DOCUMENT RESUME

ED 474 040		HE 035 673
AUTHOR	Armacost, Robert L.; Wilson, Alicia L.	
TITLE	Three Analytical Approaches for Predicting Enrollment at a Growing Metropolitan Research University. AIR 2002 Forum Paper.	
PUB DATE	2002-06-00	
NOTE	28p.; Paper presented at the Annual Meeting of the Association for Institutional Research (42nd, Toronto, Ontario, Canada, June 2-5, 2002).	
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)	
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.	
DESCRIPTORS	*Enrollment Projections; Higher Education; *Models; *Prediction; Research Universities; Urban Universities	
IDENTIFIERS	*Optimization	

ABSTRACT

In a large metropolitan research university, multiple enrollment models are required to fulfill the needs of the many constituents and planning horizons. In addition, the method of predicting enrollment in a growth environment differs from that of universities in a stable environment. Three models for predicting enrollment are discussed, with a description of the underlying methods. The models include: (1) a long-term aggregate university model; (2) a short-term detailed university model; and (3) an enhanced graduate prediction model by college. The analytical approaches discussed include an embedded optimization model used to "fit" transition factors and a Markov chain to track transition probabilities within colleges. (Author/SLD)

Three Analytical Approaches For Predicting Enrollment At A Growing Metropolitan Research University

Dr. Robert L. Armacost, Director
Alicia L. Wilson, Coordinator of Computer Applications

University Analysis and Planning Support
University of Central Florida
12424 Research Parkway, Suite 488
Orlando, FL 32826-3207
(407) 882-0285
FAX: (407) 882-0288
e-mail: armacost@mail.ucf.edu
awilson@mail.ucf.edu

Paper presented at the 42nd Annual Forum of
The Association for Institutional Research
Toronto, Canada
June 2-5, 2002

Three Analytical Approaches For Predicting Enrollment At A Growing Metropolitan Research University

Abstract

In a large, metropolitan research university, multiple enrollment models are required to fulfill the needs of its many constituents and planning horizons. Furthermore, the method of predicting enrollment in a growth environment differs from universities in a stable environment. Three models will be discussed as well as the underlying methods. The breadth of the models discussed includes a long-term aggregate university model, a short-term detailed university model, and an enhanced graduate prediction model by college. The analytical approaches discussed include an embedded optimization model used to "fit" transition factors and a Markov chain to track transition probabilities within colleges.

Three Analytical Approaches For Predicting Enrollment At

A Growing Metropolitan Research University

In a large, metropolitan research university, multiple enrollment models are required to fulfill the needs of its many constituents and planning horizons. Furthermore, the methods of predicting enrollment in a growth environment differ from universities in a stable environment. Three models will be discussed as well as their underlying methods. The first is a broad 5-year model that predicts FTEs by level and distributes them to several campuses. This model uses judgment-based estimated growth rates to predict enrollment levels, and includes historical as well as control factors for distributing growth. The second model develops short-term predictions for headcount, student credit hours, and the number of full-time equivalents at the university overall, as well as by level and classification. An embedded optimization model is used to "fit" transition factors to improve the performance of the model. This model is used to examine the effects of different admission policies. The third model predicts graduate enrollment by college. A Markov chain is used to develop transition probabilities within colleges and better capture the behavior of graduate students.

Overview

Why Do Enrollment Projections?

Students are the cornerstone of the university environment. Almost all decisions at the university-level involve student enrollment at some level. Hopkins and Massey indicate that accurate forecasts of student enrollment are needed for at least three purposes: predicting income from tuition, planning courses and curriculum, and allocating marginal resources to academic departments (Planning, 352). Models should be designed for a specific purpose and the degree of approximation that is acceptable will depend on the purpose of the model (Planning, 4).

Furthermore, all newly developed models must be validated (Planning, 4). In order to create university buy-in, evidence of credible results must be produced.

At our university, long-term enrollment planning has been used for budget requests, master planning, water use permits, and transportation studies. Short-term enrollment planning has been used for semester enrollment projections, admissions policies, course planning, retention studies, and predicting the number of graduates. This university is in a high growth environment. As such, it is essential that planning models exist in order to manage the growth. However, modeling in this environment is also more difficult than modeling in a stable environment.

Rate-of-Growth Models

Hopkins and Massy make a distinction between short-term and long-term models due to the fact that "knowledge...is likely to become increasingly vague as one moves into the future" (Planning, 229). A set of rate parameters is needed to describe the growth. A growth rate is based on an incremental rate of change, for example, new students added to the number of returning students. A growth rate range may be established within which a policy or political decision specifies the exact growth rate to be used in the model.

Optimization Models

Optimization modeling, often called mathematical modeling, is a method of using mathematical expressions to solve problems (Quantitative, 254). The most common technique is called Linear Programming where a linear objective or criterion function is optimized subject to satisfying a set of linear constraints (restrictions or requirements). More details on mathematical modeling can be found in any Operations Research or Management Science textbook. This paper only addresses the basic premise and terminology used in optimization modeling and its direct relationship to the Solver add-in that is part of Microsoft' EXCEL®.

Solver is used to find a solution by changing values in decision cells that satisfy constraints and that minimize or maximize an objective function. Input values that are fixed numbers are called parameters. Other input values that are variable are called decision variables, or in Solver, changing cells. The quantity to be minimized or maximized is called the objective function, or in Solver, Set Cell. Constraints are restrictions on the solution. A constraint may be that a specified variable must be within a certain range, or must be an integer solution in a set of values returned for the decision variables. It is a feasible solution if all of the constraints have been satisfied. An optimal solution goes one step further in that not only does the solution satisfy all of the constraints, the objective function reaches a maximum or minimum value. A global optimal solution occurs when there is only one optimal solution. A locally optimal solution occurs when there are multiple optimal solutions, such as a function with peaks and valleys (Frontline, 25-27). If the problem is nonlinear, a local optimum solution will be found based on the starting values of the solution set. Multiple starting points should be used to test to see if the solution is a global optimal solution or a local optimal solution.

Markov Analysis Models

A Markov process is described as "studying the evolution of systems over...successive time periods where the state of the system in any particular time period cannot be determined with certainty. Rather transition probabilities are used to describe the manner in which the system makes transitions from one period to the next" (Anderson, 795). In other words, "markov analysis is a technique that deals with the probabilities of future occurrences by analyzing presently known probabilities" (Quantitative, 706).

The enrollment models assume that the probability of being in a particular state for the predictive period (year) is dependent on what happened only in the period immediately preceding the predictive period (Anderson, 795). For example, suppose we only have two states, enrolled or not enrolled. Reviewing the data, it is found that 60% of students who were enrolled in the previous Fall enrolled in the Spring and 40% did not. This indicates that 40% have transitioned from enrolled to not enrolled. So, this transition probability would be used to predict the number of students next Fall that would enroll in the Spring. Examples of the Markov process are shown in the short-term model and the graduate model below.

Long-Term Funding Enrollment Model

Model Overview

The first model discussed is a very broad, high-level model used to predict the number of FTEs for the university for a period of five years. This model predicts the number of FTEs by student level (lower-level, upper-level, and graduate) for the university. It then disperses the FTEs among the main and branch campuses. This 5-year prediction is forwarded to the governing board that determines the number of funded FTEs the university will receive. For a university in a growth mode, it is important to accurately predict the enrollment at branch campuses, as well as the main campus in order to capture the necessary dollars to fund that growth.

This model uses the actual FTEs from the previous year multiplied by the estimated growth rates by level to predict enrollment for the next five years. These FTEs are then distributed to the area campuses based on historical proportions and policy factors that address growth issues.

This approach has been revised and expanded to develop longer-term student headcount predictions for facilities planning to predict the main campus enrollment over the next 20 years.

Model Details and Data Requirements

A rate of growth approach was used because:

1. Only need predictions at the aggregate level, by campus

2. Since the branch campuses are new and we are adding campuses, there is not a lot of historical data to use for prediction

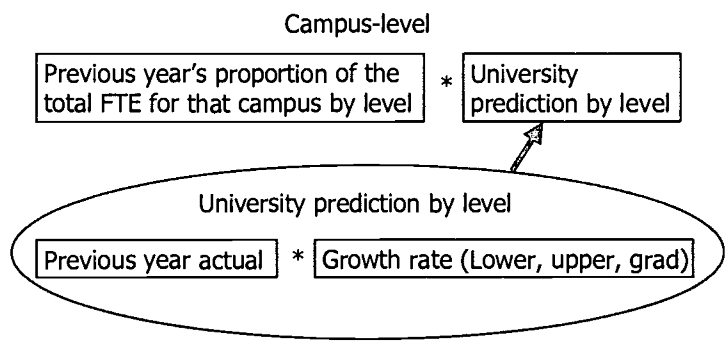
3. The university is in such a growth state, that using more than a year or two of historical data to project in the future can be unreliable. Furthermore, at some point in the next 10 years, this rapid growth will start to level out. Thus, regression models cannot reliably predict long-term in this environment.

Figure 1 below depicts the structure of this long-term model. The first stage of the model is the university-level prediction (in the circled area). A growth rate is estimated for the three levels (lower, upper, and graduate) for each of the prediction years. This growth rate is subjective based on a number of factors including the estimated number of new students that the undergraduate and graduate offices plan on admitting, external information on growth such as a decrease in high school graduates in the area three years out, and policy decisions handed down from administration. These growth rates are then applied to the previous year's enrollment for each of the three levels. Because predictions are made on the equivalent of "two year cohorts" (e.g., lower division students), growth rates need to be adjusted to reflect the combined effect on that two-year group of students. This is important when annual growth rates change as happens with predicted high school graduates.

The second stage of the model is to assign the university's enrollment to the main and branch campuses. As a default, the model assigns the predicted enrollment from stage one to the campuses based on the proportion of the total enrollment that campus had the previous year, by level. However, university policies may require adjustments to these allocations. For example, two of the branch campus allocations had to be adjusted to reflect special growth funding received by the state. There is a lag between establishment of the funds and growth in enrollment due to time required for new program development.

Figure 1

Long-term Model Structure



Model Results

For the two years that this model has been in existence, the university has been successful in capturing 96%-99% of the funding that it has requested, based on the model predictions. However, in 2001-2002, the university's actual enrollment exceeded the amount funded by 10.7%. Furthermore, the university has been able to capture a significant portion of the growth dollars for the state (approximately 20%). This was largely due to a decision to accept more new students than were anticipated at the time of the funding request. The prior year funding request sought to capture about half of the over enrollment funding in one year and the remainder over a five year period. Only part of that funding request was approved, reducing the base and continuing a significant over enrolled situation. Table 1 below provides the actual results of the long-term model.

Table 1

Long-Term Model Results

	2001-2002	2002-2003
Requested	20,840	23,599
Funded	20,630	22,645
Actual	22,836	N/A

Short-Term Enrollment Model

Model Overview

The second model provides a more focused set of short-term enrollment predictions. This model predicts not only FTEs, but also headcount and student credit hours by semester. This model predicts overall, as well as by level (eg. lower-level, upper-level) and classification (eg. freshmen, sophomores, graduates). The model is used to predict one to five years out. The results of the model are used to help administrators determine the number of new students (e.g., FTICs, transfer students, graduate students) to accept, estimate how over/under funded the university will be in a particular year, and to provide estimated Fall headcount for university relations.

This model uses historical data to predict headcount and student credit hours. However, in a growing student population, there will always be a lag in predicting the number of students who will return the following year. For this particular application at the given university, the model must take into consideration that some of the historical data collected included only funded students (such as headcount) while others included both funded and unfunded individuals (retention rates). Similarly, some of the historical data categorized students differently based on whether or not they had passed a proficiency exam. An optimization approach was used to help correct for these problems. A mathematical programming model was created using Excel's Solver tool. The objective was to minimize the sum of the squared differences between actual and predicted headcount by term (summer, fall, spring) and level (Freshmen, Sophomore, Junior, Senior, Unclassified, Graduate) for the previous year. The variables to be changed were the transition fractions by term and level. In other words, by changing the transition fractions in order to minimize the "error" in prediction the previous year, the gap due to growth can be diminished. It also corrects for any changes in transition rates from one category (Sophomore) to another (Junior) and also compensates for data from different classifications. This process resulted in a funded enrollment level (headcount) that was translated into student credit hours based on most recent behavior. The SCH predictions were then converted to predict FTEs.

Model Details and Data Requirements

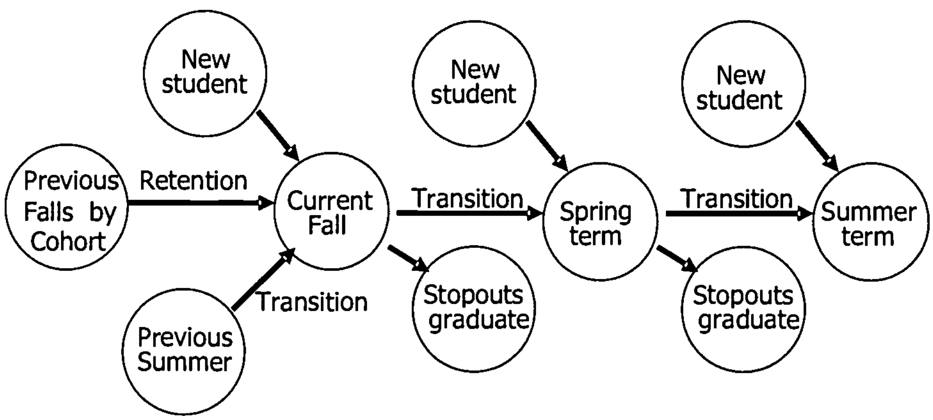
Unlike the other two models discussed in this paper, the short-term detailed model is an adaptation of a model that already existed in the Institutional Research department. The university administration requested a review of the model. An evaluation was completed comparing model output (using actual student inputs) with actual enrollment (HC and SCH). Although in general the model did pretty well, there was "no confidence" in the results. The model was not robust and there was no justification for many "adjustments" that were made from year to year.

Figure 2 below depicts the structure of the short-term model. The basic analysis is used to predict student headcount enrollments by semester. The Fall semester undergraduate prediction uses Fall cohorts with "cohort retention in class" factors (based on student file) plus new Fall students plus continuing Summer students. The Spring semester prediction uses a Fall to Spring transition rate from the previous year multiplied by Fall enrollments (modeled) by class plus new

Spring students. The Summer semester prediction uses a Spring to Summer transition rate from the previous year multiplied by previous Spring enrollments (data) by class plus new Summer students. Thus, using Markov-type analysis, the model uses historical data and transitions to predict future enrollment. Note that this is a more refined approach than the previous aggregate model that only projects annual FTEs. The short-term model uses historical undergraduate retention data to predict fall headcount and then predicts spring and summer headcounts using recent transition fractions. The graduate portion of the model only uses "continuation" fractions from the previous two years.

Figure 2

Short-term Model Structure



The predictions are made at different modeling levels. Headcount predictions are calculated by student classification (Freshman, Sophomore, Junior, Senior, Unclassified/Post Baccalaureate, and Graduate), undergraduate vs. graduate, and total enrollment. Student credit hour predictions are calculated by level (lower, upper, graduate) for each student classification listed above as well as aggregated by undergraduate, graduate, and total.

The data being used in the model had mixed definitions. Some of the data collected defined student classification by credit hours (e.g., 61-90 hours was equivalent to a Junior), while other data defined student classification with a CLAST adjustment (proficiency examination). Students were defined as a Sophomore even if they had completed over 60 hours but the CLAST requirement was not met. Retention was cohort based and used fall cohorts for ten years and tracked their progress by classification. Retention used the actual credit hour definition for classifications. However, the model predicts headcount and student credit hours using CLAST adjusted definitions with non-fundable students (e.g., state employees) eliminated.

The model uses historical data to predict headcount (HC). Student credit hours (SCH) are estimated from the predicted headcount based on previous behavior. FTE is estimated from student credit hours using 40 hours to convert undergraduate hours and 32 hours to convert graduate hours.

The problems discovered with the existing model were that there was no documentation or historical records. Due to employee turnover, formulas had been overwritten and all of the required historical data had not been updated. Furthermore, there were incomplete formulas and manual adjustments had been made to "improve" the prediction. An approach was needed that would generate appropriate adjustment factors that would be useful for prediction, independent of manual fine-tuning adjustments.

The basic conceptual structure of the model was retained. A new spreadsheet structure was developed to clearly define user inputs, historical data to be updated, and created clearly defined results pages. The data and formulas were updated. The unclassified headcount prediction was changed from a user input to a weighted formula using historical headcount.

In order to generate a systematic adjustment factor, a selection of "optimum" adjustment parameters for prediction of next yea's headcount were calculated. These parameters were applied as multiplicative factors rather than additive to help correct for the lag in growth and predictions of CLAST-adjusted headcount. In the former approach parameters were selected manually and applied as follows:

$c_iX_i + a_i$ [transition rate c_i, group size X_i, and adjustment parameter a_i].

In the new approach, adjustment parameters are selected so that the predicted headcount values for the previous year match the actual headcount values. An optimization model is used to minimize the squared deviations of the difference (predicted minus actual). This is implemented in Excel using Solver. The optimized parameters are applied as follows:

$a_ic_iX_i$ [transition rate c_i, group size X_i, and adjustment parameter a_i].

Figure 3 below shows the screen print for the optimization setup in Solver. In column L, functions have been created to sum the squared differences between predicted and actual. Since the functions are using quadratic equations, the model is nonlinear. Cell L20 sums the equations above for summer, fall, and spring. This cell is the objective function. Cells O3 through T5 are the changing cells. These cells are the "a_i"s discussed above which are the adjustment parameters used in the headcount formulas. No constraints were used in this model. Thus, the goal is to minimize the set cell (L20) by changing variable cells O3:T5. In the basic Solver installed with Excel, "linear" was deselected under options, or if using Premium Solver, Standard Nonlinear was selected for the methodology.

Figure 4 shows the results of the optimization procedure. If the model is developed correctly, Excel returns " Solver found a solution. All constraints and optimality conditions are satisfied." Select OK and the solution will be in cells O3:T5. Cell L20 will be minimized, but may not equal 0. In the model below, "unclassified" was included in the objective function but was not included in the cells to be changed. So column I did not change, but the "differences rows" in columns D through G and J are equal to 0. In other words, adjustment parameters were selected so that the predicted values for Freshmen, Sophomores, Juniors, Seniors, and Graduate for the previous year match the actual values for that year.

If you are not familiar with Solver, find someone on your campus that is. There are potential pitfalls that you will want to become aware of. One possibility is running into nonlinearity. If use a nonlinear objective function or nonlinear constraints the model can get complicated. Tolerance levels can also pose a problem. Be cautious when using Solver. It can be a very effective tool, but errors in model formulation can be difficult to spot.

Figure 3

Optimization Setup

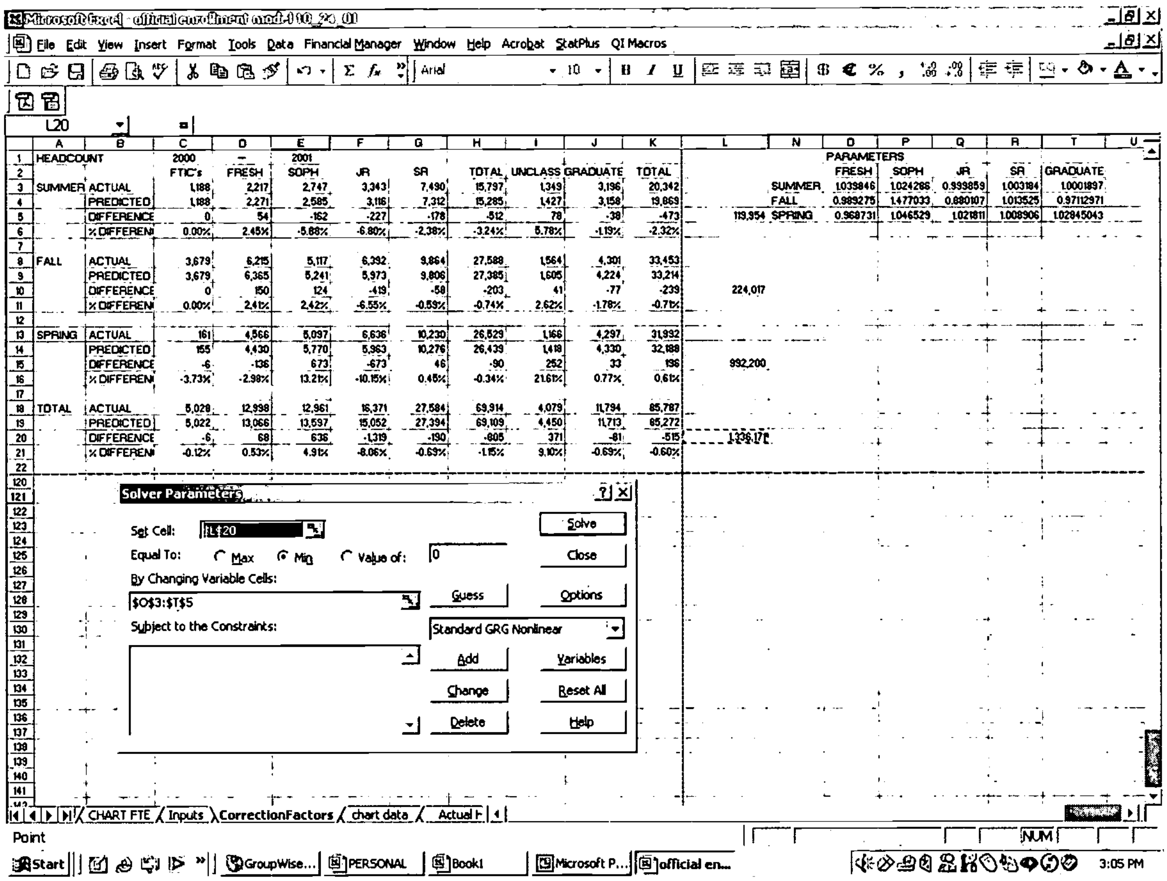
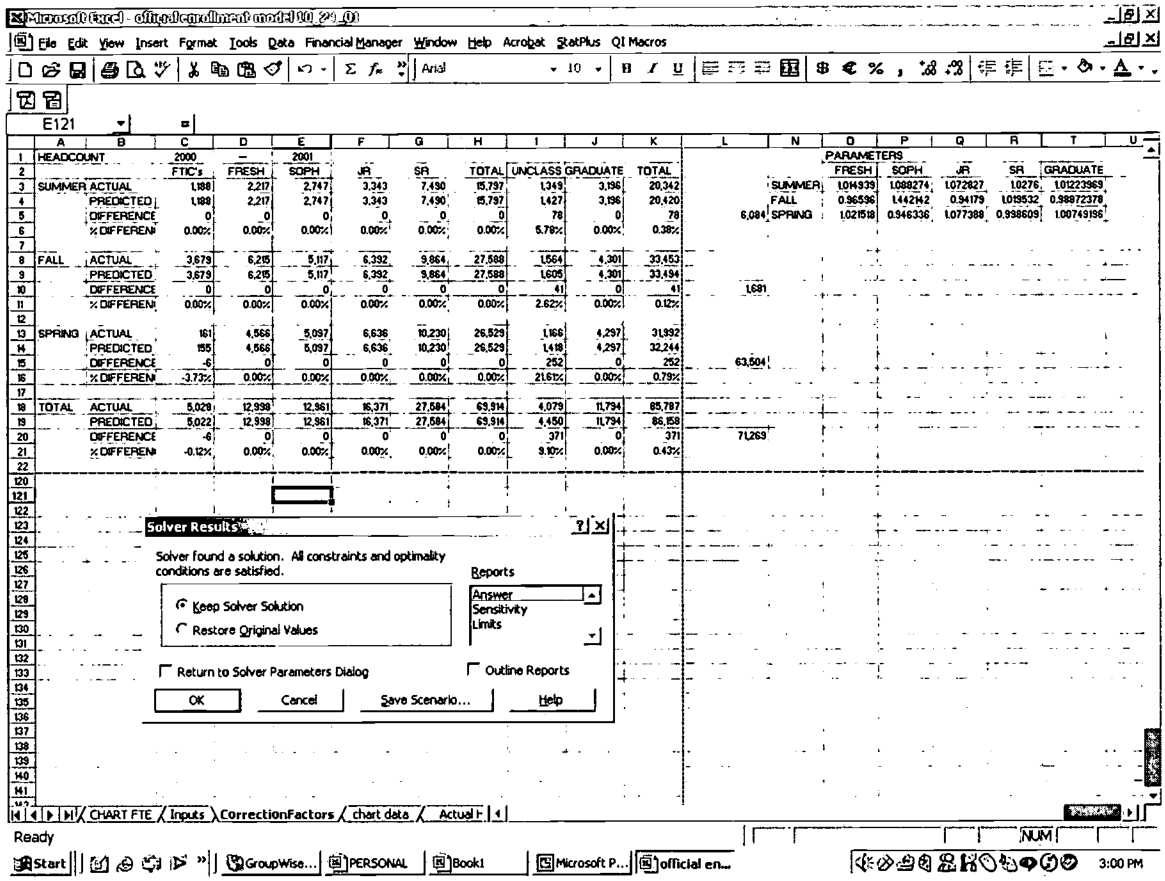


Figure 4

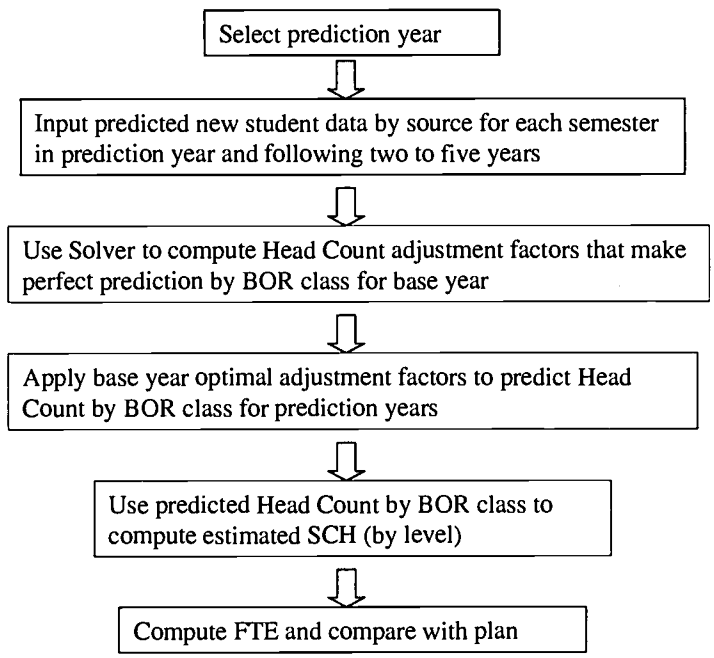
Optimization Results



The final model algorithm is shown in Figure 5 below. A prediction year is selected. The user inputs predicted new students by source for each semester to be predicted. Solver is used to compute the headcount adjustment parameters that make a perfect prediction by classification for the base year. These adjustment parameters are applied to all prediction years. The predicted headcounts are used to predict student credit hours that are also converted to FTEs. The predicted FTEs are compared to the plan to determine how much the actual enrollment will be over or under the amount funded.

Figure 5

Short-term Model Flow Chart



Model Results

The modeling approach was validated by comparing the predicted enrollment with the observed enrollment for the year following the year where the adjustment parameters were determined. The new model with adjustment factors predicted headcount better in five out of five years, error range (-0.38%, 0.54%). The new model predicted student credit hours better in three out of five years. In the other two years, the new model did almost as well as the old model, error range (-1.52%, 0.94%). In the outer-years, the headcount error range was (-.33%, 1.54%) and student credit hours error range was (-0.47%, 3.76%). The validation results indicate that the updated model is predicting very well in the short term, but begins to lose its accuracy in the out-years. The model has continued to predict fairly accurately in the short-term.

Table 2

Short-term Model Results-Predicted Headcount

DIFFERENCES (PREDICTED VALUE-ACTUAL VALUE)

	Old Model		New Model, No Correction Factors		New Model, Correction for Previous Year	
1995-1996	-1,168	-0.19%	-24,150	-3.93%	N/A	N/A
1996-1997	2,497	0.39%	-26,936	-4.16%	-4,146	-0.64%
1997-1998	-11,367	-1.69%	-20,922	-3.11%	4,638	0.69%
1998-1999	-4,414	-0.62%	-14,927	-2.09%	6,730	0.94%
1999-2000	-5,301	-0.71%	-22,226	-2.96%	2,007	0.27%
2000-2001	-23,712	-2.90%	-42,933	-5.26%	-12,401	-1.52%

Table 3

Short-term Model Results-Predicted Student Credit Hours

DIFFERENCES (PREDICTED VALUE-ACTUAL VALUE)

	Old Model		New Model, No Correction Factors		New Model, Correction for Previous Year	
1995-1996	272	0.41%	-937	-1.41%	N/A	N/A
1996-1997	826	1.19%	-1,892	-2.72%	-267	-0.38%
1997-1998	-746	-1.03%	-1,914	-2.65%	390	0.54%
1998-1999	-581	-0.76%	-1,672	-2.18%	308	0.40%
1999-2000	-96	-0.12%	-2,137	-2.66%	-78	-0.10%
2000-2001	-888	-1.04%	-2,865	-3.35%	-295	-0.34%

Table 4

Short-term Model Results-Predicted Headcount and SCH In Out-Years

HC DIFFERENCES (PREDICTED VALUE-ACTUAL VALUE)

	YEAR		YEAR+1		YEAR+2		YEAR+3	
1996-1997	-267	-0.38%						
1997-1998	390	0.54%	-237	-0.33%				
1998-1999			1,057	1.38%	399	0.52%		
1999-2000					1,240	1.54%	806	1.00%
2000-2001							296	0.35%

SCH DIFFERENCES (PREDICTED VALUE-ACTUAL VALUE)

	YEAR		YEAR+1		YEAR+2		YEAR+3	
1996-1997	-4,146	-0.64%						
1997-1998	4,638	0.69%	3,448	0.51%				
1998-1999			14,558	2.04%	17,649	2.47%		
1999-2000					17,229	2.29%	28,233	3.76%
2000-2001							-3,852	-0.47%

In general, the new model is robust. Formula cells are protected to prevent overwriting. The user has control over new student input and the duration of the historical period for estimating student credit hour conversion factors (use most recent year of data or a weighting scenario using additional years of historical data). The model is easily used and responsive to changed inputs. It is useful for "what if" analysis and easily incorporates actual data to permit revised predictions. The model is being considered for a web implementation. However, the way the model is constructed, predictions are more accurate for undergraduate because it is cohort based, than graduate. The graduate predictions are tied to current behavior so if there is a blip it is carried forward.

Graduate Enrollment Model

Model Overview

The third model is used to predict graduate enrollment by college. The university's office for graduate studies will use the model to estimate the number of post-baccalaureate, masters, and doctoral students expected each term by college. This will provide the colleges an opportunity to adjust their admission practices if necessary. This same method may be extended to predict graduate enrollment by program. If this is accomplished, the predictions could aid in course planning. This model is used only to predict one to two years out, but may be extended with future refinements.

This model is based on a Markov process approach. In this model the historical data is used to study enrollment patterns from one semester to another. This model is more refined than the previous models. Unlike the short-term model that uses retention for the fall undergraduate estimates and thereby has an annual basis for the model, the college-level graduate model captures a typical recent transition history and applies that to actual enrollments. Transition fractions are calculated to describe the percentage of students who fall into a particular state. The "states" that are used include the number of students who continue from one semester to another, those who skip one semester and come back, and those who drop out or graduate. Historical averages are used to provide an initial estimate of the number of new students who enter into a graduate program each term. This number of new students can be adjusted as needed to satisfy policy considerations.

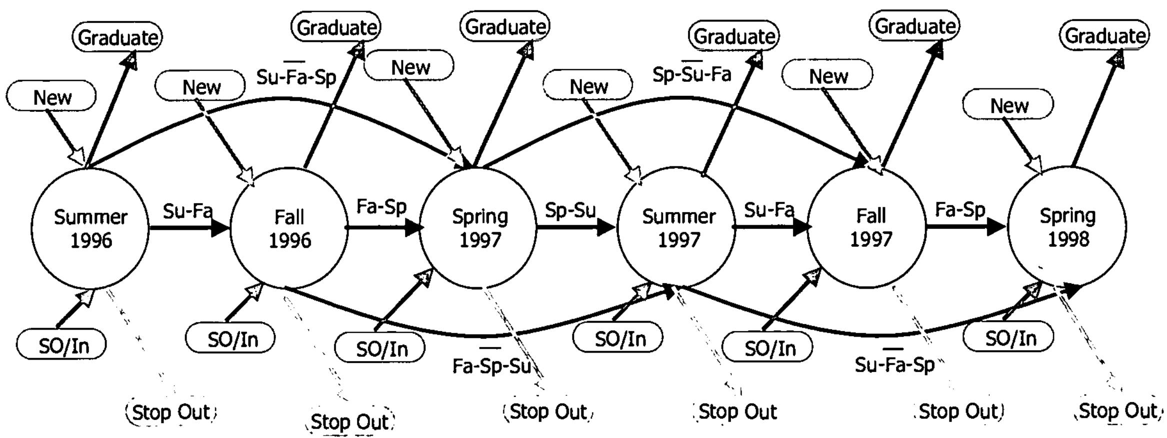
Model Details and Data Requirements

The only data used in this model is headcount data by college by semester. However, the difficulty in using this approach comes in when you have to be able to query the student database in order to track each individual's progress through the semesters using the states described below.

Figure 6 below depicts the Markov chain used in the graduate model.

Figure 6

Full graduate model

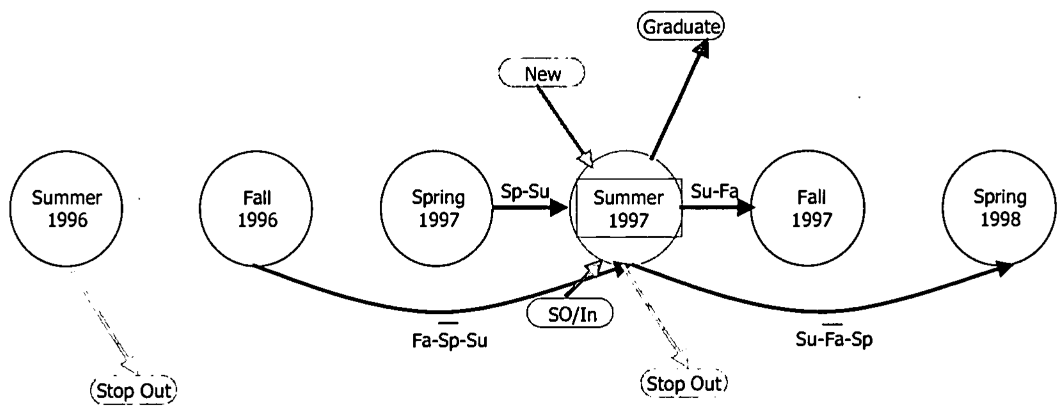


To explain this model, figure 7 shows the flow used to predict one semester, Summer 1997. Students enrolled in Summer 1997, have entered the semester in one of four states. The first is as a continuing student from the previous semester, Spring 1997 (Sp-Su). The second is as a student who skipped one semester, so was last enrolled two semesters ago, Fall 1996 (Fa-$\overline{\text{Sp}}$-Su). The third state is a new student, one who has never been enrolled as a graduate student at the university. The fourth state is a stop out-in. A student who was previously enrolled, for example in Summer 1996, but skipped more than one semester and then decided to reenroll would be classified as a stop out-in.

Students also leave a semester in one of four states. A student can continue the next semester, Fall 1997 (Su-Fa). A student can skip one semester and reenroll the next, Spring 1998 (Su-$\overline{\text{Fa}}$-Sp). A student can graduate. Or, a student can stop out for more than one semester.

Figure 7

Summer 1997 model



The prediction equation for any given semester contains two main elements: the number of students enrolled based on a given state and the transition fractions.

The number of students enrolled in a semester based on a given state can be defined as follows:

Students who continue from one semester to the next: $E_{a(yr)}=(S_{2(yr)} \rightarrow S_{3(yr)})$

Students who skip one semester: $E_{b(yr)}=(S_{1(yr)} \rightarrow S_{3(yr)})$

Stop Out/Ins: SO/IN. Estimate provided by Graduate Studies, but for model validation, actuals are used.

New Students: N. Estimate provided by Graduate Studies, but for model validation, actuals are used.

Two transition fractions are also used in the model and can be defined as follows:

Transition from one semester to the next: $T_{a(yr)}=(S_{2(yr)} \rightarrow S_{3(yr)})/\text{Total } S_{2(yr)}$

Transition from two semesters ago: $T_{b(yr)}=(S_{1(yr)} \rightarrow S_{3(yr)})/\text{Total } S_{1(yr)}$

The prediction equation can be generalized as:

$T_{a(yr-1)}{*}E_{a(yr)}+T_{b(yr-1)}{*}E_{b(yr)}+N+SO/IN$

Specific examples for each semester are shown below.

Summer 1996: $(Sp95 \rightarrow Su95)/Sp95{*}Sp96+(Fa94 \rightarrow Su95)/Fa94{*}Fa95+N_{Su96}+So/In_{Su96}$

Fall 1996: $(Su95 \rightarrow Fa95)/Su95{*}Su96+(Sp95 \rightarrow Fa95)/Sp95{*}Sp96+N_{Fa96}+So/In_{Fa96}$

Spring 1997: $(Fa95 \rightarrow Sp96)/Fa95{*}Fa96+(Su95 \rightarrow Sp96)/Su95{*}Su96+N_{Sp97}+So/In_{Sp97}$

Model Results

To validate the model, differences between predicted enrollment and actual headcount for each semester were calculated. The results showing these differences can be found in Table 5 below. In general, the model did fairly well. There were four predictions with an error rate above 10%, which can be seen in the shaded cells. Two of those, the A&S and Eng prediction for Spring 2000 can be attributed to the relocation of the computer science department from Arts & Sciences to Engineering. Overall, the average differences for each of the colleges ranged from –2.2% to 0.2%.

Table 5

Graduate Model Results-Prediction Year

Differences for Prediction Year

	A&S		Bus		Edu		Eng		H&PA	
Sum96	11	4.0%	-50	-8.8%	-19	-2.0%	14	2.6%	-18	-3.4%
Fa96	-23	-3.7%	-15	-2.0%	2	0.2%	32	4.1%	13	1.6%
Sp97	-32	-5.2%	-13	-1.9%	39	3.7%	43	5.9%	1	0.1%
1996-1997	**-44**	**-2.9%**	**-78**	**-3.9%**	**22**	**0.7%**	**89**	**4.4%**	**-4**	**-0.2%**
Sum97	-58	-16.8%	43	8.0%	-25	-2.6%	-36	-6.6%	-1	-0.1%
Fa97	10	1.5%	32	4.5%	-38	-3.6%	-57	-7.8%	-20	-2.3%
Sp98	23	3.4%	9	1.3%	-39	-3.6%	-85	-11.8%	37	4.4%
1997-1998	**-25**	**-1.4%**	**85**	**4.4%**	**-101**	**-3.3%**	**-177**	**-8.9%**	**16**	**0.7%**
Sum98	-22	-5.3%	2	0.4%	29	3.0%	1	0.2%	-37	-5.3%
Fa98	-19	-2.5%	-22	-2.9%	-25	-2.1%	-3	-0.4%	47	5.2%
Sp99	3	0.5%	-8	-1.2%	-22	-2.0%	43	6.3%	-12	-1.3%
1998-1999	**-38**	**-2.0%**	**-27**	**-1.4%**	**-18**	**-0.5%**	**41**	**2.1%**	**-2**	**-0.1%**
Sum99	-43	-8.3%	11	2.1%	20	2.0%	-13	-2.5%	-1	-0.2%
Fa99	13	1.6%	-5	-0.8%	64	5.8%	-14	-1.9%	-30	-3.1%
Sp00	121	18.8%	-5	-0.7%	21	1.8%	-155	-17.9%	-5	-0.5%
1999-2000	**91**	**4.6%**	**1**	**0.0%**	**105**	**3.3%**	**-182**	**-8.5%**	**-36**	**-1.4%**
Sum00	-13	-2.8%	6	1.2%	-17	-1.7%	-5	-0.8%	13	1.8%
Fa00	-14	-1.9%	20	2.9%	-37	-3.2%	-6	-0.7%	3	0.3%
2000-2001	**-26**	**-2.8%**	**26**	**1.2%**	**-53**	**-1.7%**	**-11**	**-0.8%**	**16**	**1.8%**
Avg Error	-3	-1.2%	0	0.2%	-3	-0.3%	-17	-2.2%	-1	-0.2%

The model was also taken out one year past the prediction year. The model used the results from the prediction year to estimate the following year. Differences between prediction and actual for year +1 are shown below in Table 6. The model did not do as well when predicting another year out. Overall, except for Engineering (error –7.6%), the average error rate ranged from –1.5% to 0.8%. However, when looking at semesters individually, many more had error rates above 10%.

Overall, the approach appears to be effective. However, this model is still in its preliminary stages and will require some fine-tuning of the model, particularly in the College of Engineering.

Table 6

Graduate Model Results-Prediction Year +1

Differences for Prediction Year+1

	A&S		Bus		Edu		Eng		H&PA	
Sum97	-60	-17.4%	-17	-3.1%	-15	-1.6%	4	0.7%	-22	-3.6%
Fa97	-36	-5.4%	8	1.1%	-10	-0.9%	2	0.3%	-4	-0.5%
Sp98	-31	-4.6%	-10	-1.5%	22	2.0%	-22	-3.1%	40	4.7%
1997-1998	**-128**	**-7.5%**	**-19**	**-1.0%**	**-4**	**-0.1%**	**-16**	**-0.8%**	**14**	**0.6%**
Sum98	-77	-18.2%	53	9.7%	-26	-2.6%	-83	-15.9%	-15	-2.1%
Fa98	10	1.4%	18	2.4%	-90	-7.8%	-111	-14.8%	48	5.3%
Sp99	44	6.0%	7	1.0%	-85	-7.5%	-83	-12.1%	47	5.3%
1998-1999	**-24**	**-1.3%**	**78**	**4.0%**	**-201**	**-6.1%**	**-277**	**-14.2%**	**80**	**3.2%**
Sum99	-65	-12.6%	7	1.4%	33	3.4%	15	3.0%	-47	-6.4%
Fa99	-5	-0.6%	-32	-4.4%	23	2.1%	12	1.6%	12	1.3%
Sp00	127	19.7%	-16	-2.5%	-15	-1.3%	-87	-10.1%	-21	-2.3%
1999-2000	**56**	**2.9%**	**-41**	**-2.2%**	**42**	**1.3%**	**-60**	**-2.8%**	**-56**	**-2.1%**
Sum00	16	3.4%	13	2.7%	20	2.0%	-117	-18.6%	8	1.1%
Fa00	86	11.7%	11	1.7%	43	3.8%	-129	-14.5%	-31	-3.0%
2000-2001	**102**	**12.4%**	**24**	**0.7%**	**64**	**1.9%**	**-246**	**-18.2%**	**-23**	**0.2%**
Avg Error	1	-1.5%	4	0.8%	-9	-0.8%	-54	-7.6%	1	0.0%

Conclusions

The three different models clearly serve different purposes. The long-term aggregate model was more policy oriented and required significantly less data. The short-term, university-level model required much more detailed data, although at an aggregate level. The model had more complex formulas and methods. The graduate "detailed" model required more specific data at the individual level. More complex methods were used and the results were at a more detailed level. All three types of enrollment models are needed to support university operations.

There is additional work ahead with respect to enrollment. An immediate need is to improve the graduate predictions in the short-term model. The detailed graduate model holds promise for using aggregated results (across colleges) to provide the structure for the total graduate enrollment prediction. In addition we need to provide better linkage between the short-term model and the aggregate enrollment model. The current models used externally generated estimates of new student input. A needed enhancement is to create models to use external data to predict "inputs" for both the short-term university-level model and the graduate model.

The three models discussed above were developed for a large, metropolitan research university in a growth mode. The models serve different needs of the several constituents and accommodate their different planning horizons. In this case, "one size does not fit all"—the different management needs required different types of models.

References

Anderson, David R., Dennis J. Sweeney, and Thomas A. Williams. (2000). An introduction to management science: quantitative approaches to decision making. 9th ed. Cincinnati: South-Western College Publishing.

Hopkins, David S. P. and William F. Massy. (1981). Planning models for colleges and universities. Stanford: Stanford University Press.

Render, Barry and Ralph M. Stair, Jr. (2000). Quantitative analysis for management. 7th ed. Upper Saddle River: Prentice Hall.

Solver user's guide. (1999). Incline Village: Frontline Systems, Inc.

NOTICE

Reproduction Basis

- [x] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").