Projection Using the Computed Regression Equation. Using the regression equation

$$\hat{Y}_1 = -194867 + 6.034\hat{X}_1 + 1793.3\hat{X}_2 + 0.787\hat{X}_3$$

with the estimates of Deflated Family Income, Consumer Price Index, and High School Graduates discussed in the previous sections, Table 14 is developed.

Table 14

ENROLLMENT PROJECTION USING
LINEAR MULTIPLE REGRESSION

Year	$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{Y}$	Y_L*	Y_U*
1970	7,425	135.2	58,799	138,733	121,245	156,221
1971	7,681	142.0	60,664	154,648	136,483	172,813
1972	7,936	149.0	63,823	170,563	149,892	191,234
1973	8,192	156.0	66,270	186,478	163,664	209,292
1974	8,447	163.0	69,269	202,393	176,236	228,550
1975	8,703	170.0	71,627	218,308	189,661	246,955
1976	8,958	177.0	73,571	234,223	203,687	264,759
1977	9,214	184.0	76,102	250,138	216,951	283,325
1978	9,469	191.0	78,412	266,053	230,215	301,891
1979	9,725	198.0	80,435	281,968	243,479	320,457
1980	9,980	205.0	82,698	297,883	256,743	339,023

*Based on accepting $\hat{X}_1$ values as true values.

It is recognized that the Consumer Price Index is a major contributor to the estimate of Total Enrollment. Figure II is offered as an alternative measure of Total Enrollment for varying values of the annual increase in the Consumer Price Index. For example, a predicted annual 5 points/year rise would result in a Predicted Total Enrollment of 262,383 in 1980.

As mentioned previously, the values of $\hat{X}_1$, $\hat{X}_2$, and $\hat{X}_3$ in Table 13 are estimates so that the prediction limits are partial rather than complete. To obtain a complete prediction limit the following steps are taken.

(1) The values $\hat{B}_i X_{iL}$ and $\hat{B}_i X_{iU}$ are computed for $i = 1, 2, 3$.

(2) The prediction interval $\hat{B}_i(X_{iU} - X_{iL})$ is computed for $i = 1, 2$.

(3) The half-interval width $\frac{1}{2}\hat{B}_i(X_{iU} - X_{iL})$ is determined for $i = 1, 2$.

(4) The lower differential interval $\hat{B}_3(\hat{X}_3 - X_{3L})$ is computed.

(5) The upper differential interval $\hat{B}_3(X_{3U} - \hat{X}_3)$ is computed.

(6) The sum $\frac{1}{2}\Sigma\hat{B}_i(X_{iU} - X_{iL}) + \hat{B}_3(\hat{X}_3 - X_{3L})$ is subtracted from Y_L in Table 13.

(7) The sum $\frac{1}{2}\Sigma\hat{B}_i(X_{iU} - X_{iL}) + \hat{B}_3(X_{3U} - \hat{X}_3)$ is added to Y_U in Table 13.

The results of these operations are shown in Table 15.

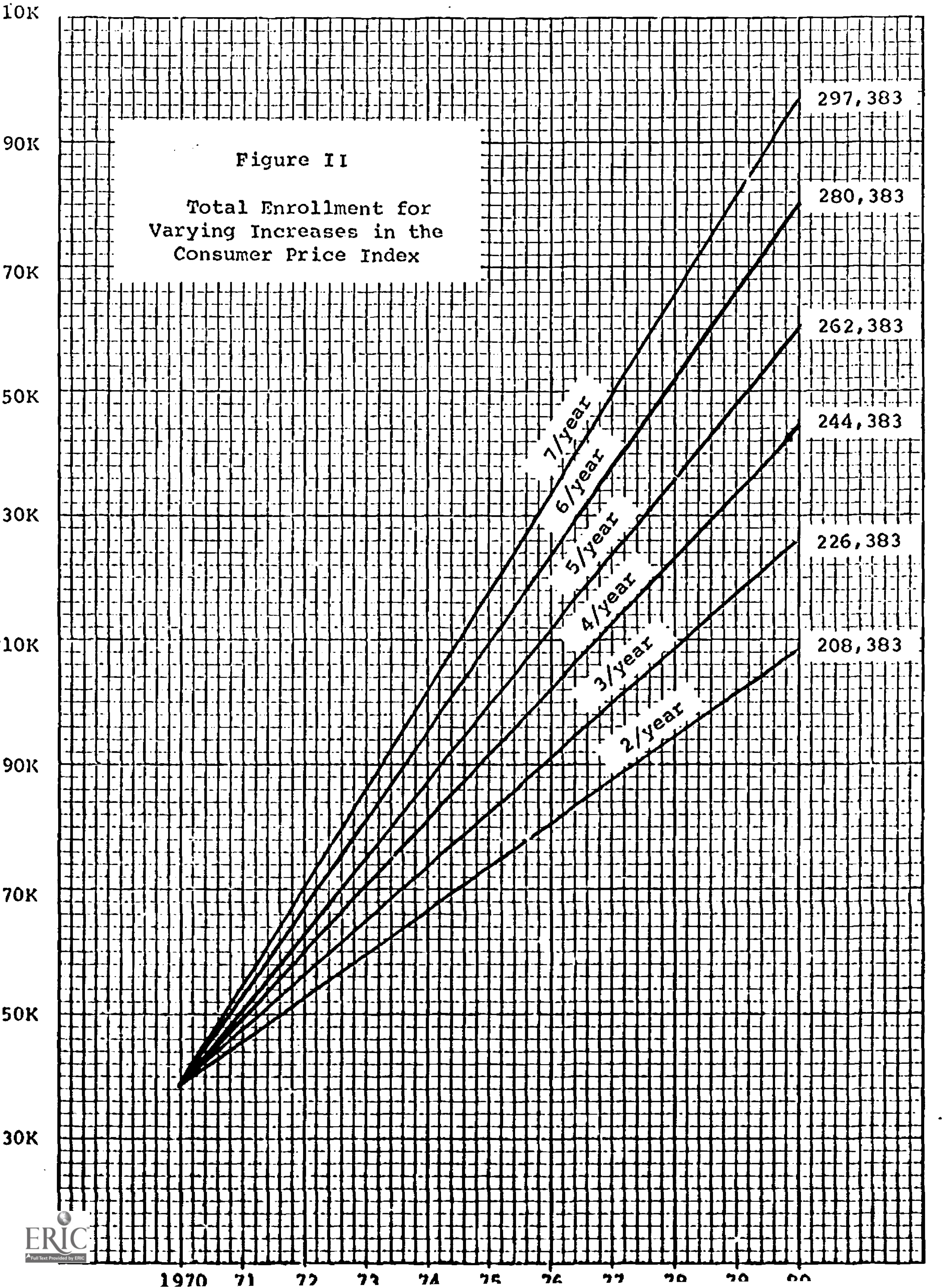

Figure II

Total Enrollment for Varying Increases in the Consumer Price Index

Table 15

TOTAL PREDICTION INTERVAL FOR LINEAR MULTIPLE REGRESSION

Year	$\hat{Y}$	Y_L	Y_U
1970	138,733	116,970	159,221
1971	154,648	131,558	176,738
1972	170,563	144,671	195,499
1973	186,478	157,319	213,868
1974	202,393	168,836	234,517
1975	218,308	182,066	254,635
1976	234,223	195,992	272,279
1977	250,135	208,631	291,870
1978	266,053	221,590	311,359
1979	281,968	234,709	330,797
1980	297,883	247,793	350,433

5.8 Comparison with Previous Projections. The various models developed and applied to this data result in the three predictions shown in Figure III below. The investigators consider the predictions made by Models I and II to be essentially identical. The models represent a continuation of the historical enrollment through the next decade. The investigators subjective feeling is that the total enrollment will not be less than the projections of Models I and II. The projection indicated by the multiple regression model is not inconceivable, but likely represents a rational upper limit of projected values.

The prediction made by the multiple regression model is quite different from that of the other two models. It is conceivable that the independent variables chosen by the investigators are inappropriate and, hence, do not directly relate to total enrollment. Future values of these variables make for potentially invalid estimates.

An additional factor is the difficulty in predicting Consumer Price Index and Deflated Family Income during the next decade. The predictions are indefensible as we extrapolate values eight, nine, and ten years from now.

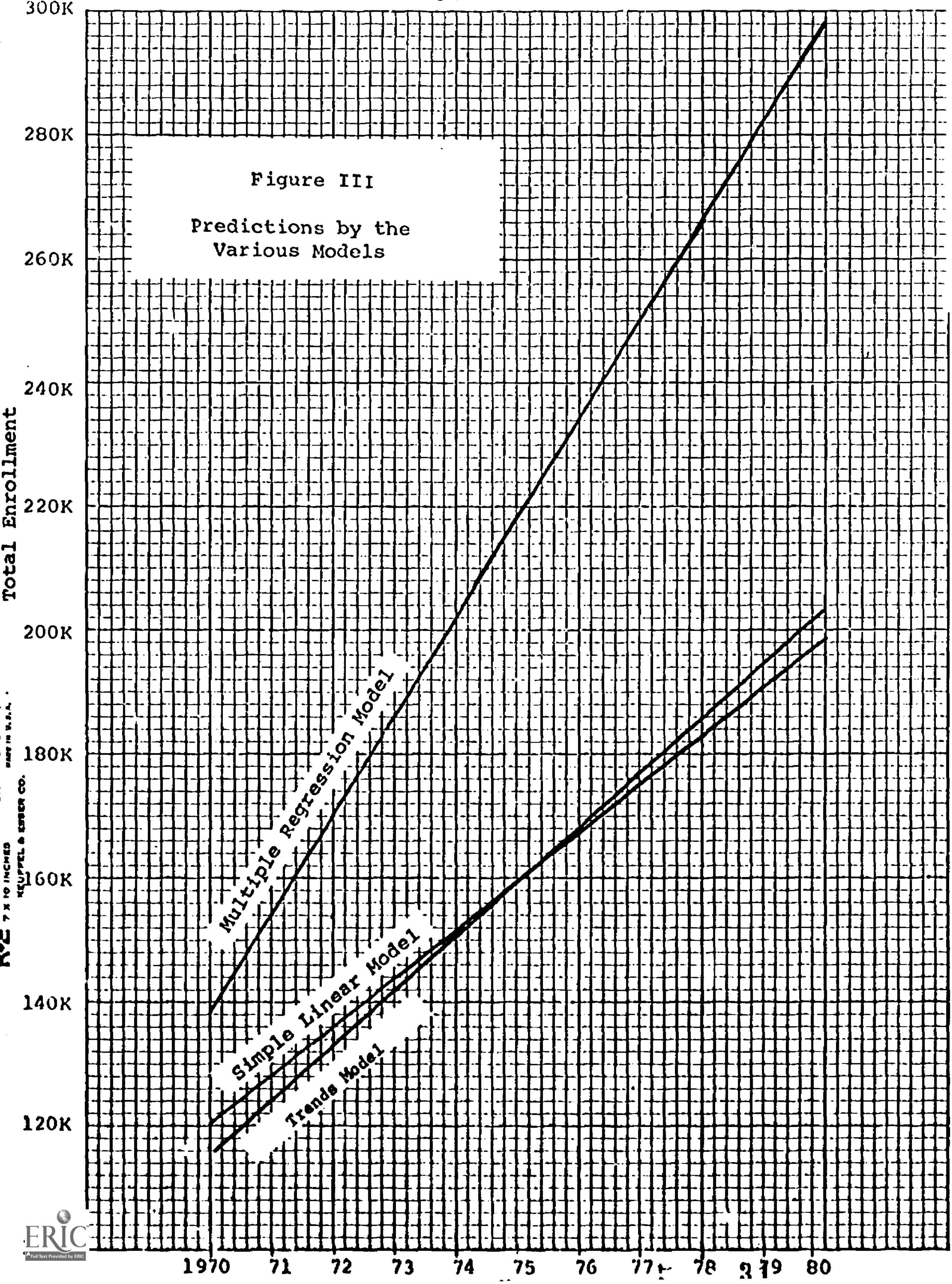
Figure III
Predictions by the Various Models
Total Enrollment
300K
280K
260K
240K
220K
200K
180K
160K
140K
120K
1970
71
72
73
74
75
76
77
78
79
80
Multiple Regression Model
Simple Linear Model
Trends Model

5.0 <u>General Model of Georgia Higher Education Enrollment</u>. Figure IV is a graphic model representing the movement of actual and potential students through the primary and secondary educational systems, and into Georgia institutions of higher education. Additionally, factors which act on candidates for higher education in Georgia are represented.

The authors feel that projections of higher education enrollment would be enhanced if the causes underlying college going were better understood. Currently, predictions are made on observation of the effects and attempting to relate one or more independent variables to the effort.

There are scattered studies in the literature that relate a small number of the blocks shown in Figure IV. A thorough investigation into the causal relationships of Figure IV would be extremely useful in predicting higher education enrollment.

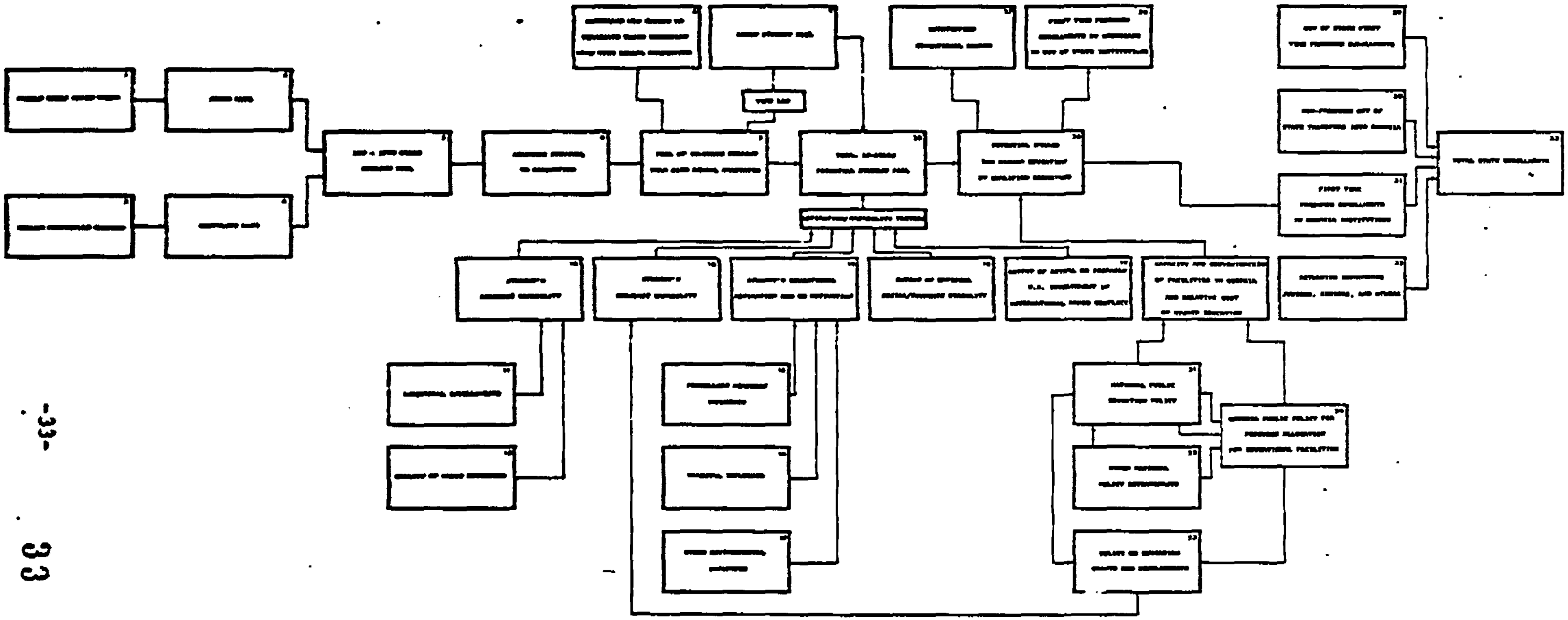

Figure IV

GENERAL MODEL OF GEORGIA HIGHER EDUCATION ENROLLMENT

References:

Draper, Norman and Smith, Harry, Applied Regression Analysis, 1966, Wiley.

Hadley, G., and Whitin, T. M., Analysis of Inventory Systems, 1963, Prentice-Hall, Englewood Cliffs.

"Regression Analysis", User's Guide, 1968, General Electric.