A MODEL FOR PREDICTION OF COURSE SIX CORE COURSE REGISTRATION

by

THOMAS WAYNE LOCKHART

Submitted in Partial Fulfillment

of the Requirements for the

Degree of Bachelor of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1972

Signature of Author . . . .
Department of Electrical Engineering, May 10, 1972

Certified by . . . .
Thesis Supervisor

Accepted by . . . .
Chairman, Departmental Committee on Theses

ABSTRACT

This thesis describes the use of linear regression models for predicting the enrollment in required Electrical Engineering courses. A brief description of regression methods, and Kalman filtering is included. Reasonable models, to be used for prediction, are selected from the models that were tried. Predictions for second term 1971-72 are made, including 80%, 90% and 95% confidence intervals, and compared with the actual enrollments. Predictions for first term 1972-73 are also made and compared with intuitive estimates of the enrollment. The models worked reasonably well considering the extremely random behavior of course enrollment.

TABLE OF CONTENTS

		PAGE
O	LIST OF FIGURES AND TABLES	4
I	INTRODUCTION	5
	a) The Problem b) Other Educational Models	
II	TECHNIQUES	12
	a) Regression b) Kalman Filtering*	
III	THE MODELS	29
	a) Introduction b) 6.01 c) 6.02 d) 6.03 e) 6.04 f) 6.05 g) 6.08 h) 6.231, 6.232, 6.233 i) 6.261 j) 6.262 k) 6.253 l) 6.28	
IV	PREDICTIONS	43
	a) Second Term 1971-72 b) First Term 1972-73	
V	CONCLUSIONS	49
VI	ACKNOWLEDGEMENTS	51
VII	BIBLIOGRAPHY	52
A	APPENDIX - The Data	53

* This section is not necessary for understanding the thesis.

0. LIST OF FIGURES AND TABLES

Figure	II-1	Graph of S602(k) vs S601(k-1)	12a
Table	III-1	The Models for 6.01	30
	III-2	The Models for 6.02	32
	III-3	The Models for 6.03	34
	III-4	The Models for 6.04	35
	III-5	The Models for 6.05	37
	III-6	The Models for 6.08	38
	IV-1	Predicted vs Actual Enrollments for Second Term 1971-72	43
	IV-2	Predictions for First Term 1972-73	46
	A-1	Enrollment Data 1966-72	53

I INTRODUCTION

I.a The Problem

Every term, at MIT and all universities, decisions must be made that allocate resources for individual courses. Teaching Assistants must be hired, money allocated and rooms selected. It would be helpful, in making these decisions, to have predictions of the enrollments in these courses for one term ahead. The purpose of this thesis is to build a model for making one term predictions of the enrollment in required Electrical Engineering courses at MIT. The model is limited to Electrical Engineering courses because most of the resource allocation for individual courses is done at the departmental level. The model is further limited to required courses because they are generally larger and it was expected that required courses would exhibit more regular enrollment behavior than elective courses. This was assumed primarily due to the fact that all of the students in the department must take the required courses.

Previous to this time estimates of course enrollment were made by looking at the enrollment data for the past several terms and noting general trends. This method works adequately well, but there is still some amount of dynamic reallocation required. That is, during the first week or so of the term, there is some shuffling of teaching staff, rooms, etc.

There are many difficulties associated with predictory enrollments. The set of required courses is constantly changing. The prerequisite structure is changing. MIT has a very flexible system in that courses do not have to be taken in any particular year of the student's course of study. It is a routine procedure to add or drop courses up until two weeks before the end of the term, and prerequisites are often waived or ignored. This flexible system provides greater freedom for the student, however, it complicates the enrollment prediction problem. Another more particular difficulty is that the Electical Engineering department at MIT is really two departments. The student can select the Computer Science option, or the Electrical Engineering option, each of which has its own set of requirements. The computer science option has only existed for a few terms and its requirements are still very volatile, there are major changes almost every year. Some of the courses have only been offered three or four times prior to second term 1971-72, and one was discontinued after second term 1970-71.

These difficulties not only add to the uncertainty of the system, but in many cases they result in very little historical data for the same prerequisite and requirement structure. In general, there was very little data to work with - eleven terms of total enrollment figures and five terms of more detailed data. Also, there was data from outside of the department that might have been useful, such as enrollment figures for non-Electrical Engineering courses that are prerequisites of the courses under consideration.

In addition to these data problems, the enrollment in courses changes so much during the term, that for each course each term, there are three values of enrollment: initial registration, fifth term enrollment and final enrollment. It is not unusual for these to vary 20-25 percent. The fifth term data was selected because the initial enrollment is not very interesting. That is, allocation decisions should be based upon the number of students who actually take the course rather than the number of students who sign up to take the course. The final enrollment was not selected because this data would not be available in time to make predictions for the next term. Also relevant in the selection of fifth term data, was that it was available.

The basic assumption that was made for the construction of the models was that enrollment in a course is linearly related to the enrollment in itself and its prerequisites during the preceding terms. Thus the enrollment in the subject Electronic Devices and Circuits, 6.02 for term k, S602(k) is assumed to be linearly related to the enrollment in Introductory Network Theory, 6.01, during term k-1, S601(k-1), because 6.01 is the prerequisite of 6.02. Thus,

$$S602(k) = aS601(k-1) + c + e$$

where S601(k-1) is the enrollment in 6.01 for term k-1

S602(k) is the enrollment in 6.02 for term k

a,c are constants to be determined by regression on data of 6.01 and 6.02 enrollments during preceding terms

e is an error term with zero mean.

This is a very simple model that actually fits the historical data fairly well and made a prediction of enrollment in 6.02 accurate to within 10%. The regression methods used to determine the constants of the model and the confidence intervals for the prediction are briefly discussed in section II-a.

I.b Other Educational Models

There are numerous accounts in the literature of student flow models and enrollment prediction models. All of this work, to the best of the authors knowledge, has been done at a very aggregated level. That is, these models were designed to predict the total enrollment in the university or perhaps the enrollment broken down by department and year. Thus the work was not directly relevant to the current problem of enrollment predictions for individual courses. It was, however, useful to see some of the approaches that are being taken in a closely related area.

Some of this work was done at the University of California by Robert M. Oliver and Kneale T. Marshall. The models they used were probabilistic and were used both for one term and long range forecasting. The results they achieved were quite good with their gross enrollment predictions to within about 5% of the actual values.

The work done at the Western Interstate Commission for Higher Education, WICHE, has a significantly different flavor than that done at the University of California. The WICHE model is much more structurally oriented, in an attempt to recreate the real situation. That is, to predict enrollments by department, the model traces the flow of students through the system, the admissions module, enrollment module, etc. The other primary difference between the WICHE and University of California models is that the WICHE model is intended to be applicable to almost any university, after the parameters are determined for that institution. This would be done primarily by regression analysis of historical data. In addition to this general character, the WICHE model has been designed specifically to interface to other WICHE resource allocation models.

II. TECHNIQUES

II.a Regression*

Regression is a statistical technique useful for determining the relationship among a number of variables. The two variable cases will be covered in detail and then extended to the multiple variable case.

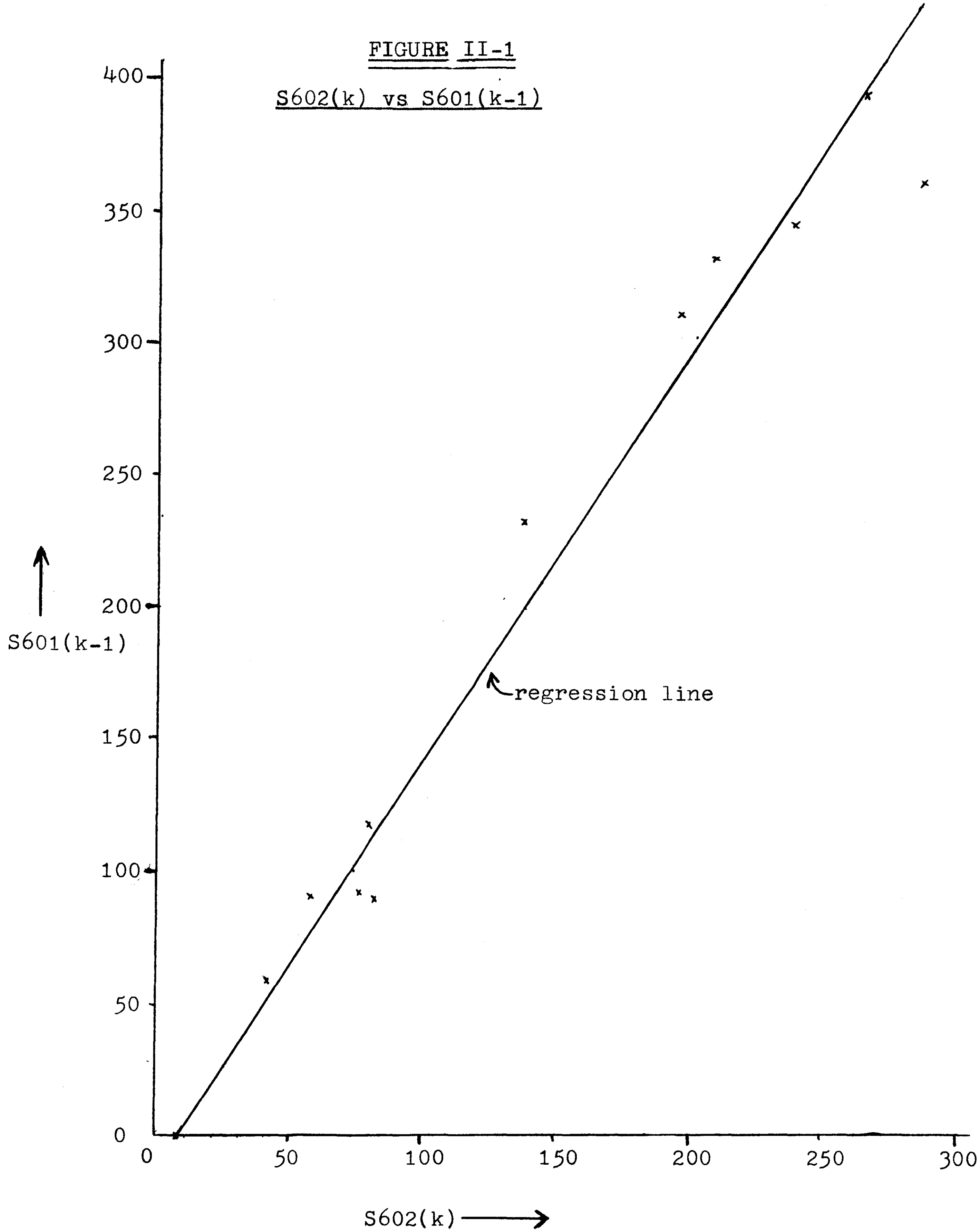
Consider the variables S602(k), the enrollment in subject 6.02 for term k, and S601(k-1), the enrollment in 6.01 for term k-1. From Figure II-1, it is clear that S602(k) and S601(k-1) are related. In fact, this relation- is very close to a straight line, which is reasonable since it says that enrollment in 6.02 is proportional to the enrollment in 6.01, its prerequisite, for the previous term. Our problem now is to find the line that best fits these points. The best line will be considered to be the line that minimizes the sum of the squared errors. If these errors are independent and have finite variance, then from the Guass-Markov Theorem we know that for the class of linear unbiased estimators this solution has minimum variance.

---

* This treatment borrows heavily from reference (6).

FIGURE II-1

S602(k) vs S601(k-1)



Thus if

$$S602(k) = a'' + b'' S602(k-1)$$

we wish to find the values of a" and b" that minimize the sum of the squared errors. To simplify the mathematics, we will translate S601(k-1) into variations from its mean, i.e. determine the new variable

$$s601(k-1) = S601(k-1) - \bar{S}601(k-1)$$

where $\bar{S}601(k-1)$ is the mean of S601(k-1)

We now have

$$S602(k) = a' + b'.s601(k-1)$$

where b' = b" but a' is a new constant

Let S602(k)' be the fitted, or calculated value of S602(k), then the sum of the squared errors is

$$\Sigma (S602(k) - S602(k)')^2$$

because each fitted value S602(k)' is on the estimated line,

$$S602(k)' = a' + b'.s601(k-1)$$

and we wish to find the a' and b' that minimize

$$J(a,b) = \Sigma(S602(k) - a' - b'. s601(k-1))^2$$

By setting

$$\frac{\partial J}{\partial a'} = \frac{\partial J}{\partial b'} = 0$$

and solving, we get

$$a' = \bar{S}602(k) \quad \text{and,}$$

$$b' = \frac{\Sigma\ S602(k) \cdot s601(k-1)}{\Sigma\ s601(k)^2}$$

We have now found a relationship between S602(k) and s601(k-1) but we know that there is some error in this relationship. This error derives both from the fact that we probably have not found the true regression line and the system was probably stocastic anyway, due to measurement errors and so on. We will assume that the S602(k) are identical independant random variables whose means are on the true regression line

$$S602(k) = a + b \cdot s601(k-1).$$

Thus,

$$E(S602(k)) = a + b \cdot s601(k-1) \quad \text{and,}$$

$$\text{variance}\ (S602(k)) = S^2$$

Our least squares estimates of the coefficients, a' and b', are then estimators of the true coefficients a and b. We can show that,

$$E(a') = a$$
$$\text{var}\ (a') = \frac{S^2}{n}$$

Where n is the number of observations

$$E(b') = b$$
$$\text{var}(b') = \frac{S^2}{\Sigma s601(k-1)^2}$$

For a'

$$a' = \bar{S}602(k) = \Sigma \frac{S602(k)}{n}$$

because the S602(k) are random variables,

$$E(a') = \frac{1}{n} \Sigma E\ (s602(k)$$

$$E(a') = \frac{1}{n} \Sigma (a + b \cdot s601(k-1)\ ,$$

$$E(a') = \frac{\Sigma a}{n} + \frac{b}{n} \Sigma s601(k-1),$$ but because

$$s601(k-1) = S601(k-1) - \bar{S}601(k-1),$$

$$\Sigma s601(k) = 0$$

Thus,

$$E(a') = a.$$

Because the S602(k) are independent,

$$\text{var}(a) = \frac{1}{n^2} \Sigma \text{var}(s602(k) = \frac{ns^2}{n^2} = \frac{s^2}{n}$$

Similarly for b',

$$b' = \frac{\Sigma S602(k) \cdot s601(k-1)}{\Sigma s601(k)^2}$$

Since each $\frac{s601(k-1)}{\Sigma s601(k)^2}$ is a constant,

$$E(b') = \frac{\Sigma s601(k-1) \cdot E\ (s602(k))}{\Sigma s601(k-1)^2}$$

$$E(b') = \frac{\sum s601(k) \cdot (a + b \cdot s601(k-1))}{\sum s601(k-1)^2}$$

$$E(b') = \frac{a \sum s601(k-1) + b \sum s601(k-1)^2}{\sum s601(k-1)^2}$$

Thus, because

$$\sum s601(k-1) = 0$$

$$E(b') = b$$

Because the S602(k) are independent,

$$var(b') = \frac{\sum s601(k-1)^2 \cdot var(S602(k))}{(\sum s601(k-1)^2)^2}$$

$$var(b') = \frac{S^2 \sum s601(k-1)^2}{(\sum s601(k-1)^2)^2} = \frac{S^2}{\sum s601(k-1)^2}$$

Thus a' and b' are unbiased estimators of a and b.

Once we have found the coefficients a' and b', the best estimate of S602(k) for a new value of s601(k-1) is

$$S602(k)' = a' + b' \cdot s601(k-1).$$

The variance of this prediction is

$$var(a') + s601(k-1)^2 \; var\,(b') + var\,(S602(k)).$$

This is because

$var(a') + s601(k-1)^2 \; var\,(b')$ is the variance of our estimate of E(S602(k)) and to this we must add the variance of the S602(k).

Therefore, the variance of our prediction is

$$S^2\left(\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1\right)$$

We can now derive the expression for the 90% confidence interval for the prediction. If we assume that S602(k)', our predicted distribution for S602(k) is gaussian, we can normalize S602(k)' to

$$Z = \frac{S602(k)' - S602(k)}{S^2\left(\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1\right)}$$

Now Z is a normal distribution with a mean of zero and a variance of 1. Because we do not know S^2, the variance of the S602(k), we estimate it with S'^2 where

$$S'^2 = \frac{1}{n-2}\Sigma(S602(k) - S602(k)')^2$$

The $\Sigma(S602(k) - S602(k)')^2$ is just the sum of the squared errors of the regression, and the $\frac{1}{n-2}$ is used to make S'^2 an unbiased estimator of S^2. When the S'^2 is substituted into the expression for Z, the result is no longer normal but has the t distribution,

$$t = \frac{S602(k)' - S602(k)}{\sqrt{S1^2\left(\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1\right)}}$$

where t has n-2 degrees of freedom, the same as S'^2.

If we let $t_{.05}$ be the t value that cuts off 5% of the distribution in both tails, then

$$Pr(-t_{.05} < t < t_{.05}) = .90,$$

substituting for t

$$Pr\left(-t_{.05} < \frac{S602(k) - S602(k)}{S'^2\left(\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1\right)} < t_{.05}\right) = .90$$

and,

$$Pr\Big(S602(k)' - t_{.05}\cdot S'\sqrt{\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1} < S602(k)$$

$$< S602(k)' + t_{.05}\cdot S'\sqrt{\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1}\Big) = .90$$

Therefore the 90% confidence interval for a prediction of S602(k) is:

$$S602(k) = S602(k)' \pm t_{.05}\cdot S'\sqrt{\frac{1}{n} + \frac{s601(k-1)^2}{\Sigma s601(k-1)^2} + 1}$$

Now that we have considered the two variables case we can easily extend the ideas to the multiple variable case, which is called multiple regression. For example, we may wish to consider S602(k) as a linear combination of S601(k-1), S602(k-1), and S601(k-2),

$$S602(k) = b_1 + b_2\cdot S601(k-1) + b_2\cdot S602(k-1) + b_3\cdot S601(k-2)$$

Plus an error term e(k).

If we let

$\underline{Y}$ = column vectors of the observed S602(k)

X = an (n x k) matrix where n= the number of observations

and k= the number of independent variables

$$X = \begin{bmatrix} 1 & S601(2) & S602(2) & S601(1) \\ 1 & S601(3) & S602(3) & S601(2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & S601(n+1) & S602(n+1) & S601(n) \end{bmatrix}$$

$\underline{B}$ = a column vector of the b_i's, and

$\underline{e}$ = a column vector of the e(k)'s with $E(\underline{e}) = 0$ and $\text{cov}(\underline{e}) = E(\underline{e}\underline{e}^t) = S^2$,then

$$\underline{Y} = X \cdot \underline{B} + \underline{e}$$

Again, to calculate the coefficient vector, $\underline{B}$, we minimize J, the sum of the squared errors.*

$$J = (\underline{Y} - X\underline{B})^T (\underline{Y} - X\underline{B})$$

$$J = \underline{Y}^T Y - 2\underline{Y}^T X\underline{B} + \underline{B}^T X^T X\underline{B}$$

Setting the vector of partial derivatives of J with respect to the coefficients to zero,

$$0 - 2X^T\underline{Y} + 2X^T X\underline{B} = 0$$

---

* Superscript T indicates transpose.

Thus $\underline{B}' = (X^T X)^{-1} X^T \underline{Y}$

where $\underline{B}'$ is our estimate of the coefficient vector $\underline{B}$.

To show that $\underline{B}'$ is an unbiased estimator of $\underline{B}$, we note that

$$E(B') = E\,((X^T X)^{-1} X^T \underline{Y}) = (X^T X)^{-1} X^T E(\underline{Y})$$

However, because

$$E(\underline{Y}) = X\underline{B}$$

$$E(\underline{B}') = (X^T X)^{-1} X^T X \underline{B} = \underline{B}$$

Thus $\underline{B}'$ is an unbiased estimator of $\underline{B}$.

The covariance of $\underline{B}'$ is just

$$\text{cov}\,(B') = ((X^T X)^{-1} X^T)\ \text{cov}\,(\underline{Y})((X^T X)^{-1} X^T)^T$$

But

$$\text{cov}\,(\underline{Y}) = \text{cov}\,(\underline{e}) = S^2 I \quad \text{so,}$$

$$\text{cov}\,(B') = S^2 (X^T X)^{-1}$$

If we now obtain a new set of values for the independent variables, $\underline{n}$ and wish to calculate the mean and covariance of the prediction Y_o corresponding to $\underline{n}$, we get

$$E\,(Y_o) = \underline{n}\ \underline{B}'$$

$$\text{var}\,(Y_o) = \underline{n}\ \text{cov}\,(B')\ \underline{n}^T + S^2$$

$$\text{var}\,(Y_o) = S^2\ (\underline{n}(X^T X)^{-1} \underline{n}^T) + S^2$$

Similarly to the two variable case we substitute

$$S'^2 = \frac{1}{n-k} \left(\underline{e}\ \underline{e}^T\right) \qquad \text{for } S^2$$

to get the 90% confidence interval for the prediction Y_o,

$$Y_o = \underline{n}\ B' \pm S'\ t_{.05}\sqrt{\underline{n}\ (X^TX)^{-1}\underline{n}^T + 1}$$

where t has n-k degrees of freedom, the same as S'. Note that n is the number of observations and k is the number of independent variables.

We now know enough about linear regression for the purposes of this thesis.

All of the regressions done for this thesis were done with the Econometric Software Package, ESP, available at the MIT Information Processing Center.

ESP contains standard features that do multiple regressions, giving the vector of coefficients $\underline{B}'$ the standard errors and t statistics for $\underline{B}'$, the covarience matrix $S'^2(X^TX)^{-1}$ and S', the standard error of the regression. In addition to this, a special program was written that calculated the 80%, 90% and 95% confidence intervals for the predictions of the models.

II.b Kalman Filtering*

The basic idea behind the Kalman filter is to update the estimate of a state vector on the basis of a noisy measurement of a known function of the state vector. The new estimate of the state vector is to be optimal. We will first consider the case of a static system and then extend this for single-stage linear transitions, and linear multistage processes. The possible application of Kalman filtering to the enrollment prediction problem will then be discussed.

Consider a static system with the n-component state vector x. We have an estimate of x with

$$E(x) = \bar{x} \quad \text{and,}$$

$$E((x-\bar{x})\ (x-x)^T) = \text{cov}\ (x) = M, \text{ a known } (nxn) \text{ positive matrix}$$

We then wish to get a new estimate for x based on the old estimate and a p-component measurement vector z, where,

$$z = H\ x + v$$

---

* Knowledge of the material contained in this section is not necessary for understanding the rest of the thesis. It borrows heavily from reference (1).

H is a known (pxn) matrix

is a p-component error vector for the measurement, with

$E(v) = 0$

$E(vv^T) = \text{cov}(v) = R$ a known (pxp) matrix.

A good estimate of x is the weighted-least-squares estimate. Thus our new estimate of x, which we call x" will be the value of x that minimizes

$$J = \frac{1}{2}\left[(x-\bar{x})^T M^{-1}(x-\bar{x}) + (z-Hx)^T R^{-1}(z-Hx)\right]$$

Note that as M, the covariance of x gets large the error, $(x-\bar{x})$ becomes less important. Similarly, as R gets large, the measurement error $v = z-Hx$ becomes less important.

To minimize J and find x", take the differential of J

$$dJ = dx^T \left(M^{-1}(x-\bar{x}) + H^T R^{-1}(z-Hx)\right)$$

and set the coefficient of dx^T equal to zero

$$M^{-1}(x'' - \bar{x}) + H^T R^{-1}(z-Hx'') = 0$$

collecting the x"

$$(M^{-1}+H^T R^{-1} H)\, x'' = M^{-1}\bar{x} + H^T R^{-1} z$$

$$= M^{-1}\bar{x} + H^TR^{-1}H\bar{x} + H^TR^{-1}z - H^TR^{-1}H\bar{x}$$

$$= (M^{-1} + H^TR^{-1}H)\ \bar{x} + H^TR^{-1}\ (z-H\bar{x})$$

let $P = (M^{-1}+H^TR^{-1}H)^{-1}$ and premultiply

$$x'' = \bar{x} + PH^TR^{-1}\ (z-H\bar{x})$$

We have found x", the new estimate of the state vector, the claim now is, that the covariance matrix of the new estimate is just P where again,

$$P = (M^{-1}+H^TR^{-1}H)^{-1}$$

To show this, let

$e = x''-x$, the error in the estimate

$$e = \bar{x} - x + x'' - \bar{x}$$

using our equation for x" and letting $K = PH^TR^{-1}$

$$e = \bar{x} - x + K\ (z-H\bar{x})$$

using the definition of z

$$e = (\bar{x}-x) + K\ (v+Hx-H\bar{x}) = (\bar{x}-x) + k(v - H(\bar{x}-x))$$

$$e = (I - kH)\ (\bar{x}-x) + Kv$$

because $(\bar{x}-x)$ and v are independent,

$$cov(e) = E(ee^T) = (I-kH)\ M\ (I-kH)^T + KRK^T$$

Remembering that

$$P^{-1} = (M^{-1} + H^TR^{-1}H)$$

and premultiplying by P and postmultiplying by M, yields

$$M = P + PH^TR^{-1}HM = P + KHM$$

or,

$$(I - KH)\, M = P$$

Substituting for (I-KH) M in the equation for cov(e),

$$\text{cov}(e) = P(I-KH)^T + KR^{\ T}$$

$$= P - PH^TK^T + PH^TR^{-1}RR^{-1}HP$$

$$= P - PH^TR^{-1}HP + PH^TR^{-1}HP = P$$

Thus cov(e) = P

From $P = (M^{-1} + H^TR^{-1}H)^{-1}$

and the fact that $H^TR^{-1}H$ is at least a positive semi-definite matrix, it is obvious that P the error covariance matrix after measurement is never larger than M, the error covariance matrix before measurement. It is interesting to note that it can be shown that e and x" are uncorrelated.

To extend these results for single-stage linear transitions, consider then a system which experiences a discrete change from state 0 to state 1, described by the equation

$$x_1 = \Phi_o\, x_o + F_oW_o \qquad \text{where,}$$

$_o$ is a known (nxn) transition matrix,

F_o is a known (nxr) matrix

$$E(W_o) = \overline{W}_o$$

$$\text{cov}(W_o) = E(W_o - \overline{W}_o)(W_o - \overline{W}_o)^T = Q_o$$

The state x_o is a random vector with mean x_o" and covariance P_o. This, along with the fact that x_o and W_o are independent permit us to write

$$E(x_1) = \overline{x}_1 = \Phi_o x_o'' + F_o \overline{W}_o$$

$$\text{cov}(x_1) = M_1 = \Phi_o P_o \Phi_o^T + F_o Q_o F_o^T$$

If we make a measurement, z_1, after the transition to state 1 we can update the estimate of x_1 on the basis of z_1 and the results we obtained before, to get

$$x_1'' = \overline{x}_1 + P_1 H_1^T R_1^{-1}(z_1 - H_1 \overline{x}_1),$$

$$P_1 = (M_1 + M_1^T R^{-1} H_1)^{-1}$$

Noting that $\overline{x}_1$ and M_1 are the estimate and covariance matrix of x_1 before measurement and x_1" and P_1 are the estimate and covariance matrix of x_1 after measurement, we can easily see how, for a multi-stage process, x_1" and P_1 could be used to find $\overline{x}_2$ and M_2. With another observation z_2 we could get x_2" and P_2 and the procedure could continue as long as we had measurements.

It was originally hoped that this technique of Kalman filtering might have been used to update our predictions of enrollment. We soon realized, however, that in order to use Kalman filtering, we would have to have a measurement of the enrollment after the system had changed to that state. For example, suppose it is now term 12 and we want to predict S602(13). We can find an estimate based on S601(12), but we cannot use a Kalman filter to improve this estimate of S602(13) until we have a measurement of S602(13) which of course cannot happen until term 13. Thus the Kalman filter cannot improve the estimate of the state for future time periods.

In order for the use of Kalman filters to make any sense at all we would have to hypothesize a variance for the observed value of S602(13). This is reasonable, since some errors are likely to be made in the collection of the data. This variance is,however, much less than the variance of our prediction, so that once we have a measurement of S602(13), that measurement is essentially our best estimate of the state. For these reasons, it was decided not to use Kalman filtering techniques in the rest of the thesis. The Kalman filter is especially useful in applications where the measurement errors are of the same order as the prediction errors, which is not the case for the predictions of the models of course enrollment.

III THE MODELS

III.a Introduction

This section describes, by course, the models that were constructed. For example, the models of 6.01 are presented, along with the values of these coefficients, and the standard error for each regression. The possible physical meaning of each model is considered. In some cases, the signs of the coefficients are not what would be expected from the physical situation. In other cases, the coefficients are exactly what one would expect. In the light of this physical interpretation and consideration of the standard error of the regression, good models are selected for predicting the enrollment for the second term 1971-72.

III.b 6.01

Introductory Network Theory, 6.01, is the first required electrical engineering course. It has two prerequisites, Physics II, 8.02 and Differential Equations, 18.03. Both 8.02 and 18.03 are very large courses. 8.02 is an institute requirement and 18.03 is required by a lot of departments, thus they would not be expected to closely correlate with 6.01. For this reason, and the fact that the data for 8.02 and 18.03 were not readily available, 6.01 was not regressed with either 8.02 or 18.03. The enrollment in 6.01 would be likely to be correlated with its enrollment the previous term, however, because an approximately constant number of people take the course each year. Most EE students take the course in the first term of their sophomore year, which results in a large first term enrollment and a small second term enrollment. That is, S601(k) oscillates with a period of one year. This oscillation reflects itself in other courses for which 6.01 is a prerequisite.

TABLE III-1

The Models for 6.01

Model No.	Dependent Variable	Independent Variables	Coefficients b_1	b_2	b_3	b_4	Standard Error, S
1	S601(k) =	$b_1 \cdot S601(k-1)$	.535				227
2	S601(k) =	$b_1 + b_2 \cdot S601(k-1) + b_3 \cdot S601(k-2)$	95	-.168	.696		35
3	S601(k) =	$b_1 + b_2 S601(k-1)$	401	-.901			34
4	S601(k) =	$b_1 + b_2 S601(k-2)$	24	.857			32

Model 1 in Table III-1 has an extremely large standard error. This is reasonable because the model claims that S601(k) is proportional to S601(k-1). In fact, because there is no constant term in the model, the constant of proportionality must be positive. It is absurd, however, to claim that S601(k) will be larger if S601(k-1) is larger. Model 1 is therefore rejected.

Model 2 is much more reasonable. It makes physical sense insofar as S601(k) is negatively correlated with S601(k-1), which would be expected. Because it has a larger standard error and is more complex, Model 2 was rejected in favor of Models 3 and 4.

Model 3 makes good sense in that it reflects the idea that the total number of people taking 6.01 over two terms is a constant, with a magnitude of about 400. Model 4 is good also, however, and it has a smaller error term than Model 3. Because Models 3 and 4 were both reasonable, they were both used for prediction.

III.c 6.02

As mentioned in Section I, 6.02, Electronic Devices and Circuits, has only one prerequisite, 6.01.

Thus it is reasonable to assume that S602(k) is correlated with S601(k-1).

TABLE III-2

The Models for 6.02

Model No.	Dependent Variable		Independent Variable	b_1	b_2	b_3	b_4	Standard Error, S
1	S602(k)	=	$b_1S601(k-1)$	.698				18
2	S602(k)	=	$b_1+b_2S601(k-1)$	7	.673			19
3	S602(k)	=	$b_1+b_2S601(k-1)$	150	.337	-.470		16
			$+b_3S602(k-1)$					
4	S602(k)	=	$b_1+b_2S601(k-1)$					
			$+b_3S601(k-2)+$					
			$b_4S601(k-3)$ +	54	1.022	.574	-.066	
			$b_5S602(k-1)$ +					
			$b_6S602(k-2)$ +					
			$b_7S602(k-3)$	-1.720	-.108	.948		15
5	S602(k)	=	$b_1+b_2\cdot S601(k-1)$					
			$+b_3\cdot S601(k-2)$					
			$+b_4\cdot S602(k-1)$	-17	.261	.106	-.036	
			$+b_5\cdot S602(k-2)$	.592				11
6	S602(k)	=	$b_1+b_2S601(k-1)$					
			$+b_3S601(k-2)$	-19	.704	.069		14
7	S602(k)	=	$b_1+b_2S601(k-1)$					
			$+b_3S602(k-1)$					
			$+b_4S601(k-2)$	14	.633	-.395	.266	12

From Table III-2 we see that S602(k) and S601(k-1) are indeed related. Model 1 demonstrates that S602(k) and S601(k-1) are very highly correlated because of its low error term. Despite the fact that Model 2 has a higher error term than Model 1, it is favored because it is quite likely that there are factors not considered with such simple models that would create an error term with non-zero mean.

Model 3 was introduced to see what effect adding S602(k-1) to the regression would have. In that it lowered the error term, the addition was helpful. The 4th and 5th models were tried to see if a lot of variables was better than a few. It appears, though, that the added complexity and loss of useable data points more than counteracts the slightly lower error terms. Note that both models have coefficients that are close to zero. Those terms could probably be ignored. In Model 6, again, one of the coefficients was very close to zero, implying that term was probably insignificant. It was concluded that Model 7 was probably the best, because it has enough complexity to have a low standard error, but none of its coefficients are close to zero. Models 2, 3, and 7 were used for prediction.

III.d 6.03

Electromagnetic Fields and Energy, 6.03, has one prerequisite, 6.01. However, it is almost always taken immediately following 6.02.

TABLE III-3

The Models for 6.03

Model No.	Dependent Variable	Independent Variables	Coefficients b_1	b_2	b_3	b_4	Standard Error, S
1	S603(k) =	b_1+b_2·S601(k-1) $+b_3$·S603(k-1)	171	-.106	-.553		15
2	S603(k) =	b_1+b_2·S601(k-1)	162	-.321			16
3	S603(k) =	b_1+b_2·S601(k-2) $+b_3$·S603(k-1)	82	.195	-.305		15
4	S603(k) =	b_1+b_2S601(k-2)	30	.297			14
5	S603(k) =	b_1+b_2·S601(k-1) $+b_3$S601(k-2)+ b_4S603(k-1)	-45	.604	.516	-1.056	14
6	S603(k) =	b_1+b_2S602(k-1)	24	.459			7

We can see from Table III-3 that Model 6 has a significantly lower standard error than the other models. The first two models, 1 and 2, have S603(k) negatively correlated with S601(k-1). This outcome does not make much sense physically because courses should be positively correlated with their prerequisite. Models 3, 4, and 5 all make physical sense, but Model 6 is better because of its lower error term, thus Model 6 was the only 6.03 model used to make predictions.

III.e 6.04

Electrodynamics, 6.04, recently replaced two EE core subjects, 6.06 and 6.07. Thus there are only 4 terms of data for the course prior to second term 1971-72. 6.04 is the only required EE course that has two prerequisites, 6.03 and 6.05 within the department. Thus, S604(k) was regressed against S605(k-1), S603(k-1), and both of them together.

TABLE III-4

The Models for 6.04

Model No.	Dependent Variable	Independent Variable	Coefficients b_1	b_2	b_3	b_4	Standard Error, S
1	S604(k) =	b_1+b_2S605(k-1)	179	-.778			26
2	S604(k) =	b_1+b_2S603(k-1)	3	.762			5
3	S604(k) =	b_1+b_2S603(k-1) $+b_3$S605(k-1)	-40	.878	.240		3

As can be seen from Table III-4, Model 1, with just S605(k-1) and a constant for the independent variables is not very good. That is, it has a high error term and it says that S604(k) is negatively correlated with its prerequisite, S605(k-1). This does not make physical sense, so Model 1 was rejected. Both Models 2 and 3 make good physical sense, so they were both used to make predictions.

III.f 6.05

Despite the fact that 6.05, Signals and Systems, is a required EE course, and has a prerequisite, 6.01, it was very difficult to find a model that fit the data reasonably well. This is due in part to the large proportion (20%) of enrollment from outside of the department, as well as the fact that 6.05 is generally not taken the term immediately following 6.01.

TABLE III-5

The Models for 6.05

Model No.	Dependent Variable	Independent Variables	Coefficients b_1	b_2	b_3	b_4	Standard Error
1	S605(k) =	$b_1 \cdot$S601(k-1)	.529				74
2	S605(k) =	$b_1+b_2 \cdot$S601(k-2) $+b_3 \cdot$S601(k-3)+ $b_4 \cdot$S601(k-4)	714	-.633	-1.272	-.706	20
3	S605(k) =	$b_1+b_2 \cdot$S601(k-1) $+b_3 \cdot$S601(k-2)+ $b_4 \cdot$S605(k-1)	159	-.022	-.138	.134	34
4	S605(k) =	b_1+b_2S601(k-2) $+b_3 \cdot$S601(k-3)	506	-.873	-.79		21
5)	S605(k) =	$b_1+b_2 \cdot$S601(k-1) $+b_3 \cdot$S601(k-2)	187	-.059	-.150		31
6	S605(k) =	$b_1+b_2 \cdot$S601(k-1)	122	-.095			28
7	S605(k) =	$b_1+b_2 \cdot$S602(k-1)	175	-.198			25
8	S605(k) =	$b_1+b_2 \cdot$S601(k-2) $+b_3 \cdot$S601(k-3)+ $b_4 \cdot$S605(k-1)	411	-.106	-.830	1.07	12

Table III-5 shows that Model 1 can be rejected just from its large error term while Models 3, 5 and 6 can be rejected for the very small coefficient for S601(k-1). Model 2 does not seem reasonable since it implies that 6.05

is negatively correlated to its prerequisite for the last four terms (except the most recent). The same negative coefficients appear in both Models 4 and 8. The error term for Model 8 is significantly lower than for any of the other models, hence it was selected for prediction, as was Model 7. Model 7 was formulated in recognition of the fact that it is quite often the case that 6.05 is taken the term after 6.02. The coefficient for S602(k-1) is negative however, which is odd, but the error is relatively low.

III.g 6.08

Like 6.01, the prerequisites of 6.08, Statistical Mechanics and Thermodynamics, are outside of the EE department. They are 8.04, Principles of Quantum Physics, or 8.211, Introduction to Quantum Physics. The same arguments hold in this case as for 6.01, so 6.08 was regressed against itself and against a constant.

TABLE III-6

The Models of 6.08

Model No.	Dependent Variable	Independent Variables	Coefficient b_1	b_2	b_3	b_4	Standard Error, S
1	S608(k) =	$b_1+b_2 \cdot S608(k-1)$ $+b_3 S608(k-2)$	59	.059	.034		10
2	S608(k) =	$b_1+b_2 \cdot S608(k-1)$	72	-.121			9
3	S608(k) =	b_1	67				9

Table III-6 shows that while the error of the regression is about the same in the three cases, two coefficients in Model 1 are almost zero. Thus Models 2 and 3 were used for the predictions.

III.h 6.231, 6.232, and 6.233

6.231, Programming Linguistics, 6.232, Computation Structures, and 6.233, Information Systems, are relatively new series required by all computer science option students. 6.233 had only been offered four times prior to second term 1971-72. 6.231 was the prerequisite for 6.232, which in turn is the prerequisite for 6.233. This would probably make for meaningful and accurate models of these courses. But recently 6.231 was discontinued; 6.251, the new prerequisite for 6.232, has been around a long time and has a farily large student population that will not take 6.232. There was no point in trying to model 6.231 since it doesn't exist anymore. There is also not much point in modeling 6.232 either, because there are only one or two terms of data available with its new prerequisite. 6.233, however, had a couple of nice models using 6.232 from prior terms.

Model 1: $S6233(k) = b_1+b_2 \cdot S6232(k-1)$

where $b_1 = 1$, $b_2 = .962$ and the standard error, $s = 20$

Model 2: $S6233(k) = b_1+b_2 \cdot S6232(k-1)+b_3 \cdot S6232(k-2)$

where $b_1 = 7$, $b_2 = .360$, $b_3 = .794$ and the standard error = 2.

Because Model 1 had such a small constant, almost zero, and because the error of Model 2 was so much smaller, Model 2 was selected for prediction purposes.

III.i <u>6.261</u>

Introduction to Modern Algebra, 6.261, has no prerequisites other than 18.03, and is required for all 6-3 students. The only model tried was

$S6261(k) = b_1+b_2 \cdot S6261(k-1)$

where $b_1 = 79$, $b_2 = -.206$ and the standard error, $S = 43$

Because of the exceptionally large error term (50%), this model would probably not predict very well. It was not used in the predictions, because 6.261 wasn't offered second term 1971-72.

III.j 6.262

The big problem with building a model of 6.262, Computability, Formal Systems and Logic, was that it was only offered for three terms prior to second term 1971-72. 6.262 has one prerequisite, 6.261.

The only model tried,

$$S6262(k) = b_1 + b_2 \cdot S6261(k-1)$$

where $b_1 = 42$, $b_2 = .168$ and $S = 31$,

made use of this fact. This model was used for the predictions.

III.k 6.253

6.261 is also the prerequisite for 6.253, Theoretical Models for Computation. 6-3 students are required to take one of 6.253 or 6.262. The two models tried for 6.253 where

Model 1: $S6253(k)=b_1 + b_2 \cdot S6261(k-1)$

where $b_1 = 47$, $b_2 = -.160$, and $S = 15$

Model 2: $S6253(k)=b_1 + b_2 \cdot S261(k-1)+b_3 \cdot S6261(k-2)$

where $b_1 = 20$, $b_2 = .023$, $b_3 =.269$, and $S= 9$

No prediction was made because 6.253 was not offered second term 1971-72.

III.1 6.28

Probabalistic Systems Analysis, 6.28, has no prerequisites other than second term Calculus, 18.02. Thus, it was merely regressed against itself for one, and two terms previously. The resulting models were:

Model 1: $S628 = b_1 + b_2 \cdot S628(k-1) + b_3 S628(k-2)$

where $b_1 = 48$, $b_2 = .683$, $b_3 = .019$ and $S = 12$

Model 2: $S628(k) = b_1 + b_2 \cdot S628(k-1)$

where $b_1 = 25$, $b_2 = .865$, and $S = 13$

Because the b_3 coefficient of Model 1 was so close to zero, Model 2 was selected for the predictions of the enrollment for second term 1971-72. These predictions are treated in the following section.

IV PREDICTIONS

IV.a Second Term 1971-72

In Table IV-1 of this section, the models used to predict the enrollments for second term 1971-72, the predictions and the 80% 90% and 95% confidence intervals are compared with the actual enrollments for the same period.

TABLE IV-1

Predicted vs Actual Enrollments for 2nd Term 1971-72

Model No.	Course	Independent Variables	Enrollment Act	*Predct	Confidence Intervals ±80%	90%	95%
1	6.01	1,S601(k-2)	149	* 123	±20	27	34
2	6.01	1,S601(k-1)	149	* 197	±15	20	25
3	6.02	1,S601(k-1), S602(k-1), S601(k-2)	143	* 156	±29	39	50
4	6.02	1,S601(k-1)	143	* 159	± 8	11	14
5	6.02	1,S601(k-1), S602(k-1)	143	* 188	±24	32	39
6	6.03	1,S602(k-1)	74	* 61	± 4	5	7
7	6.04	1,S603(k-1)	76	* 79	± 5	8	12
8	6.04	1,S603(k-1), S605(k-1)	76	* 93	±18	38	77
9	6.05	1,S601(k-2), S601(k-3), S605(k-1)	158	* 238	±35	49	64
10	6.05	1,S602(k-1)	158	* 159	±14	19	24
11	6.08	1,608(k-1)	85	* 65	± 4	6	7
12	6.08	1	85	* 67	± 5	7	8
13	6.233	1,S6232(k-1), S6232(k-2)	83	* 112	±12	24	48
14	6.262	1,S6261(k-1)	50	* 56	±65	135	271
15	6.28	1,S628(k-1)	138	* 171	±12	16	20

It may at first seem unusual that of the fifteen predictions, for only seven the actual value fell within the 95% confidence interval. This is easier to take, however, when one sees that three of the actual values just missed the 95% confidence interval by a few students, and three of the remaining five "bad" predictions had claimed 95% confidence intervals of 7, 8, or 14. These small confidence intervals seem a bit preposterous considering the stocastic nature of the enrollment process. That is, fluctuations of 10-15 would not be unusual for a class with around 70 people in it.

The predictions of the two remaining models,

$$S602(k) = b_1+b_2 \cdot S601(k-1) \quad \text{and}$$

$$S628(k) = b_1+b_2 \cdot S628(k-1),$$

just seem to be bad. There is always, of course, the 5% chance that the actual value will fall outside of the 95% confidence interval. But, at least for the 6.01 model, I do not think that this was the case because the 95% confidence intervals for the predictions of the two 6.01 models do not overlap. At least one of the 6.01 models had to be outside of its 95% confidence interval. This is probably due to the inability of the models to fit the real situation.

There are other places where error has crept into the calculation of the confidence interval. For example, the errors e_i (difference between fitted and actual value) are assumed to be normal which may not be true. And, the enrollments in a particular course, for different terms, are assumed to be independent. That is, S601(k) is assumed to be independent of S601(k-1), S601(k-2), etc., which is clearly not the case. This inaccuracy in the assumptions necessary for regression could well have reflected itself in the confidence intervals for the predictions. There are no predictions for 6.261 or 6.253 in Table IV-1 because they were not offered second term 1971-72.

IV.b First Term 1972-73

Predictions for First Term 1972-73, with their 80%, 90%, and 95% confidence intervals, are presented in Table IV-2.

TABLE IV-2

Predictions for First Term 1972-73

Model No.,	Course	Independent Variables	Enrollment Proj*Predct	Confidence Intervals 80%	90%	95%
1	6.01	1,S601(k-2)	210 * 220	±14	19	23
2	6.01	1,S601(k-1)	210 * 262	±17	23	28
3	6.02	1,S601(k-1), S602(k-1), S601(k-2)	100 * 106	±11	15	19
4	6.02	1,S601(k-1)	100 * 106	± 9	12	15
5	6.02	1,S601(k-1), S602(k-1)	100 * 113	±18	24	29
6	6.03	1,S602(k-1)	90 * 91	± 3	5	6
7	6.04	1,S603(k-1)	60 * 59	± 4	6	8
8	6.04	1,S603(k-1), S605(k-1)	60 * 59	± 6	9	13
9	6.05	1,S601(k-2), S601(k-3), S605(k-1)	140 * 217	±46	62	78
10	6.05	1,S602(k-1)	140 * 147	±11	14	17
11	6.08	1,S608(k-1)	65 * 63	± 9	12	14
12	6.08	1	65 * 68	± 5	7	8
13	6.28	1,S628(k-1)	140 * 160	±13	15	19

These predictions were based on all of the data available through second term 1971-72, as opposed to the predictions of the last section which were only based on the data through first term 1971-72. That is, all of the coefficients of the models were redetermined from this larger data base and, of course, the predictions had to use the second term 1971-72 enrollment figures for all of the (k-1) terms in the prediction equations. It is suggested that, should these models be used for further predictions, the coefficients of the models be redetermined each term to take into account the extra data points.

At the time of this writing there are no actual enrollment figures for first term 1972-73 to compare the predictions with. Instead, there is an item in Table IV-2 called "Projected Enrollment". These are my guesses for the enrollment, obtained by noting general trends in the enrollment patterns. It is interesting to note that the projections and predictions for Models 2, 9, and 15 differ significantly and that it was just these models that made "bad" predictions for second term 1971-72. These predictions could probably be ignored. It seems like a good idea to compare common sense projections with the calculated predictions in order to detect wildly aberrant predictions. Except for Models 2, 9 and 15, as previously noted, the predictions of Table Iv-2 seem very reasonable. Predictions were not made for 6.233 or 6.262 because they are second term courses, and predictions were not made for 6.253 or 6.261 because 6.261 was not offered the previous term.

V CONCLUSIONS

This thesis has shown that despite the extremely random nature of the enrollment process, reasonably good predictions for required courses can be made using linear regression techniques. The best predictions can be made for the courses such as 6.02, whose prerequisites are usually taken the term immediately before the term of interest. This is reasonable because it is in these cases that the models make the most sense. Not suprisingly, the worst predictions are made for courses such as 6.01, or 6.28, which have no prerequisites, or at least no specific prerequisites. That is, both 6.01 and 6.28 have prerequisites which are institute requirements.

It is not likely that useful prediction models could be built for most non-required Electrical Engineering courses. There are two reasons for this; the enrollment in most of these courses is very small (i.e., less than 50), which would make them subject to much larger percentage fluctuations; and, so far as resource allocation is concerned, it does not matter whether 20 or 25 people are expected to take the course. There are a few non-required courses, such as 6.00 or 6.14, for which the enrollment is large enough that an accurate enrollment prediction could help the resource allocation problem. The difficulty with these courses is that they usually don't have specific prerequisites, which, as was mentioned earlier, would cause difficulties in finding physically meaningful models.

An interesting possibility for further research would be to fit the WICHE student flow model to MIT, (if this has not already been done), and use its predictions of Electrical Engineering enrollment, in addition to the methods described in this thesis, to predict course enrollment. This would be especially helpful in the case of a course with no prerequisites. Another possibility for further research would be to build a model that uses the predictions of course enrollment to make resource allocation decisions.

VI ACKNOWLEDGEMENTS

I am indebted to the many people who helped with this thesis. I would like to take this opportunity to thank: Professor J.D. Bruce, Associate Dean of the School of Engineering at M.I.T., for providing me with information on the student flow models of the University of California and WICHE; Professor L.D. Braida, Executive Officer of the Department of Electrical Engineering, for providing the data used as well as several helpful discussions about course enrollment; Aleco Sarris and Nathaniel Mass for informal discussions of the models, regression and ESP; and of course, my thesis advisor Professor S.K. Mitter, who thought of the topic and without whom the thesis would have been impossible. I would also like to mention the MIT Information Processing Center, whose implimentation of ESP was used for the calculations.

VII BIBLIOGRAPHY

1. Bryson, Arthur E. Jr., and Ho, Yu-Chi. Applied Optimal Control. Waltham, Blaisdell, 1969.

2. Galiana, Francisco D. An Application of System Identification and State Prediction to Electrical Load Modelling and Forecasting. MIT PhD Thesis, 1971.

3. Lovell, C.C. Student Flow Models A Review and Conceptualization. Boulder, Co.,WICHE,1971.

4. Marshall, Kneale T., and Oliver, Robert M. A Constant Work Model for Student Attendance and Enrollment. Office of the Vice President - Planning and Analysis,University of Califronia, 1969.

5. Oliver, Robert M. Models for Predicting Gross Enrollments at the University of California. Office of the Vice President- Planning and Analysis,University of California, 1968.

6. Wonnacott, Ronald J., and Wonnacott, Thomas H. Econometrics. New York, John Wiley & Sons, 1970.

A APPENDIX - The Data

The enrollment data used in this thesis came from Electrical Engineering Memorandum 4017E, April 10, 1972. Only the relevant portions of that document are reproduced here.

TABLE A-1

Enrollment Data 1966-72

Course	1966-67		1967-68		1968-69		1969-70		1970-71		1971-72	
	1	2	1	2	1	2	1	2	1	2	1	2
6.01*	361	59	382	93	346	94	327	92	310	115	226	149
6.02*	25	286	42	269	62	235	75	205	89	195	81	143
6.03*	180	32	164	49	151	60	129	56	116	64	100	74
6.04*	-	-	-	-	-	-	-	101	41	91	57	76
6.05*	64	189	95	162	132	143	111	136	131	168	193	158
6.08	92	57	68	73	62	59	52	63	61	83	65	85
6.231	-	23	-	65	52	36	141	92	131	78	-	-
6.232	-	-	18	23	70	59	54	68	95	103	67	-
6.233	-	-	-	14	-	49	-	72	-	96	1	83+
6.253	41	44	62	71	63	38	29	22	51	22	59	-
6.261	-	-	-	31	64	-	111	34	137	49	85	-
6.262	-	-	-	-	-	44	-	86	-	49	-	50
6.28	74	73	98	130	122	132	125	130	146	157	168	138

* Includes enrollment for 6.0X3, the 6-2 version of 6.0X. 6.01 includes enrollment in 6.001.

\+ Correction to Memorandum 4017E