PREDICTING TOTAL STUDENT CREDIT HOURS PRODUCTION BY COHORT STRATIFICATION

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

COLLEGE OF ARTS AND SCIENCES

BY

CAROLINA DOMINGUEZ SHEEDER

DENTON, TEXAS

AUGUST 2013

DEDICATION

For my husband, Ward, and my children, Sophia and Christian,

thank you for your patience and love.

ACKNOWLEDGMENTS

I would like to thank Dr. Mark Hamner for all his time, dedication, and for making this thesis a learning process in many areas. I am grateful to Dr. David Marshall and Dr. Ann Wheeler, who served as members of my committee. I am also grateful to Dr. Hamner for the wonderful experience I had in his classes. As a teacher he inspired me to be a better teacher, and as a student he shed light on probability concepts that had been in the dark since I “learned” them in College. I am also grateful to Dr. Wheeler for teaching Symbolic Logic the way she did, I really enjoyed that class. She helped us develop a good understanding of proofs and encouraged us in taking challenges.

Finally I would like to thank my children for their patience, their unconditional love, and for trying to understand when I could not play with them because I had to work on my “book”. And last, but not least, I want to thank my husband Ward for taking over cleaning the kitchen and making dinner and the lunches every night, and taking care of the bedtime routine in many occasions as well. I am so thankful for his support and his love.

ABSTRACT

CAROLINA DOMINGUEZ SHEEDER

PREDICTING TOTAL STUDENT CREDIT HOURS PRODUCTION
BY COHORT STRATIFICATION

AUGUST 2013

The objective of this study is to develop predictive models of total student credit hours (SCH) prior to the fall semester of interest by using preregistration data from Texas Woman's University (TWU). We developed two different approaches to predict SCH for undergraduate and SCH for graduate students separately. Our first approach is based on the patterns of weekly counts of SCH observed over time. For our second approach, we developed a model that relies on an average of SCH and a total headcount. This research presents a self-contained procedure to predict headcount and includes a criterion to select a prediction model for the average of SCH. After explaining the development of each of our SCH prediction models, we compare the results and discuss their strengths and weaknesses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Texas Higher Education Coordinating Board is charged with developing recommendations for improvements regarding state funded institutions of higher education to the governor and state legislatures. In carrying out its duties, the Coordinating Board reviews and recommends changes in formula funding that provide the allocation of state funds to public institutions and ensures an effective and efficient system of higher education. They do this by controlling costly duplication of academic programs and unnecessary construction projects. The report submitted in January 2011to the 82nd Texas Legislature by this Coordinating Board explains how nearly 54 percent of state appropriations for general academic institutions are allocated via two funding formulas and two supplements: the Instruction and Operations Formula, the Infrastructure Formula, the Teaching Experience Supplement, and the Small Institution Supplement (Legislative Primer, 2011).

For an institution of higher education to make appropriate budgeting decisions for each academic year, it is important to understand the underlying mechanism by which formula funding is provided by the state. For example, the Instruction and Operations

Formula funding mechanism created by the state relies on semester credit hours (SCH) production generated by institutions. Thus, budget planning at an institution of higher education can be greatly facilitated, if the institution has a viable way of predicting the semester credit hours they will have prior to their fiscal year. Interestingly, however, the literature reveals few articles dealing with semester credited hour projections. Winona State University (Ed Callahan, 2011), the University of Florida ("Overview of a Detailed Enrollment Prediction Model", 2011), and the University of Baltimore ("Headcount and Student Credit Hour Projections in support of the Master Facility Plan 2008-2018") address SCH projection using current enrollment numbers, predicted retention, advancement rates by class, and credit hour averages but do not expand on the predictive accuracy of their modeling technique. On the other hand, the literature reveals many articles about enrollment projection models (Guo, 2002; Nandeshwar and Chaudhari, 2009; Armstrong and Wenckowski, 1981). Although enrollment is positively correlated with semester credit hour production, the aforementioned models do not make such a connection. Other aspects of enrollment such as influential factors that increase or decrease the enrollment or retention (Cameron and McLaughlin (2008); Gao, Hughes, O'Rear, and Fendley, (2002); Luo, Williams, and Vieweg (2007)) are also commonly explored in the literature.

The objective of this study is to develop predictive models of total student credit hours (SCH) prior to the fall semester of interest by using preregistration data from Texas Woman's University (TWU). To predict total SCH at any time *t*, where *t* represents some point in time within 23 weeks prior to the start of the semester of interest, we will use prior fall historical patterns of preregistration. The assumption is that the historical preregistration data at time t relative to the respective prior fall semesters will provide relevant patterns for predicting total SCH for the semester of interest. In this research, we will test the predictive accuracy of our models by using cross-validation.

The idea of a preregistered student and preregistered SCH begins with understanding that the student accesses the TWU website to preregister for a class or classes within a period of 23 weeks before the beginning of the semester of interest or the official census day, which is referred to as the 12th day. If a student then completes their payment and is enrolled on the 12th day, the student now becomes part of the official headcount. In addition the number of SCH the student has on the 12th day become part of the official total SCH on 12th day. The total SCH on 12th day is

$$T = \sum_{k \in \mathrm{P}} i_k x_k \tag{1.1}$$

where x_k = SCH of individual $k \in \mathbf{P} = \{1, 2, \ldots, N\}$ such that $\mathbf{P}$ represents the index of individuals who preregister for the semester of interest. The magnitude of the set $\mathbf{P}$ is represented by the following notation $|\mathbf{P}| = N$, where $|\cdot|$ is the magnitude (i.e., the number of elements) of a set. In this case, N is the total number of preregistered students during the 23 week period prior to the census day (i.e., 12th day). Furthermore, associated with each unit $k \in \mathbf{P}$ is the value i_k defined as

$$i_k = \begin{cases} 0 & \text{if a preregistered student does not complete the registration process} \\ 1 & \text{if a preregistered student completes the registration process} \end{cases},$$

where the registration process is assumed complete if a student completes payment for their SCH.

When we use Equation 1.1 at any time *t* during the 23 week time period prior to the census day, we will have observed only those students who preregistered up to that time period, but there are also preregistered students we expect to observe after time *t*. Notationally, we will let t', which is read as *t* complement, represent the time period after *t*. For this study, we will partition the time interval into weekly periods such that each time *t* represents one of the 23 weeks prior to the semester of interest. Using 23 weeks prior to the start of any fall semester has the predictions for the SCH starting around the month of April. To visualize the partition of the time interval by making a prediction at time *t*, see Figure 1.1 below.

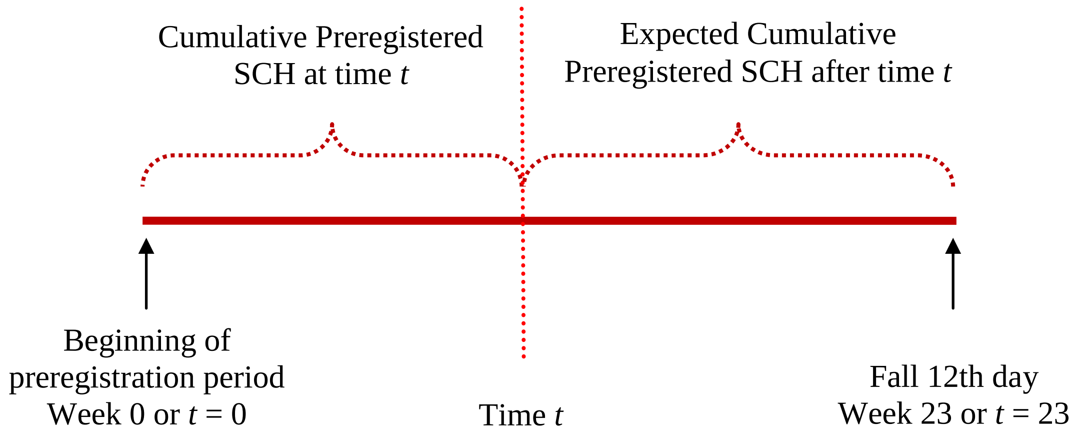

Figure 1.1: Time partition for cumulative parallel patterns

At any time t, prior to the 12th day of the fall semester of interest, we only observe part of the total number of preregistered students. Let n^t represent the total number of preregistered students at time t, then there are still $n^{t'} = N - n^t$ students that will preregistered after time t. Thus,

$$\mathbf{P} = \mathbf{P}_t \cup \mathbf{P}_{t'} \qquad (1.2)$$

where $\mathbf{P}_t \cap \mathbf{P}_{t'} = \varnothing$, $\mathbf{P}_t$ is the set of labels for the n^t students that preregistered before time t, and $\mathbf{P}_{t'}$ is the set of labels for the $n^{t'}$ students that preregistered after time t. Thus, grouping the right hand components of Equation 1.1 according to Equation 1.2, we can rewrite Equation 1.1 as

$$T = T_t + T_{t'} \ . \qquad (1.3)$$

The typical undergraduate SCH load will differ from the typical graduate SCH load since a full-time undergraduate student takes at least 12 hours, whereas a full-time graduate student takes at least 9 hours. Considering that the requirement for a full-time graduate student differs from the full-time requirement of an undergraduate student, we will develop separate predictive models for undergraduate SCH versus graduate SCH. It is also important to note that the funding of an undergraduate SCH differs from the funding of a graduate SCH. Accordingly, we will let *U* represent the set of undergraduate students that preregister prior to the semester of interest and *G* represents the set of graduate students that preregister prior to the semester of interest.

Considering that predictions are made at some time *t*, we can define

$$\mathbf{P}_t = \mathbf{P}_{U^t} \cup \mathbf{P}_{G^t} \text{ and } \mathbf{P}_{t'} = \mathbf{P}_{U^{t'}} \cup \mathbf{P}_{G^{t'}}$$

Thus, we can rewrite Equation 1.2 as

$$\begin{aligned} \mathbf{P} &= \mathbf{P}_t \cup \mathbf{P}_{t'} \\ &= \left(\mathbf{P}_{U^t} \cup \mathbf{P}_{G^t}\right) \cup (\mathbf{P}_{U^{t'}} \cup \mathbf{P}_{G^{t'}}) \\ &= \left(\mathbf{P}_{U^t} \cup \mathbf{P}_{U^{t'}}\right) \cup (\mathbf{P}_{G^t} \cup \mathbf{P}_{G^{t'}}) \end{aligned} \tag{1.4}$$

Now, the components of Equations 1.3 are defined as $\mathbf{P} = \mathbf{P}_{U^t} \cup \mathbf{P}_{U^{t'}}$ and $\mathbf{P} = \mathbf{P}_{G^t} \cup \mathbf{P}_{G^{t'}}$ for each time *t* during the 23-week preregistration period. Accordingly, Equation 1.1 can be rewritten as

$$T = T_U + T_G \tag{1.5}$$

where

$$T_U = \sum_{k \in \mathbf{P}_{U^t}} i_k x_k + \sum_{k \in \mathbf{P}_{U^{t'}}} i_k x_k \tag{1.6}$$

and

$$T_G = \sum_{k \in \mathbf{P}_{G^t}} i_k x_k + \sum_{k \in \mathbf{P}_{G^{t'}}} i_k x_k \tag{1.7}$$

In Chapter-2 of the Thesis, we will expand our literature review and discuss the different approaches to predicting Equation 1.1.

In Chapter-3 and Chapter-4, we will develop two different approaches to predict Equation 1.6 and Equation 1.7. Our first approach to predict Total SCH uses an enrollment model framework presented by Dr. Mark Hamner and Preet Ahluwalia in their presentation at the 2007 TAIR Conference "Predicting Real-Time Percent Enrollment Increase". In this presentation, the objective was to predict student enrollment at time *t* using applicant data, where *t* is defined as the time when the prediction is made, and t'(*t* complement) is defined as the time between the projection and the final actual count. They found that the graphs of weekly counts of applicants year to year have the same slope, so they could assume that the counts of applicants after time *t* would behave similarly to the counts of applicants before time *t*. Our first modeling technique will model these types of patterns but not for headcount. Instead, this technique will use SCH cumulative patterns. This model was the first of its kind for enrollment then and has not been used to predict SCH until now. We will then compare the results to our second approach.

Our second approach to predict Total SCH will be our own modified version of the model that the University of Baltimore presents in their paper ("Headcount and Student Credit Hour Projections in support of the Master Facility Plan 2008-2018"). This model predicts student credit hours through 2018 by modeling the weighted average of credit hours and multiplying its output by a headcount predicted through the model created by Maryland Higher Education Commission. In our second approach to modeling Equations 1.6 and 1.7, we will use our own version of their approach to predict SCH. Their technique requires a headcount total, but they do not address headcount prediction. Instead, they borrowed enrollment projections of total headcount. In our research, we will develop a self-contained procedure for predicting headcount. Since the weights of the weighted average were not used explicitly in their model, we will use a regular average instead of a weighted average of credit hours for TWU data and include a criterion to select a prediction model for such average.

After explaining the development of each of our SCH prediction models, we will compare the results and discuss their strengths and weaknesses.

CHAPTER II

LITERATURE REVIEW

The literature shows that universities have been actively pursuing models that can predict enrollment and retention. However, in the state of Texas, having a model that can predict semester credit hours will be particularly helpful to administrator's trying to create an accurate budget. One of the biggest benefits of having a good model to predict enrollment is the ability to plan and administrate resources for the upcoming semester and anticipate future needs.

The objective of our research is to construct a model to predict the total count of credit hours. In the process of constructing this model, we will consider the advantages and disadvantages of models demonstrated in previous studies. There is a long list of works dedicated to enrollment modeling, and in this chapter we intend to discuss some these works and make a comparison between them and the present research paper. We will also include references that discuss the retention of students. This topic is often associated with headcount modeling, and it is a precedent for the topic of this paper.

We realize that the average amount of credit hours per semester is different between undergraduates and graduate students. Thus, to improve our prediction, we will only specify strata of students as undergraduates and graduate students; and we will not distinguish between new students, transfers, or continuing students; nor do we differentiate between part-time and full-time students.

Stratification of the data is a common factor among several sources that discuss models to project enrollment or retention. The usual stratification consists in students that are in college for the first time (First Time In College or FTIC), transfers, continuing; and between undergraduates and graduates. The reason behind this separation is that each of these groups behaves in different ways as enrollment and retention is concerned.

Gao, Hughes, O'Rear, and Fendley (2002) approached some of these differences in their article. In this paper, they used Structural Equation Models to identify factors linked to high graduation or retention rates distinguishing between native students (first-time freshmen) and transferred students. They concluded that the number of hours transferred in is a strong predictor of transfer student graduation, and first year performance is obviously linked to graduation and retention rates.

Colleges with higher retention rates for first-year students tend to have higher graduation rates. For this reason, there is a large amount of literature about mathematical models to predict the behavior of freshman students. Cameron and McLaughlin (2008) used decision trees to recognize primary influences in the success of freshmen transfer students. They defined success as retention in their article "Modeling Success of

Freshmen Transfer Students." They used decision trees to analyze academic characteristics of this cohort and followed the bracket with higher retention rates. Students that returned after the first quarter were first divided into the group that passed 47% of their classes or more and the complement of this group. The subgroup that passed 47% of their classes or more had a higher retention rate and was then split between full time and part time students. The full time subgroup had a higher retention rate and was then divided between students that transferred in with more than 12 hours and its complement. The group with greater number of transferred in hours had the highest retention rate. They ended the tree by analyzing the ACT English scores of this last group and concluded that the subgroup who passed 47% of their classes or more, considered full time, transferred in with more than 12 hours, and an ACT English score greater than 29 was the subgroup with the highest retention rate.

Luo, Williams, Vieweg (2007) also wrote about first year retention in their article. They used sequential sets of logistic regression analyses on blocks of variables applied to student groups of different transfer status to analyze patterns of interactive factors that influence transfer student's first-year retention.

The studies mentioned previously were focused on explaining and predicting retention rates of students that are FTIC. These studies and our study share the same motivation, which is the anticipation of resources needed by a university to efficiently serve their students. The afore mentioned studies addressed this objective by projecting student headcount only, while this research is centered on the projection of student credit hours using two different models.

Our first approach, which will be referred as Model-1, follows a similar methodology to the enrollment model presented by Dr. Mark Hamner and Preet Ahluwalia in their presentation at the 2007 TAIR Conference “Predicting Real-Time Percent Enrollment Increase”. In their presentation, they defined Total Enrollment is equal to the Enrollment as of time *t* plus the Enrollment after time *t*, where time *t* is the time of prediction. This model was the first of its kind to predict headcount. In this research, we will borrow their enrollment model framework to predict total SCH in a particular semester of interest.

A different framework to predict total SCH is illustrated by North Carolina General Assembly in their Final Report to the Joint Legislative Program Evaluation Oversight Committee (2010). In this report, they exemplified the model to predict SCH used by the University of North Carolina (UNC). This article, besides discussing a different approach to model total SCH, is also a reference for the impact an accurate or inaccurate SCH projection has on the funding. In this report, the North Carolina General Assembly also discussed the relevance of the total SCH in the context of their funding formula.

Their funding formula to calculate enrollment growth has three components. First, the number of credit hours that will be taken at each institution is projected based on enrollment data from the fall semester. Second, the number of additional instructors needed to serve the projected enrollment is estimated based on various ratios of credit hours per instructor. Finally, the funds needed to cover the additional salaries, academic costs; library services and general institutional support are calculated. The SCH model works as follows. Credit hours offered at each university are classified into one of 12 categories reflective of the area of instruction (four possible categories) and level of instruction (undergraduate, master's and doctoral). Projections for the number of credit hours that will be taken in each category are then developed. These are estimated for each cell in the matrix and are expressed as an increase or decrease in the number of student credit hours from the prior year.

Below is a hypothetical example of the projected SCH for a campus presented by North Carolina General Assembly, 2010. In this example, the campus estimates 4,700 additional SCH for the next academic year.

Hypothetical Example of the Projected SCH for a Campus of UNC

	Instructional Level		
Instructional Category	**Undergraduate**	**Master's**	**Doctoral**
Category I	1,000	200	100
Category II	1,000	200	-100
Category III	1,000	200	50
Category IV	1,000	50	0

	Instructional Level		
	Undergraduate	**Master's**	**Doctoral**
Total by level	**4,000**	**650**	**50**
		Institution Total	4,700

The second approach we will use to build a predictive model for total count of SCH, which will be referred to as Model-2, combines a model that predicts headcount and a model that predicts the average of SCH. The headcount projection and average SCH projections are multiplied by each other to obtain an estimate of total SCH. Ed Callahan (2011) of Winona State University presented a projection model for Student-Credit Hour load in his article. This model also used an estimated average credit load and an estimated headcount to predict total Student-Credit Hour load. The estimated count uses current enrollment numbers, predicted retention, and advancement rates by class (freshmen, sophomore, juniors, and seniors).

The University of Central Florida (UCF) also uses a headcount model to predict student credit hours. UCF posted on their website the "Overview of a Detailed Enrollment Prediction Model" that estimates headcount (HC) and student credit hours. The headcount model takes the Spring and Summer enrollment and multiplies it by the previous year's semester transition fraction, built with retention or returning rates for undergraduates and graduates from the previous ten years and two years, respectively. They then added the estimated number of new students. Because the retention and transition parameters can vary, the model uses a set of multiplicative adjustment parameters computed so that the model, based on the previous year's data, "fits" the actual enrollment from the previous year perfectly. The resulting model with the adjustment parameters is then used with current year enrollment and the expected new students to predict the following year enrollment by classification. The predicted headcounts are used to estimate the fundable student credit hours by semester and the annual SCH are used to estimate the fundable full-time equivalent students by level.

The idea of using an average-SCH was also used by Campbell and Doan (1982) in their article. They did two regressions to predict total SCH, one between headcount and the number of credit hours; and the second between headcount and student average load. The difference in the results between the two regressions was due to the fact that the average (credit hours/total of students) decreases as the total headcount increases because the number of credit hours has an upper limit. They later found out that the average of the results from the two predictions resulted in a value that was closer to the actual observed total SCH.

The modeling process we will use to calculate the average of SCH in Chapter-2 was presented by the Office of the Provost Institutional Research at the University of Baltimore in their draft "Headcount and Student Credit Hour Projections in support of the Master Facility Plan 2008-2018." The model that the University of Baltimore presents in their paper predicts student credit hours through 2018 by modeling the weighted average of credit hours and multiplying its output by a headcount borrowed from the Maryland Higher Education Commission's Enrolment projections. The article discusses the importance and relevance of both, and Enrollment Projection model and the Projection of student credit hour loads. Their projection of student credit hour separates students by colleges and level. To model the weighted average of each cohort, they used polynomial, exponential, logarithmic regressions, and the curve of the cumulative distribution of a Weibull variable to fit their historic data and a reasonable prediction of a weighted average of SCH for future years, considering a limit on credit hour loads based on historically high trends or values. However, the University of Baltimore did not provide the criterion used to select the models used to fit the pattern of historic data of weighted averages of SCH and to predict weighted averages of SCH in future years. In this study, we will discuss a criterion that will be used to select the prediction models for the averages of SCH in future years. Once we obtain a predicted average of SCH, we will develop a prediction model for headcount. This headcount model is based on the enrollment model presented by Dr. Mark Hamner and Preet Ahluwalia in their presentation at the 2007 TAIR Conference and also used as the framework in the approach under our Model-1. The combination of the predicted average SCH and the predicted total headcount will result in our unique version of the University of Baltimore's SCH prediction model.

Data mining has been a popular approach to develop headcount or enrollment projection models by analyzing data from different perspectives and summarizing it into useful information. Nandeshwar and Chaudhari (2009) used data mining to build models to predict enrollment using the student admissions data, evaluate the models using cross-validation, win-loss tables and quartile charts. These authors also discuss previous applications of data mining such as Enrollment management, Graduation, Academic performance, and Retention.

Both approaches in our study will be limited to one input variable, time. In the attempt to offer an accurate representation of the population being studied, other models used many characteristics of freshman students, such as SAT scores, GPA, age, and many other variables that seem to influence the decision of whether or not to enroll, continuing in the same degree, or changing colleges. In the work of Guo (2002), three different enrollment projection models and their application in six Community Colleges are compared. Guo also lists a set of factors that need to be considered in forecasting, such as time frame, cost, the availability of data, data patterns, and the ease of operation and understanding. The conclusion was that a complex model may not be necessarily better than a simpler model.

Armstrong and Wenckowski Nunley (1981) also compared two Enrollment Projection models. One model was based on Curve fitting, the other one was based on

Yield from population components. They emphasized the importance of direct involvement of key administrators in discussing the reasonableness of the assumptions associated with the projections.

Another way to determine the reasonableness of a prediction model is to test the predictive accuracy of such a model. To test Model-1 and Model-2 developed in Chapter-3 and Chapter-4 of this study, we will predict the total count of SCH in 2011 using preregistration data from 2008 through 2010 and compare the predicted count of SCH with the actual count obtained from the actual fall data of 2011. Then, we will predict the total count of SCH in 2012 using preregistration data from 2009 through 2011 and compare the predicted count of SCH with the actual count obtained from the actual fall data of 2012. A similar test is used by Tsui, Murdock, and Mayer (1997). They examined whether the use of trend analysis combined with analysis of persistence variables can be used to establish a model to forecast the first-year persistence of college freshmen. This paper uses linear regression, hypothesis testing, and confidence intervals. A linear model was created using data on 2,603 first-time freshmen at a moderate-sized comprehensive university from fall 1989 through fall 1993. To test the accuracy of the model, they used linear regression for both scales of percent and a number of campus residents' first-year persistence from fall 1989 to fall 1993. They then used hypothesis testing at 0.01 level, and both forecast equations were statistically significant. The forecast equations were also tested by predicting the first-year persistence rate for freshmen newly enrolled in fall

1994 and both models came to have the same significance level but different accuracy of forecast, both within 4%. The study concluded that trend analysis is an effective method to discover a relationship between students' retention and categorical factors. Furthermore, the methodology of combining trend analysis and significant persistent variables provides a potentially more accurate method to predict continuous enrollment.

CHAPTER III

NOTATION AND PREDICTIVE MODEL 1

In this chapter, we will introduce additional notation and extend on the notation presented in Chapter-1. This notation is necessary to present the first modeling method, which we will refer to as Model-1, which we used to predict the total count of SCH on 12th day defined in Equation 1.1.

The first model we will develop to predict T is similar to the model framework presented by Hamner and Ahluwalia (2007). In their presentation, they showed that the graphs of weekly counts of applicants year to year have the same slope, which is visually represented by parallel lines over time. Given this pattern, they could assume that the counts of applicants after time t would behave similarly to the counts of applicants before time t. Our Model-1 is based on these types of patterns but not for headcount. Instead, this technique will use SCH cumulative patterns over time.

This modeling approach requires a partition of the time interval into weekly periods, such that each time period t_w is time in terms of the number of weeks from the beginning of the prediction period, which starts 23 weeks prior to the semester of interest. At the end of each time period t_w, there is an aggregate total of observed SCH, T^{t_w}.

However, since preregistration is still ongoing, we know there are students who will preregister after time t_w. Define $t_{w'}$ to be the time after time t_w, such that $t_w + t_{w'} = 23$. Accordingly, the cumulative SCH observed after time t_w is represented by $T^{t_{w'}}$. Using this notation, we can rewrite Equation 1.3 as

$$T = T^{t_w} + T^{t_{w'}} \quad . \tag{3.1}$$

To visualize the partition of the time interval, see Figure 3.1 below. This figure illustrates the partitioning of time if we wanted to make a prediction at week 11, t_{11} =11. Since the prediction period lasts 23 weeks, there are still 12 weeks until the start of the semester of interest or $t_{11'}$ = 12.

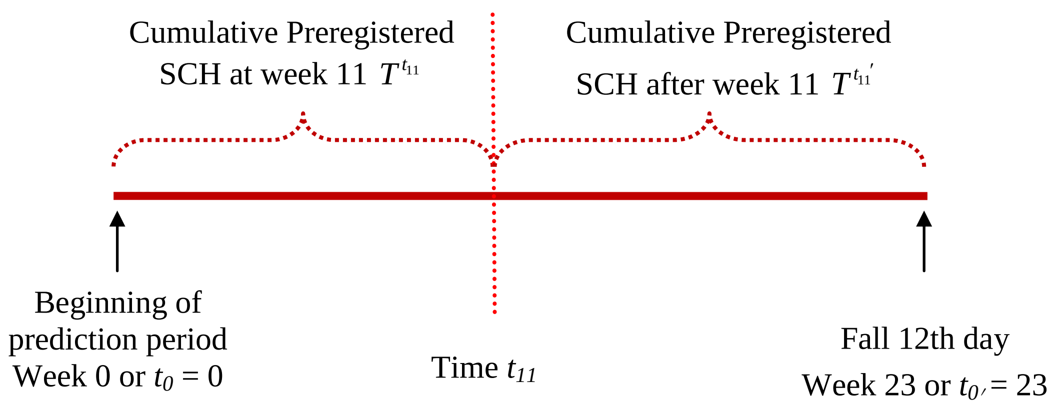

Figure 3.1: Time partition for Model-1

For the reasons mentioned in Chapter-1, the preregistered data for undergraduates and graduates is stratified to predict T_U and T_G. Using Equation 3.1, for any time t_w, we rewrite Equations 1.6 and 1.7 as

$$T_U = T_U^{t_w} + T_U^{t_w'} \tag{3.2}$$

$$T_G = T_G^{t_w} + T_G^{t_w'} \quad . \tag{3.3}$$

Figure 3.2 below illustrates a graph of the points $\left(t_w, T_G^{t_w}\right)$ for 2009 and 2010 graduate strata. Included is this graph is the observed total count SCH, T_G, represented by the horizontal lines.

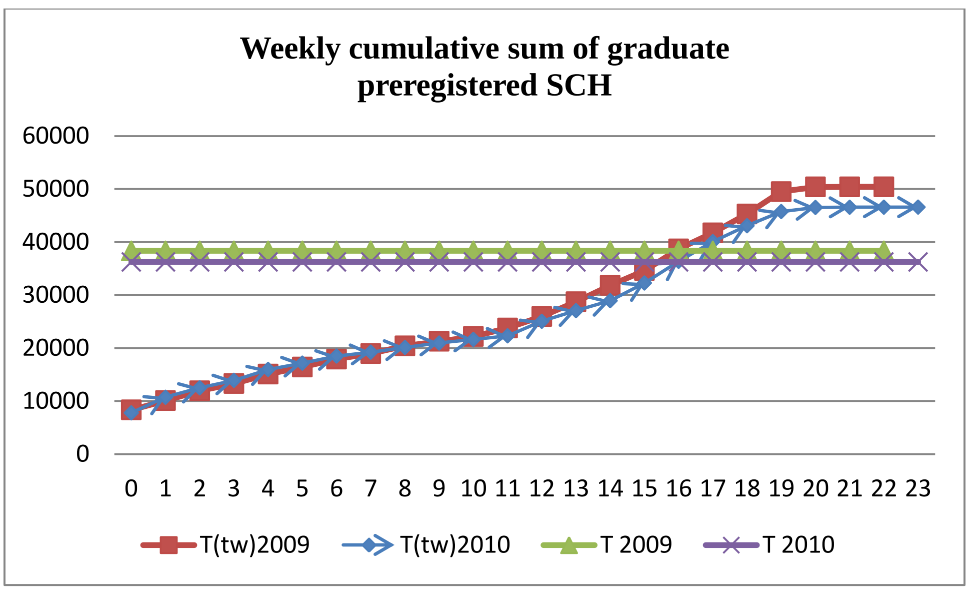

Figure 3.2: Weekly sum of graduate preregistered SCH

To predict fall 2011 SCH, we noted that the historical patterns of the two preceding years, see figure 3.2, follow a similar pattern. The viability of the modeling approach assumes these types of patterns will hold for the subsequent years. In fact, with any predictive modeling approach there is an implicit assumption that the data used to build the model, particularly the parameters that define the model, would be similar to the parameter values we would generate with the unobserved data over the period being predicted. Given the consistency of the pattern over the last two years, the assumption of consistent patterns in the future is viable, particularly if all the other factors that affect enrollment remain unchanged for the upcoming predicted year. For example, if the institution made significant changes to the cost of tuition, then the future pattern of SCH could be altered from the previous year.

We now introduce some fundamental notation in order to distinguish between an estimated or predicted value, and an actual or observed value. For example, T^{t_w}, the first component of Equation 3.1, represents the actual cumulative sum of SCH obtained from the pre-registration data observed as of time t_w. It is worth noting that the first component, T^{t_w}, is known at the time of prediction. However, we need to develop a modeling process that will predict the second component of Equation 3.1, $T^{t_w\prime}$, which is unknown at the time of prediction. The prediction of $T^{t_w\prime}$ will be denoted by $\hat{T}^{t_w\prime}$. The framework to predict $T^{t_w\prime}$ for a particular year will be based on modeling past patterns of $T^{t_w\prime}$. Solving for $T^{t_w\prime}$ in Equation 3.1, we obtain the following equation

$$T^{t_w\prime} = T - T^{t_w}. \tag{3.4}$$

The value of Equation 3.4 can be thought of as the prediction error associated with the cumulative SCH, T^{t_w}, at time t_w as an estimate of T. This error concept is illustrated in Figure 3.3 below. In figure 3.3, you can see that the error at time t_8 = 8 is represented by $T_G^{t_8'} = T_G - T_G^{t_8}$. In general, at each time t_w = 0, 1, 2, …, 23 there is a corresponding error for T^{t_w}.

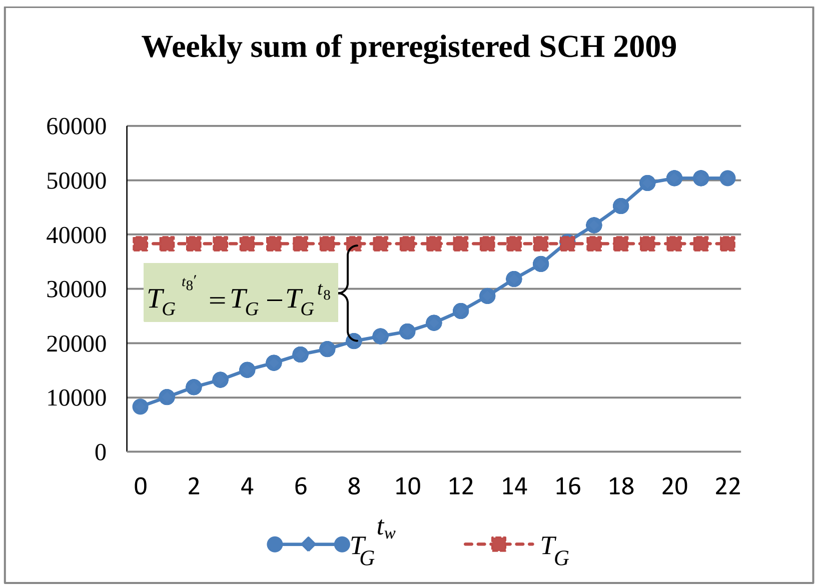

Figure 3.3: Weekly sum of preregistered SCH 2009

In order to predict Equation 3.4 values for the semester of interest, we will use the error patterns of the most recently observed SCH for the cohort and semester of interest. For example, if we were trying to predict $T^{t_{w'}}$ for fall 2013 during the spring semester of 2013, then the most recent data for which the pattern of $T^{t_{w'}}$ is complete and observed is fall 2012 $T^{t_{w'}}$. By using the most recently observed data for $T^{t_{w'}}$, which comes from the SCH pattern of the prior year, we are assuming that last year's data is more closely related to the pattern we would expect for the current pattern. If extraneous factors exists that would alter the pattern of the current year from the patterns from previous years, we expect these changes would occur slowly so that the current year's pattern is going to have the least amount of variation to the SCH pattern observed most recently. Notice in Figures 3.2 that the pattern from 2009 to 2010 changes lightly, and we would expect a small variation as well between 2011 and 2012.

To illustrate how we model Equation 3.4 for say, the fall 2010 undergraduate cohort, we will model the $T^{t_{w'}}$ patterns for undergraduates for fall 2009 (see Figure 3.4). In Figure 3.4, we graphed these 23 error values, $T^{t_w}{}'$, for 2009 then fit a polynomial trend line to fit this error pattern. This polynomial model

$$\hat{T}^{t_w}{}'(t_w) = f(t_w) = 3.5468t_w{}^4 - 189.32t_w{}^3 + 3579.7t_w{}^2 - 32293t_w + 120535$$

will be used as the estimator, $\hat{T}^{t_w}{}'$, for fall 2010 $T^{t_{w'}}$ and is a function with respect to time. Accordingly, $\hat{T}^{t_w}{}'$ can be used to predict fall 2010 $T^{t_{w'}}$ during any time of the prediction period. For example, suppose we wanted to predict the fall 2010 error, $T^{t_{w'}}$, at week $t_2 = 2$. Then, using the Equation above the predicted error is,

$$\hat{T}^{t_w\prime}(2) = f(2) = 3.5468(2)^4 - 189.32(2)^3 + 3579.7(2)^2 - 32293(2) + 120535$$

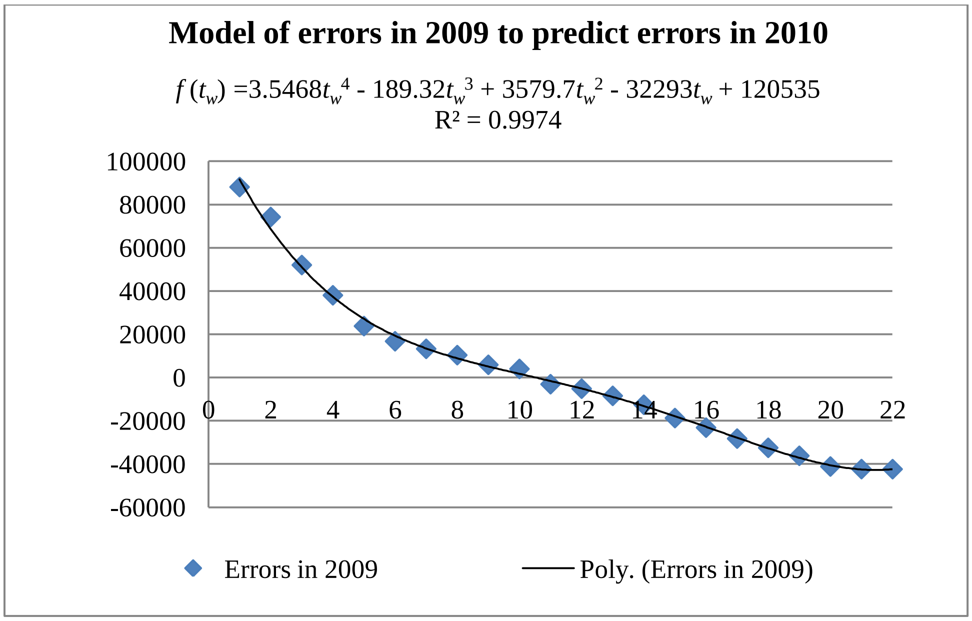

Figure 3.4: Model of 2009 SCH errors for undergraduates

With this error approach to predicting $T^{t_w\prime}$, we now have a methodology to predict total SCH, denoted as *T* in Equation 1.1. Using the $\hat{T}^{t_w\prime}$ as our estimate for $T^{t_w\prime}$, our weekly estimate for Equation 3.1 becomes

$$\hat{T}_{t_w} = T^{t_w} + \hat{T}^{t_w\prime} \ . \tag{3.5}$$

To illustrate how well this predictive methodology works, Table 3.1 shows the results of predicting Equation 3.1 for fall 2012 SCH using the error patterns for fall 2011. Notice how well this prediction method predicted the actual SCH, *T*, which would not be known at the time of prediction. In retrospect, you can see that the predicted value, Equation 3.5, was within 2% or less of the actual total SCH, *T*, 16 out of the 23 weeks of the prediction period or 70 % of the time; and 44% of the time the prediction method was within 1% or less of the actual total SCH. Similar results can be seen for 2011 and 2012 prediction of undergraduate and graduate SCH totals.

Table 3.1

Projection of Total SCH for Fall 2012 Graduate Students under Model-1

Week	Sum T^{t_w}	Predicted Diff $\hat{T}^{t_w}{}'$	Predicted $\hat{T}$	Actual T	Off $\hat{T} - T$	% Off from *T*
0	9087	26913.00	36000.00	35987	13.00	0%
1	12361	23482.67	35843.67	35987	-143.33	0%
2	14374	21057.46	35431.46	35987	-555.54	-2%
3	15916	19381.06	35297.06	35987	-689.94	-2%
4	17753	18224.58	35977.58	35987	-9.42	0%
5	19220	17386.63	36606.63	35987	619.63	2%
6	20339	16693.22	37032.22	35987	1045.22	3%
7	21084	15997.88	37081.88	35987	1094.88	3%
8	22542	15181.54	37723.54	35987	1736.54	5%
9	23393	14152.63	37545.63	35987	1558.63	4%
10	24196	12847.00	37043.00	35987	1056.00	3%
11	24979	11227.98	36206.98	35987	219.98	1%
12	26885	9286.34	36171.34	35987	184.34	1%
13	29250	7040.33	36290.33	35987	303.33	1%
14	31323	4535.62	35858.62	35987	-128.38	0%
15	33337	1845.38	35182.38	35987	-804.63	-2%
16	37754	-929.82	36824.18	35987	837.18	2%
17	40397	-3661.89	36735.11	35987	748.11	2%
18	42697	-6195.34	36501.66	35987	514.66	1%
19	44359	-8347.18	36011.82	35987	24.82	0%
20	44588	-9907.00	34681.00	35987	-1306.00	-4%
21	44612	-10636.91	33975.09	35987	-2011.91	-6%

Week	Sum T^{t_w}	Predicted Diff $\hat{T}^{t_w}{}'$	Predicted $\hat{T}$	Actual T	Off $\hat{T} - T$	% Off from T
22	44615	-10271.58	34343.42	35987	-1643.58	-5%
23	44627	-8518.20	36108.80	35987	121.80	0%

The projections for undergraduate students for each week of 2011 and 2012, as well as the projections for graduate student for each week of 2011, can be found in Appendix A.

CHAPTER IV

NOTATION AND PREDICTIVE MODEL 2

In this chapter, we will introduce additional notation and extend on the notation presented in Chapter-1. This notation is necessary to present the second modeling method, which we will refer to as Model-2, developed in this chapter to predict the total SCH, Equation 1.1.

The second model we will develop is our own modified version of the modeling process introduced by the University of Baltimore (UB). The modeling process used by UB relies on a weighted average and a total headcount; however, they did not use the weights explicitly. Our modified version of this model will use a regular average instead, and total head count in order to predict Equation 1.1. Their approach relied on using Maryland Higher Education Commission's (MHEC) enrollment projections of total headcount. In our research, however, we will use an alternative method to predict headcount, since that methodology was not developed sufficiently or discussed in detail in their research. The University of Baltimore simply provides a link to the MHEC projections (Headcount and Student Credit Hour Projections, p.1), which simply provides calculated numbers of enrollment for the prediction years. Our study requires projections for Texas institutions. Accordingly, we will develop and discuss a self-contained forecasting methodology for headcount that eliminates the reliance on MHEC forecasting.

The University of Baltimore did not provide the criterion applied to select the models used to fit the pattern of historic data of weighted averages of SCH and to predict weighted averages of SCH in future years. In this study, we will discuss a criterion that will be used to select the prediction models for the averages of SCH in future years.

Conceptually, the Model-2 approach is derived from a simple idea stemming from the equation of an average. For example, let μ represent the average SCH of the total number of preregistered students, *N*. By definition, this average can be written as

$$\mu = \frac{\sum_{k=1}^{N} i_k x_k}{N} \quad . \tag{4.1}$$

Using Equation 1.1, we can rewrite Equation 4.1 as

$$\mu = \frac{\sum_{k=1}^{N} i_k x_k}{N} = \frac{T}{N} \quad . \tag{4.2}$$

Therefore, the total SCH can be obtained by multiplying Equation 4.2 by N to obtain

$$N \cdot \mu = \frac{T}{N} \cdot N \quad .$$

Solving for T in the equation above, we obtain the following formula for total SCH

$$T = N\mu \ . \tag{4.3}$$

Accordingly, to predict the total SCH, *T*, we need to predict-the right hand components of Equation 4.3.

To predict *T* in Equation 4.3, we begin by stratifying the indexes of **P** into graduates and undergraduates, since each group has different criteria for classification as a full-time student. Recall from Chapter-1, the total SCH can be written as

$$T = T_U + T_G,$$

Using Equation 4.3 for each strata, we can rewrite Equations 1.6 and 1.7 as

$$T_U = \mu_U \cdot N_U \tag{4.4}$$

$$T_G = \mu_G \cdot N_G \tag{4.5}$$

where $N_U = |\mathbf{P}_U|$ and $N_G = |\mathbf{P}_G|$. Recalling from Chapter-1, $\mathbf{P} = \mathbf{P}_U \cup \mathbf{P}_G$, where $\mathbf{P}_U$ is the set of indices for preregistered undergraduates and $\mathbf{P}_G$ the set of indices for preregistered graduates.

In order to formulate a prediction for either Equation 4.4 or Equation 4.5, we require an estimate of the components of the right hand side of the equations. In general, this method will use an average to predict μ for the particular stratification of interest. In addition, we will develop a prediction method for the total headcount, *N*, which we will discuss later in this chapter.

Using the same notation from Chapter-3 to distinguish between an estimated or predicted value and an actual or observed value, we let μ represent the actual average of SCH obtained from all data after pre-registration data has been completed for a semester prior to the semester of interest, and the predicted value or estimated value will be denoted as $\hat{\mu}$, the estimated average of SCH for the future semester of interest. Similarly, N represents the actual total headcount for a particular stratification and $\hat{N}$ represents the predicted total headcount. Finally, to predict actual total SCH, *T*, we will use estimates in Equations 4.4 and 4.5

$$\hat{T}_U = \hat{\mu}_U \cdot \hat{N}_U \tag{4.6}$$

$$\hat{T}_G = \hat{\mu}_G \cdot \hat{N}_G \tag{4.7}$$

To predict μ, with $\hat{\mu}$, we explored the patterns of the observed μ three years prior to the semester of interest. Next, we superimposed a trend line that fit the graphed pattern. Using the equation of the fitted graph, we can then predict $\hat{\mu}$ for the year of the semester of interest. In particular $\hat{\mu}_{2011}$, which is read as the predicted average of the SCH for 2011, was derived from calculating and graphing μ_{2008}, μ_{2009}, and μ_{2010}. Figure 4.1 below graphs the observed pattern of the weighted SCH for 2008-2010. This is simply a graph of the following three points: $(1, \mu_{2008})$, $(2, \mu_{2009})$, $(3, \mu_{2010})$. Accordingly, the idea is to find an appropriate model that fits this three year pattern.

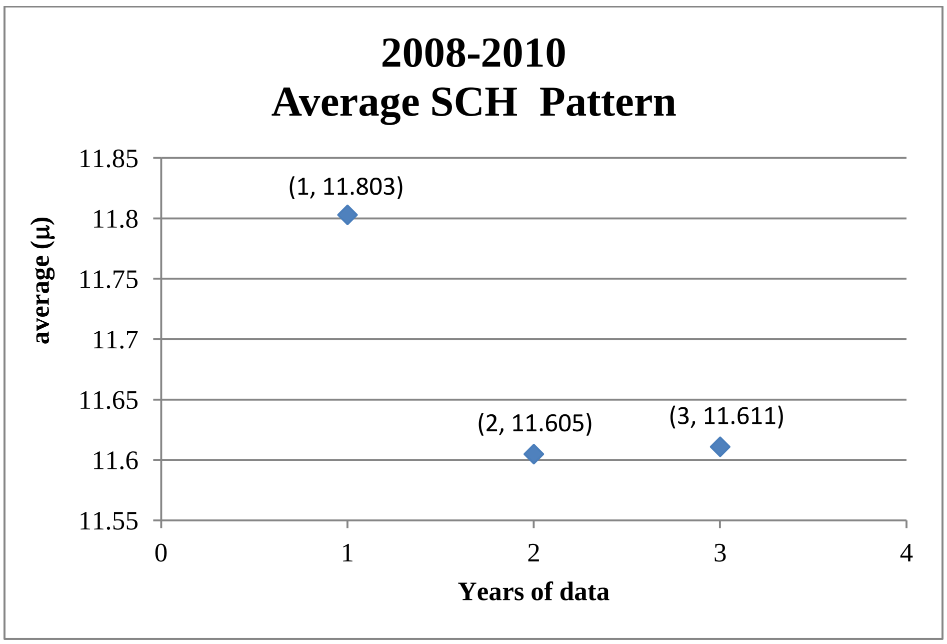

Figure 4.1: Average SCH pattern

Table 4.1 shows the observed μ for each undergraduate and graduate strata for each fall semester from 2008 through 2010.

Table 4.1

Average SCH 2008-2010

	Fall		
Average SCH	2008	2009	2010
μ_U	11.803	11.605	11.611
μ_G	6.808	6.971	6.359

In order to predict μ for say fall 20ll, we need to select a model that fits the pattern of the graphed points in figure 4.1. We will explore the fit of a power, an exponential, and a natural logarithmic trend over the graphed points. We then will evaluate the fit of the pattern and select one from these three possible models to predict the average of SCH for 2011 using a set of criteria described later in this chapter. In the following discussion, we will develop notation to represent a time component in the model specification.

This modeling approach requires the time to be measured in years. Accordingly, let t_y represents the time in years. The next four figures below show μ as a function of t_y.

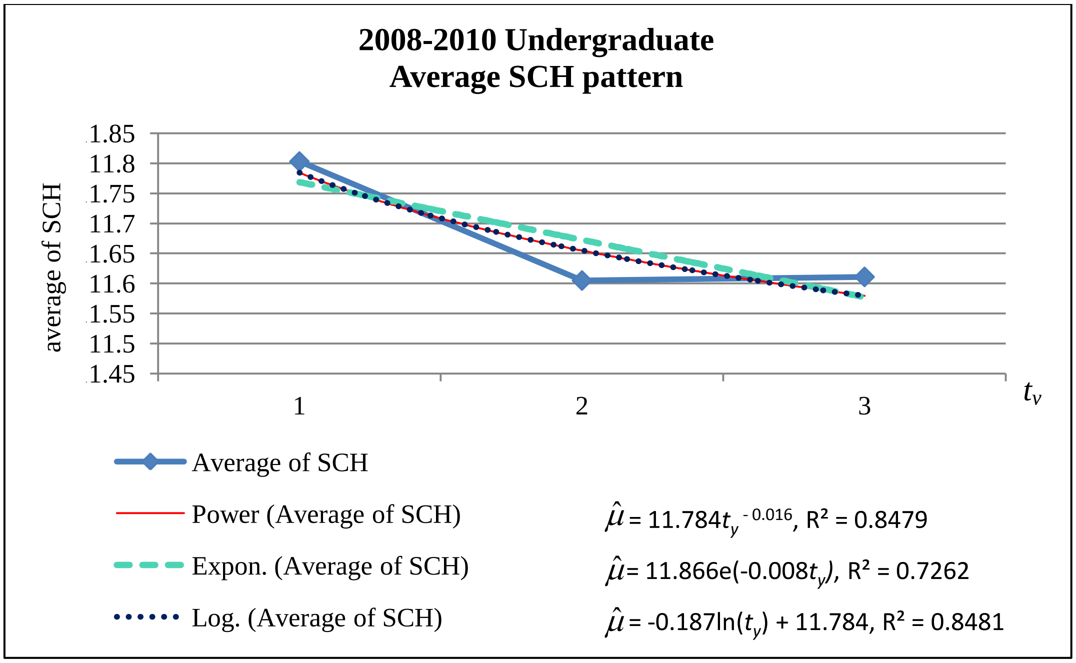

Figure 4.2: 2008-2010 Undergraduate average SCH pattern

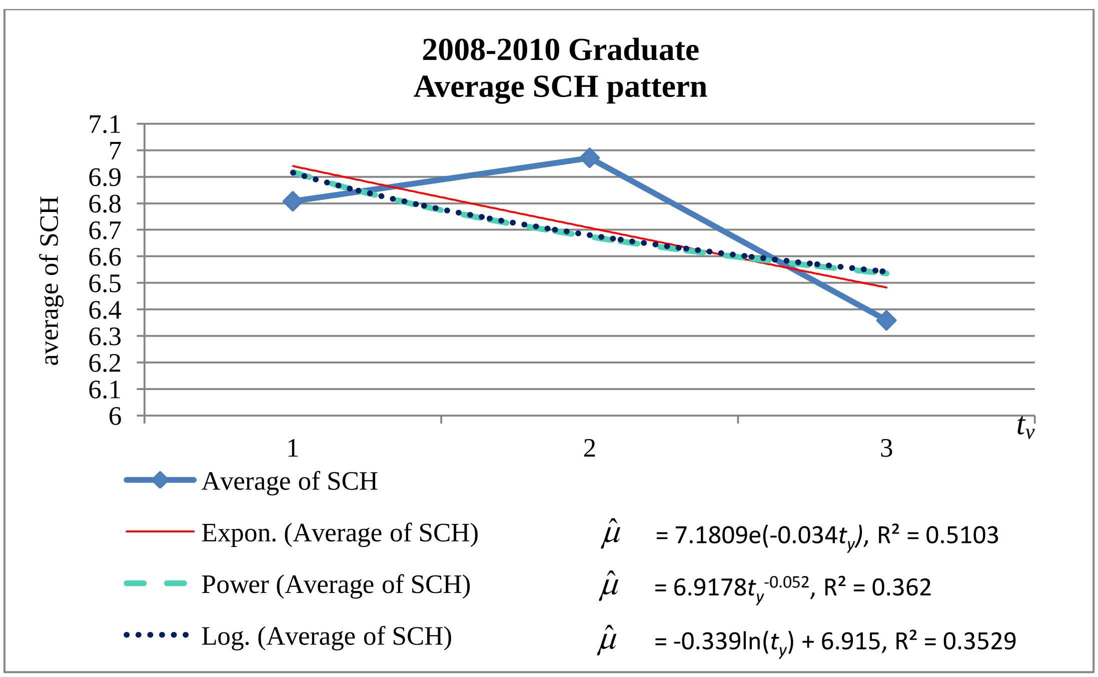

Figure 4.3: 2008-2010 Graduate average SCH pattern

We used each of the equations in Figure 4.2 and 4.3 to predict the average of SCH for 2011, and we obtained the averages shown below. The highlighted rows indicated the selected model used in order to predict the actual average SCH. In the tables below, you can see the difference between the predicted and the actual value. However, it is worth noting that a prediction happens before realizing the actual value. Thus, having a selection criterion for selecting a prediction model is an important discussion we will address later in this chapter. In the meantime, Tables 4.2-Tables 4.5 highlight the selected models we would have used at the time of the prediction.

Table 4.2

Fall 2011 Undergraduates Model Fit Average of SCH

Fall 2011 Undergraduates Model fit average of SCH	Predicted $\hat{\mu}$	Observed μ	Difference $\hat{\mu}-\mu$
Exponential	11.49230	11.63785	-0.14555
Power	11.52550	11.63785	-0.11235
Logarithmic	11.52476	11.63785	-0.11308

Table 4.3

Fall 2011 Graduates Model Fit Average of SCH

Fall 2011 Graduates Model fit average of SCH	Predicted $\hat{\mu}$	Observed μ	Difference $\hat{\mu}-\mu$
Exponential	6.26780	6.35879	-0.09100
Power	6.43666	6.35879	0.07787
Logarithmic	6.44505	6.35879	0.08625

We then graphed μ from each fall from 2009 through 2011 and fit a power, an exponential, and a natural logarithmic trend line to predict the average of SCH for 2012.

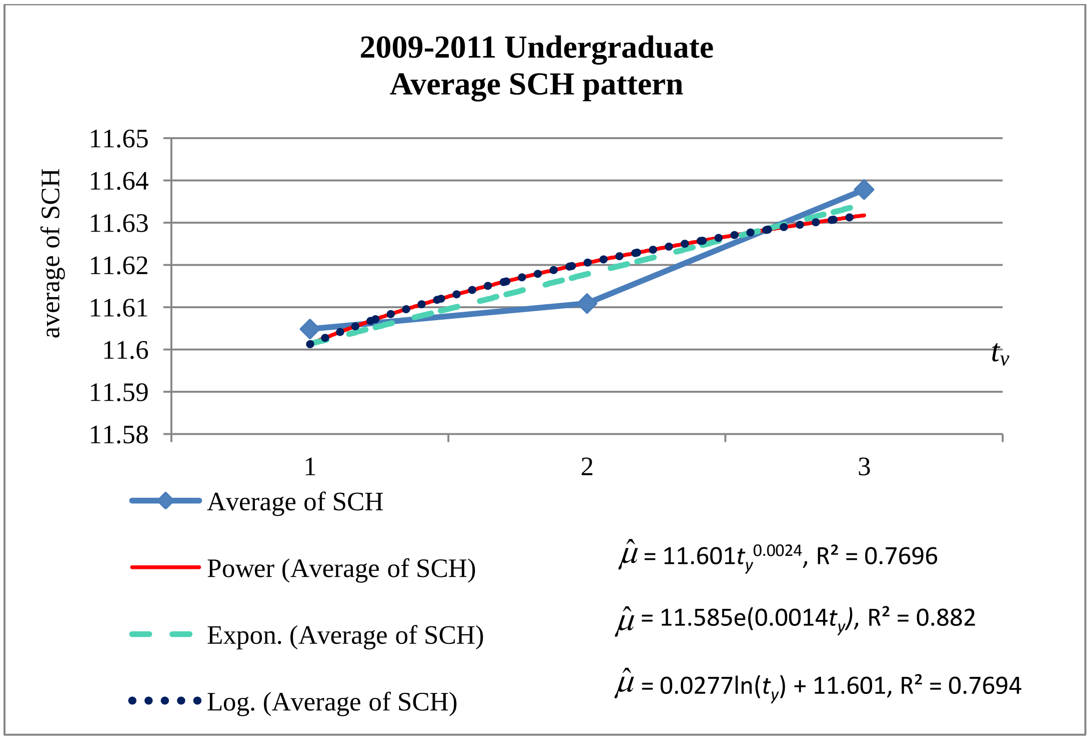

Figure 4.4: 2009-2011 Undergraduate average SCH pattern

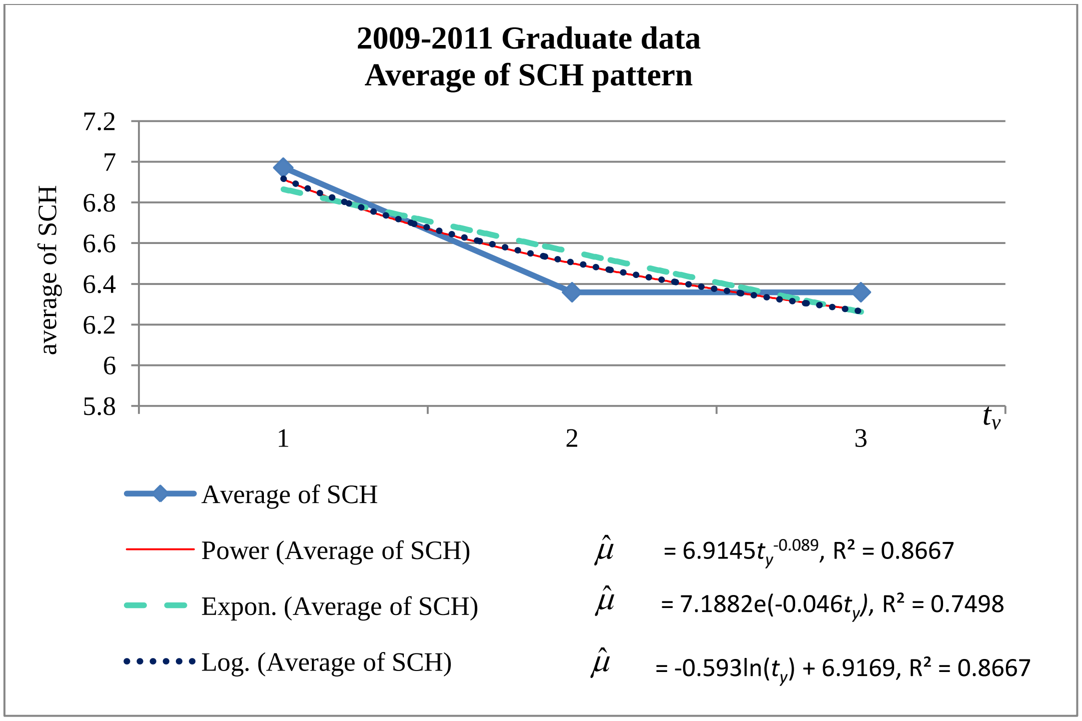

Figure 4.5: 2009-2011 Undergraduate average SCH pattern

We used each of the equations shown in figure 4.4 and 4.5 to predict the average of SCH for 2012, and we obtained the predictions shown below.

Table 4.4

Fall 2012 Undergraduates Model Fit Average of SCH

Fall 2012 Undergraduates Model fit average of SCH	Predicted $\hat{\mu}$	Observed μ	Difference $\hat{\mu}-\mu$
Exponential	11.65006	11.59451	0.05554
Power	11.63966	11.59451	0.04515
Logarithmic	11.56260	11.59451	-0.03191

Table 4.5

Fall 2012 Graduates Model Fit Average of SCH

Fall 2012 Undergraduates Model fit average of SCH	Predicted $\hat{\mu}$	Observed μ	Difference $\hat{\mu}-\mu$
Exponential	5.98012	6.31683	-0.33671
Power	6.11192	6.31683	-0.20492
Logarithmic	6.09483	6.31683	-0.22201

In our previous discussion, we developed several competing models for predicting the parameter μ in Tables 4.2-Tables 4.5. As with any prediction modeling process, at the time you make the prediction you have only observed part of the data and therefore you do not have the luxury of knowing which competing model will actually provide a prediction closest to the actual value of the parameter. In the following discussion, we will specify the selection criterion and the decision making process we used to select the highlighted model to predict μ. Our decision making process is going to involve two factors.

The first factor involves evaluating how well the models fit the general pattern. In each case, we are fitting a model on three years of data. Considering the fact that we are only trying to model the average SCH over a three year period, we could theoretically always find a polynomial model that would have perfect fit (i.e., go through all three points). Accordingly, fitting a polynomial would generate a coefficient of determination, R^2, equal to 1. In general, the coefficient of determination provides the proportion of the total variation that is explained by the fitted model. For a more detailed explanation of the coefficient of determination, see Ranney and Thigpen (1981). Although a polynomial model would generate a $R^2 = 1$ or would explain 100% of the total variation, for such few data points the polynomial model would bring about a problem well known in statistics referred to as overfitting (Vaughan and Ormerod, 2005). A model that overfits the data produces a curve that fits a particular data well but does not model the underlying trend well. For this reason, we did not consider a polynomial model as an option to predict μ. For the other non-polynomial models, we want to consider models that have the highest R^2 value. Although the R^2 value is not the sole criteria for model selection, we specify how to judge this criterion in the following discussion.

In statistics, the correlation is categorized as weak, moderate, or strong using the boundaries shown in Table 4.6. Using these values and their corresponding range for R^2, we will evaluate how well the three possible models capture the 3-year pattern of averages of SCH.

Table 4.6

Boundaries for Correlation Values

Correlation description	r	R^2
Weak	0 – 0.39	0 – 0.16
Moderate	0.4 – 0.69	0.16 – 0.49
Strong	0.7 – 1	0.49 – 1

When two or more competing models fall into the highest R^2 grouping as specified in Table 4.6, then we will use a second factor in order to facilitate a decision on which model to choose.

The second factor to consider when selecting a model for estimating μ involves an intuitive notion that comes with financial planning. That is, when planning a budget you want to make sure that you have the necessary finances or money to pay for the expenses you will incur. For planning purposes, this means you should not spend more money than you will generate. Thus, the second factor we consider when selecting an appropriate model is to select the model that will predict more conservatively.

With the assumption that university administrators prefer a conservative projection, we will choose the model that has the least rate of at the last observed year of data, which is year three. Conceptually, this means that we want the prediction to stay as flat as possible from the average SCH in year three to the average SCH of the predicted year or year four. In a mathematical context, the least rate of change translates to consider selecting the model for μ with the first derivative, evaluated at $t_y = 3$, closest to zero. Without loss of generality, suppose we have three models for estimating μ: $\hat{\mu}_1\left(t_y\right)$, $\hat{\mu}_2\left(t_y\right)$,and $\hat{\mu}_3\left(t_y\right)$.

Then, the selected model with the least rate of change is

$$\min\left\{\hat{\mu}_1'(3), \hat{\mu}_2'(3), \ldots, \hat{\mu}_3'(3)\right\} \tag{4.8}$$

where $\hat{\mu}'(3)$ represents the first derivative of the respective model evaluated at $t_y = 3$.

Criterion 4.8 can be illustrated for the three Undergraduate competing models of μ in Figure 4.2. Using this criterion, we selected the logarithmic model to predict μ for Undergraduates in 2011 using the 2008-2010 pattern. The R^2 value of both, the logarithmic equation ($R^2 = 0.8481$) and the power equation ($R^2 = 0.8479$), fall in the strong category according to Table 4.6. Because they both had the same rate of change $\hat{\mu}'(3) = -0.062$ at 2010, we made our decision purely based on the greater R^2 value, which is larger for the logarithmic model. Similarly, the logarithmic model fit to the 2009-2011 Undergraduate pattern was selected to predict μ for fall 2012, see figure 4.4. The R^2 value of both, the logarithmic equation ($R^2 = 0.7694$) and the power equation ($R^2 = 0.7696$) in figure 4.4, fall in the strong category. Therefore, we considered the second factor of the criterion: the logarithmic model had a lesser rate of change ($\hat{\mu}'(3) = 0.0092$) than the power model $(\hat{\mu}'(3) = 0.0093)$, indicating that the projection under the logarithmic model is more conservative.

In a similar way, we applied the criterion above for each model for Graduate students and decided that the exponential model was the best fit to predict μ for 2011 using the 2008-2010 pattern, see Figure 4.3. In Figure 4.3, the R^2 value of the exponential equation (R^2 = 0.5103) was the only model in the strong category, according to Table 4.6, so we selected the exponential model without considering criterion specified in 4.8. To predict μ for Graduates in 2012 using the 2009-2011 pattern, as shown in Figure 4.5, we selected the power model. In Figure 4.5, the R^2 value of both the logarithmic equation (R^2 = 0.7694) and the power equation (R^2 = 0.7696) fall in the strong category. Therefore we considered the second criterion specified in Equation 4.8. Using this additional criterion, the power model had a lesser rate of change ($\hat{\mu}'(3) = -0.186$) than the logarithmic model ($\hat{\mu}'(3) = -0.198$), indicating that the projection under the power model is more conservative and hence the desired model for predicting μ. The derivative calculations of the criterion in Equation 4.8 can be found in Appendix-B for the examples mentioned above.

Equations 4.9 and 4.10 show the equations of the logarithmic trend lines used to predict μ for undergraduates in Figures 4.2 and 4.4, respectively. The input for these equations is t_y. Since figure 4.2 is a pattern of 2008-2010 undergraduate average SCH patterns used to predict μ for fall 2011, the subscript of the estimator $\hat{\mu}$ in Equation 4.9, in this case U-2011, indicates the cohort of interest and the predicted year. Thus, $\hat{\mu}_{U-2011}$, is used to predict the average SCH, μ, for fall 2011 by evaluating this function at t_y = 4. Similarly, $\hat{\mu}_{U-2012}$ evaluated at t_y = 4 provides the predicted average SCH for 2012.

$$\hat{\mu}_{U-2011} = -\ 0.187 * ln(t_y) + 11.784 \tag{4.9}$$

$$\hat{\mu}_{U-2012} = -\ 0.0277 * ln(t_y) + 11.601 \qquad (4.10)$$

Equations 4.11 and 4.12 show the equations of the exponential and power trend lines used to project μ for graduate students in 2011 and 2012 respectively.

$$\hat{\mu}_{G-2011} = 7.1809 e^{(-0.034 t_y)} \qquad (4.11)$$

$$\hat{\mu}_{G-2012} = 6.9145 * t_y^{(-0.089)}\ . \qquad (4.12)$$

Now that we have described a method to predict μ in Equation 4.3, we now focus on developing a model to predict *N* in Equation 4.3. To develop a model to predict N, which is total headcount, we will use a similar framework to the one described in Chapter-3 for our Model-1to predict the total count of SCH.

The modeling technique to predict N, requires a partition of the time interval into weekly periods t_w, as described in detail in Chapter -3. In this case, at the end of each time period t_w, there is cumulative headcount of preregistered students n^{t_w} known before time t_w . However, since preregistration is still ongoing, we know that there are students who will preregister after time t_w. Let $n^{t_w'}$ represent the expected cumulative headcount of students that will preregistered after time t_w. Using this notation, we can define N, the first component of Equation 4.3 as

$$N = n^{t_w} + n^{t_{w'}} \quad . \tag{4.13}$$

For the reasons mentioned in Chapter-1, the preregistered data is stratified for undergraduates and graduates to predict N_U and N_G. Accordingly, N_U and N_G can be written as

$$N_U = {n_U}^{t_w} + {n_U}^{t_{w'}}$$

$$N_G = {n_G}^{t_w} + {n_G}^{t_{w'}} \quad .$$

Figure 4.6 illustrates a graph of the points $\left(t_w, {n_U}^{t_w}\right)$ for 2009 and 2010 undergraduate strata. Included in these graphs is the observed total headcount, N_U, represented by the horizontal line.

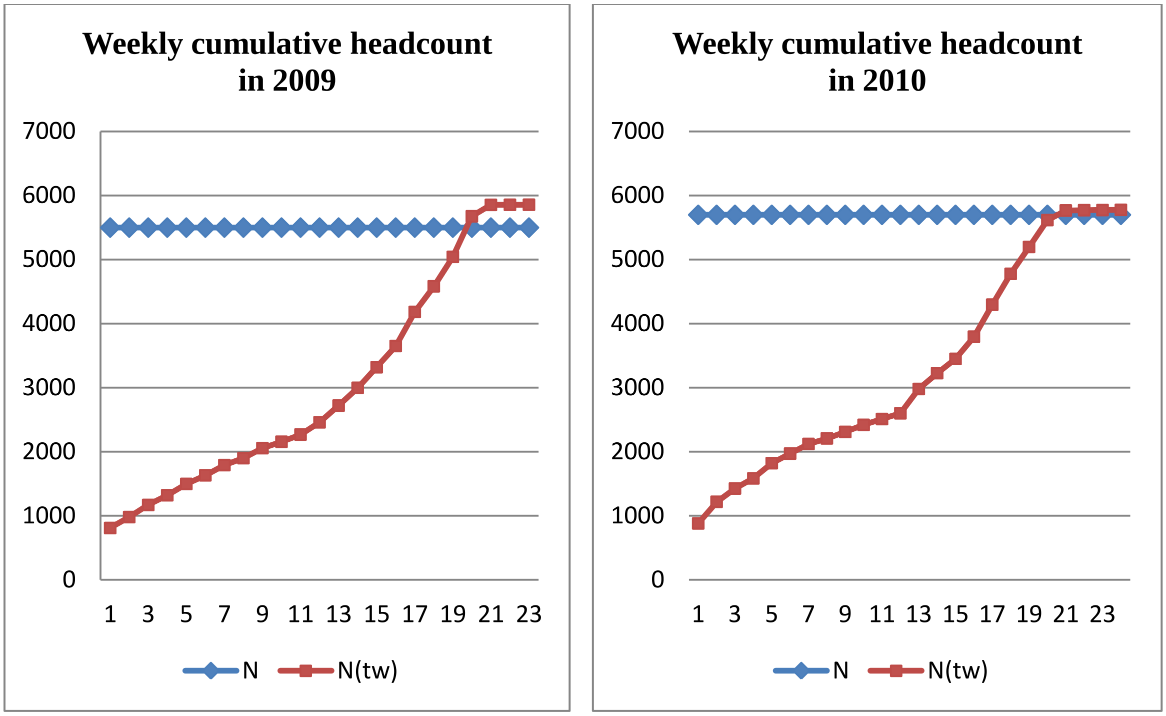

Figure 4.6: Weekly cumulative headcount

To predict fall 2011 headcount, we noted that the historical patterns of the two preceding years follow a similar pattern. Therefore, we assumed that the counts of preregistered students after time t would behave similar to the counts of preregistered students after time *t* of the prior year, if all the other factors that affect enrollment remain unchanged for the upcoming predicted year (see Figure 4.6).

While the first, n^{t_w}, component of Equation 4.13 is known at the time of prediction t_w, the second component, $n^{t_w'}$, is unknown at the time of prediction t_w.

Therefore, in order to predict *N*, we first need to develop a model to predict $n^{t_w'}$. The prediction of $n^{t_w'}$ will be denoted by $\hat{n}^{t_w'}$. The framework to predict $n^{t_w'}$ for a particular year will be based on modeling past patterns of $n^{t_w'}$. Solving for $n^{t_w'}$ in Equations 4.13, we obtain the following equation

$$n^{t_w'} = N - n^{t_w} \; . \qquad (4.14)$$

The value of Equation 4.14 can be thought of as the prediction error associated with the cumulative headcount of preregistered students, n^{t_w}, at time t_w as an estimate of *N*. In general, at each time $t_w = 0,1, 2, \ldots, 23$ there is a corresponding error for n^{t_w}.

To illustrate how we model Equation 4.14 for say, the fall 2010 graduate cohort, we will model the $n^{t_w'}$ patterns for graduates for fall 2009 (see Figure 4.7). In Figure 4.7, we graphed these 23 error values, $n^{t_w'}$, for 2009 then fit a polynomial trend line to fit this error pattern. This polynomial model

$$\hat{n}^{t_w'}(t_w) = f\ (t_w) = 0.1389t_w^{\,4} - 6.5809t_w^{\,3} + 93.022t_w^{\,2} - 602.14t_w + 5354.8$$

will be used as the estimator, $\hat{n}^{t_{w'}}$, for fall 2010 $n^{t_{w'}}$ and is a function with respect to time. Accordingly, $\hat{n}^{t_{w'}}$ can be used to predict fall 2010 $n^{t_{w'}}$ during any time of the prediction period.

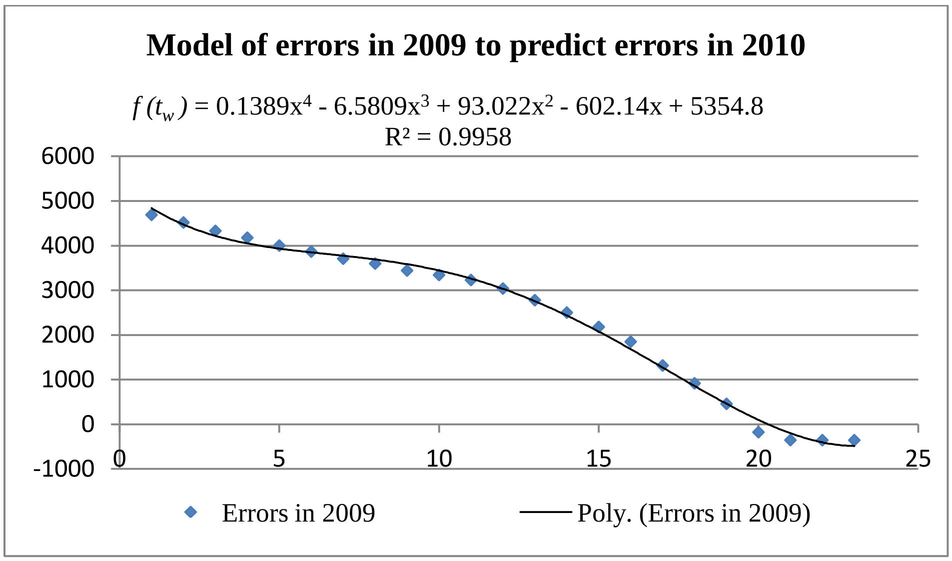

Figure 4.7: Model of 2009 headcount errors for graduates

Using Equation 4.13, we defined an estimate for $\hat{N}$ as

$$\hat{N} = n^{t_w} + \hat{n}^{t_{w'}} . \tag{4.15}$$

Similar to the idea presented in Figure 4.7, we used the pattern of errors in 2011 to predict the errors in 2012, $\hat{n}^{t_{w'}}$. Table 4.7 illustrates Equation 4.15 for graduate students for each preregistration week of 2012. Using this methodology, the predicted headcount $\hat{N}$ is within 2% or better of the actual headcount N 57% of the time.

Table 4.7

Headcount Projection for Fall 2012 Graduate Students

Week t_w	Sum n^{t_w}	Actual N	Predicted Difference $\hat{n}^{t_w'}$	Predicted Headcount $\hat{N}$	Off by $\hat{N} - N$	Off by %
0	1053	5697	4686.90	5739.90	42.90	1%
1	1436	5697	4280.36	5716.36	19.36	0%
2	1664	5697	3995.84	5659.84	-37.16	-1%
3	1860	5697	3802.29	5662.29	-34.71	-1%
4	2085	5697	3671.91	5756.91	59.91	1%
5	2266	5697	3580.16	5846.16	149.16	3%
6	2397	5697	3505.81	5902.81	205.81	4%
7	2484	5697	3430.86	5914.86	217.86	4%
8	2661	5697	3340.61	6001.61	304.61	5%
9	2752	5697	3223.61	5975.61	278.61	5%
10	2847	5697	3071.70	5918.70	221.70	4%
11	2944	5697	2879.98	5823.98	126.98	2%
12	3180	5697	2646.83	5826.83	129.83	2%
13	3474	5697	2373.88	5847.88	150.88	3%
14	3731	5697	2066.07	5797.07	100.07	2%
15	4001	5697	1731.56	5732.56	35.56	1%
16	4526	5697	1381.83	5907.83	210.83	4%
17	4872	5697	1031.60	5903.60	206.60	4%
18	5213	5697	698.88	5911.88	214.88	4%
19	5499	5697	404.93	5903.93	206.93	4%
20	5551	5697	174.30	5725.30	28.30	0%

Week t_w	Sum n^{t_w}	Actual N	Predicted Difference $\hat{n}^{t_w'}$	Predicted Headcount $\hat{N}$	Off by $\hat{N} - N$	Off by %
21	5557	5697	34.81	5591.81	-105.19	-2%
22	5558	5697	17.53	5575.53	-121.47	-2%
23	5560	5697	156.84	5716.84	19.84	0%

The modeling equations used to predict total headcount in 2011 and 2012 for undergraduate and graduate students are shown in Appendix-C.

Finally, multiplying our estimators $\hat{\mu}$ and $\hat{N}$, we provide the estimate to predict total SCH *T*, by rewriting Equation 4.3 as

$$\hat{T}_{t_w} = \hat{N}_{t_w} \cdot \hat{\mu} \tag{4.16}$$

where $\hat{T}_{t_w}$ is the weekly estimate of total SCH, *T*; $\hat{N}_{t_w}$ is the weekly estimate of total headcount, *N*; and $\hat{\mu}$ is the estimate of the average SCH, μ for the year of interest. With this approach, we define and alternative equation to find the total SCH, *T*, in Equation 1.1.

Table 4.8 below illustrates this alternative approach to predict 2012 graduate student total SCH, Equation 4.7, for each week of 2012. Using this methodology, the predicted total SCH, $\hat{T}$ is within 2% or better of the actual SCH, *T*, 65% of the time.

Table 4.8

Projection of Total SCH for Fall 2012 Graduate Students under Model-2

Week	$\hat{N}_G$	$\hat{\mu}_G$	$\hat{T}_G$	T_G	Off $\hat{T}_G - T_G$	Off %
0	5739.90	6.11192	35081.79	35987	905.21	3%
1	5716.36	6.11192	34937.90	35987	1049.10	3%
2	5659.84	6.11192	34592.50	35987	1394.50	4%
3	5662.29	6.11192	34607.46	35987	1379.54	4%
4	5756.91	6.11192	35185.73	35987	801.27	2%
5	5846.16	6.11192	35731.26	35987	255.74	1%
6	5902.81	6.11192	36077.47	35987	-90.47	0%
7	5914.86	6.11192	36151.12	35987	-164.12	0%
8	6001.61	6.11192	36681.31	35987	-694.31	-2%
9	5975.61	6.11192	36522.43	35987	-535.43	-1%
10	5918.70	6.11192	36174.60	35987	-187.60	-1%
11	5823.98	6.11192	35595.69	35987	391.31	1%
12	5826.83	6.11192	35613.09	35987	373.91	1%
13	5847.88	6.11192	35741.78	35987	245.22	1%
14	5797.07	6.11192	35431.19	35987	555.81	2%
15	5732.56	6.11192	35036.95	35987	950.05	3%
16	5907.83	6.11192	36108.17	35987	-121.17	0%
17	5903.60	6.11192	36082.33	35987	-95.33	0%
18	5911.88	6.11192	36132.90	35987	-145.90	0%
19	5903.93	6.11192	36084.32	35987	-97.32	0%
20	5725.30	6.11192	34992.56	35987	994.44	3%
21	5591.81	6.11192	34176.65	35987	1810.35	5%
22	5575.53	6.11192	34077.19	35987	1909.81	5%
23	5716.84	6.11192	34940.83	35987	1046.17	3%

The projections for undergraduate students for each preregistration week of 2011 and 2012, as well as the projections for graduate students for each preregistration week of 2011, can be found in Appendix-D.

CHAPTER V

RESULTS AND FUTURE RESEARCH

In this research, we developed two modeling approaches to predict total SCH, *T*, in Equation 1.1. These modeling approaches were used to predict total SCH by undergraduates and graduate stratification. In Chapter-3, we illustrated the predictive accuracy of Model-1, whereas in Chapter-4 we illustrated the predictive accuracy of Model-2. In both Chapter-3 and Chapter-4, however, we only illustrated the predictive accuracy using the respective models on the graduate student cohort. In this chapter, we will extend this discussion by comparing the predictive accuracy of the two modeling approaches we developed for both the undergraduate and graduate strata. In addition, we will discuss the strengths and weaknesses of using each modeling approach. Finally, we will discuss future research regarding the development of alternative modeling techniques for predicting total SCH, *T*.

One of the major contributions of this research is the development of two modeling approaches to predict total SCH by relying on parallel patterns of cumulative preregistration data. Using preregistration data for both modeling approaches provides a viable approach to predict total SCH, since this type of data should be readily available to all institutions of higher education. This is in contrast to the University of Baltimore's

SCH model; their model relied on enrollment projections from MHEC, a source not available to institutions outside of Maryland.

An advantage of using the Model-1 approach to predict total SCH, is that it only requires using one year of historical data, although you can certainly use multiple years of data, in order to predict total SCH. In addition, the Model-1 approach relies solely on using a simple pattern of the following points $\left(t_w, T^{t_w}\right)$, over weekly periods of time, with the fixed historical value of *T*.

The model developed in Chapter-4, Model-2, relies on estimating the average SCH, μ, and total headcount, *N*, to predict *T*. The model to predict total headcount relied on using a simple error technique, developed for Model-1, but on cumulative parallel patterns of preregistered headcount instead of cumulative parallel patterns of SCH. In addition, Model-2 relied on modeling historical average SCH patterns over multiple years. In our research, we used three years of data for the projection of the average SCH due to the limited number of years of data that we had available at the start of this research. Modeling historical patterns of μ over a longer period may add more accuracy to this prediction method. Another major contribution of this research, is the inclusion of a decision making process to select a predictive model for the average SCH, which was not addressed in the modeling approach by the University of Baltimore.

To compare the predictive accuracy of the two modeling approaches presented in Chapter-3 and Chapter-4, we will look at the deviation of the weekly predicted total SCH, $\hat{T}$, with the actual total SCH, *T*, over the 23 weeks period. Using notation from Equations 3.5 and 4.16 this deviation is defined as

$$dev_{t_w} = \left(\hat{T}_{t_w} - T\right) \text{ for } t_1, t_2, \ldots, t_{23}.$$

In addition, we are interested in the weekly percent deviation which is defined as

$$\%dev_{t_w} = \left(\frac{dev_{t_w}}{T}\right) \times 100\%.$$

To judge the models predictive accuracy we will compare the following two statistics:

$$\mu_{|dev|} = \frac{\sum_{w=1}^{23} \left|dev_{t_w}\right|}{23}. \quad (5.1)$$

and

$$\mu_{|\%dev|} = \frac{\sum_{w=1}^{23} \left|\%dev_{t_w}\right|}{23} \quad (5.2)$$

Using Equations 5.1 and 5.2, Tables 5.1 and 5.2 show the comparative predictive results of each modeling approach on undergraduates and graduates total SCH for fall 2011 and fall 2012. In general Model-2 outperforms Model-1 in predicting total SCH for each stratum. Yet, we can see that both models are within 4.5% of the actual total SCH for each year of prediction and corresponding cohort. To put the predictive accuracy of these models in perspective, we will focus on the greatest value of $\mu_{|dev|}$ for 2012 undergraduates, which is 4791.28. This means that on average the SCH weekly predictions for 2012 were 4791.28 hours off from the actual total SCH for 2012 undergraduates; this is equivalent to about 400 full-time equivalent (FTE) students, assuming 12 SCH for a full-time undergraduate. Although this model was our most inaccurate, it was only off on average 4.38% SCH, as defined by Equation 5.2, from the actual total SCH over the 23 week prediction period. However, over the two year prediction of total undergraduate SCH our most accurate prediction, by Model-2, was only off on average 1.55% SCH from the actual total SCH over the 23 week prediction period.

Table 5.1

Comparison of the Results of Model-1 and Model-2 for Undergraduates

	2011		2012	
	$\mu_{\|dev\|}$	$\mu_{\|\%dev\|}$	$\mu_{\|dev\|}$	$\mu_{\|\%dev\|}$
Model 1	2041.32	1.95%	4791.28	4.38%
Model 2	1621.26	1.55%	3425.23	3.13%

Table 5.2

Comparison of the Results of Model-1 and Model-2 for Graduates

	2011		2012	
	$\mu_{\lvert dev \rvert}$	$\mu_{\lvert \% dev \rvert}$	$\mu_{\lvert dev \rvert}$	$\mu_{\lvert \% dev \rvert}$
Model 1	1474.69	4.06%	723.78	2.01%
Model 2	851.44	2.35%	674.75	1.87%

The actual value of interest is a prediction of total SCH,*T* in Equation 1.1, which requires a combination of the predictions by stratum. Therefore we rewrite Equation 1.5 as

$$\hat{T} = \hat{T}_U + \hat{T}_G \tag{5.3}$$

In Table 5.3 below, we show how far off Equation 5.3 is from the actual total SCH, $T_,$ for each modeling approach using Equations 5.1 and 5.2.

Table 5.3

Comparison of the Results of Model-1 and Model-2 for the Total Population

	2011		2012	
	$\mu_{\lvert dev \rvert}$	$\mu_{\lvert \% dev \rvert}$	$\mu_{\lvert dev \rvert}$	$\mu_{\lvert \% dev \rvert}$
Model 1	2765.32	1.96%	4926.29	3.39%
Model 2	1614.44	1.14%	3665.14	2.52%

Table 5.3 shows the comparative predictive results of each modeling approach on the total population, combining undergraduates and graduates total SCH, for fall 2011 and fall 2012. Both modeling approaches predict fairly accurately and over a two year period, no model prediction is off by more than 3.5%. In general Model-2 outperforms Model-1 in predicting total SCH.

We are also interested in whether the prediction method has a tendency to over predict versus under predict total SCH, *T*. Assuming that Equation 5.1 and 5.2 are low for each modeling method, then we would likely favor the method that has a tendency to under predict. The following indicator function allows us to determine if a method under predicts.

$$f\left(dev_{t_w}\right) = \begin{cases} 0 & if\ dev_{t_w} > 0 \\ 1 & if\ dev_{t_w} < 0 \end{cases}$$

Thus, to determine the number of times a method undercounts during the 23 week period we use the following sum

$$undercounts = \sum_{w=0}^{23} f\left(dev_{t_w}\right) \tag{5.4}$$

Table 5.4

Undercounts of Model-1 and Model-2

	Undergraduates		Graduates	
	2011	2012	2011	2012
Model 1	9	17	3	9
Model 2	16	15	13	15

Table 5.4 above illustrates Equation 5.4 for undergraduates and graduates for each of the predicted years. As we can observe, Model-2 has a higher tendency to under predict in every example, which would make it the most conservative model as well as the most accurate.

To see the details of the weekly projections of total SCH over a 23 week period for fall 2011 graduate, 2011 undergraduate, and 2012 undergraduate students see Appendix A (projections under Model-1) and Appendix C (projections under Model-2).

Future Research

We noticed that the University of Baltimore developed the concept of a weighted average SCH but did not use it in their projections of total SCH. One of the differences between the model that the University of Baltimore developed and our own version of this approach, is that we used a regular average of SCH, instead of a weighted average SCH. In the following discussion, we will develop a notation for this kind of weighted average SCH and introduce an implicit application of a weighted average SCH for predicting a weighted average SCH that can be developed in future research.

Let μ represent the weighted average SCH of the total number of students. To calculate μ, we started by gathering observed data. Let x represent the sum of the SCH each student completed in a particular semester. For example, suppose student A took a three-hour course, a one-hour course, and a four-hour course, then for student A, $x = 8$. However, another student, student B, may only take a one-hour course in a particular semester so $x = 1$ for student B. In our exploratory analysis, we found that $x = 1, 2\ldots 24$. Thus, we define the sample space of x, the set of all possible values for x, as $S_x = \{1, 2, \ldots, 24\}$. In other words, for any semester of interest, you can find a student whose course load is anywhere from 1 to 24 SCH. In the discussion that follows, we will develop notation to represent the number of students in a semester that take 1 to 24 SCH in a semester.

Let c_x represent the number of students taking x amount of SCH for a particular semester of interest. Table 5.5 illustrates the x and c_x values for fall 2009. As can be seen in Table 5.5, these values of x and c_x, are stratified by undergraduates (U) and graduate students (G) since the enrollment patterns of these two strata differ. Using our notation, $c_1 = 17$ under the c_x column for undergraduates, indicates there were 17 undergraduate students who earned 1 SCH in the fall of 2009. The total number of students for each level in 2009 were N_U = 7,825, and N_G = 5,497 for graduate students.

Table 5.5

Counts and Weights for Each Number of SCH (x) for Undergraduate Students

Number of SCH (*x*)	c_x for 2009 Undergraduates	d_x for 2009 Undergraduates	c_x for 2010 Undergraduates	d_x for 2010 Undergraduates
0	0	0.00000	6	0.00071
1	17	0.00217	13	0.00153
2	10	0.00128	6	0.00071
3	357	0.04562	373	0.04399
4	109	0.01393	115	0.01356
5	33	0.00422	21	0.00248
6	673	0.08601	716	0.08444
7	217	0.02773	206	0.02430
8	110	0.01406	125	0.01474
9	578	0.07387	636	0.07501
10	251	0.03208	289	0.03408
11	143	0.01827	164	0.01934
12	1672	0.21367	1822	0.21488
13	1187	0.15169	1381	0.16287
14	661	0.08447	779	0.09187
15	918	0.11732	978	0.11534
16	511	0.06530	506	0.05968
17	183	0.02339	161	0.01899
18	143	0.01827	134	0.01580
19	48	0.00613	42	0.00495
20	2	0.00026	4	0.00047
21	0	0.00000	2	0.00024
22	1	0.00013	0	0.00000
23	1	0.00013	0	0.00000
Total	N_U =7825		N_U =8479	

Let d_x represent the *weight of each x* which is defined as

$$d_x = \frac{c_x}{N} \tag{5.5}$$

where $0 \leq d_x \leq 1$ and

$$\sum_{x \in S_x} d_x = 1 .$$

Equation 5.5 can be thought of as the probability of a student taking *x* SCH. For example, for an undergraduate taking 12 hours in fall 2009, Equation 3.8 becomes $d_{12} = \frac{c_{12}}{N_U} = \frac{1672}{7825} = 0.21$. Thus, if you randomly select an undergraduate from fall 2009, the probability they would take 12 SCH is 0.22 (or 22%).

Figures 5.1 and 5.2 below show the distribution of d_x for 2009 and 2010 respectively from Table 5.5.

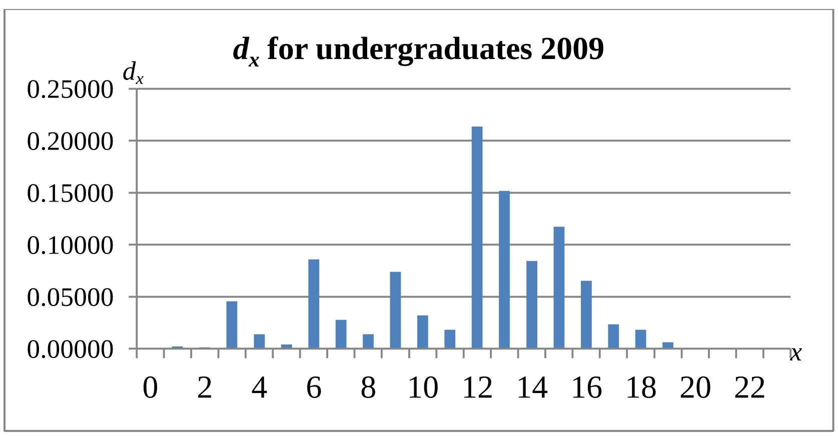

Figure 5.1: Distribution of d_x for 2009 undergraduates

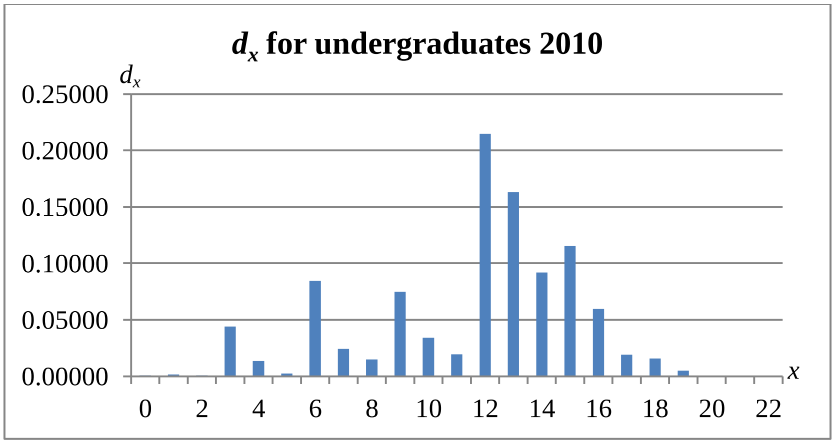

Figure 5.2: Distribution of d_x for 2010 undergraduates

The weight calculations for each number of SCH (*x*), for 2009 and 2010 graduate students can be found in Appendix-E.

Using Equation 5.5, we can rewrite Equation 4.1 as

$$\mu = \sum_{x=1}^{24} x \; d_x \quad . \tag{5.6}$$

Equation 5.6 represents the weighted average of the random variable x where the weight of each x is represented by d_x. Below, we illustrate formula 5.6 for the undergraduate group in Table 5.5 as

$$\mu_{U-2009} = 1 \cdot \left(\frac{17}{7825}\right) + 2 \cdot \left(\frac{10}{7825}\right) + \ldots + 23 \cdot \left(\frac{1}{7825}\right) = 11.605$$

Thus, the observed average semester credit hour taken by undergraduates in fall 2009 was 11.605 hours.

Figure 5.1 and Figure 5.2 show that the distributions of semester credit hours are consistent from one year to the next. Given this consistency we can use this probability distribution of X, represented by the patterns of d_x, to formulate an alternative approach to predict μ by using the observed sample mean, $\bar{X}$, during the weekly prediction period. From the Central Limit Theorem we know that the sampling distribution of the sample mean is centered at the μ of interest. This means that the weekly sample mean values we observe, can be used to make an inference about μ.

REFERENCES

Armstrong, D., Wenckowski Nunley, C. (1981). *Enrollment Projection Within a Decision-Making Framework.* The Journal of Higher Education, Vol. 52, No. 3, May – June 198. Page 295-309.

Callahan, E. (2011). *Enrollment Projection Model.* Office of Institutional Planning, Assessment and Research form Winona State University.

Cameron, J. and McLaughlin, G. (2008). *Modeling Success of Freshmen Transfer Students.* DePaul University; Chicago, Illinois.

Campbell, W., Doan, H. (1982). *Better Managerial Effectiveness Through Improved Planning and Budgeting in the Academic Area.* Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, March 19-23, 1982).

Financing Higher education in Texas. Legislative Primer (January 2011). http://www.lbb.state.tx.us/Higher_Education/Financing%20Higher%20Education%20in%20Texas%20Legislative%20Primer%20112011.pß

Gao, H., Hughes, W., O'Rear, M., Fendley, W. Jr., (2002). *Developing structural equation models to determine factors contributing to student graduation and retention: Are There Differences for Native Students and Transfers?* Paper presented at the Annual Research Forum of the Association for Institutional Research (43rd, Toronto, Ontario, Canada, June 2-5, 2002).

Guo, S., (2002). *Three Enrollment Forecasting Models: Issues in Enrollment Projection for Community Colleges.* Presented at the 40th RP Conference. Research and Planning Chancellor's Office California Community Colleges.

Hamner, M. and Ahluwalia, P. (2007). *Predicting Real-Time Percent Enrollment Increase*. Presentation for the 2007 Texas Association for Institutional Research (TAIR) Conference.

Headcount and Student Credit Hour Projections in support of the Master Facility Plan 2008-2018. Maryland Higher Education Commission’s (MHEC) and the Office of the Provost Institutional Research, University of Baltimore. http://www.ubalt.edu/downloads/Enrollment%20Projections%20for%20the%20Master%20Plan-2-2009.docx

Luo, Williams, Vieweg, (2007). *Transitioning transfer students: Interactive factors that influence first-year retention.* College & University (Volume 83, No. 2).

Nandeshwar, A. and Chaudhari, S. (2009). *Enrollment Prediction Models Using Data Mining*. http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Project.pdf

Overview of the Detailed Enrollment Prediction Model. Office University Analysis & Planning Support at the University of Central Florida. Page last updated in 2011. http://uaps.ucf.edu/enrollment/methods_detailed.html

Ranney, G.B. and Thigpen, C.C. (1981). *The Sample Coefficient of Determination in Simple Linear Regression.* The American Statistician , Vol. 35, No. 3 (Aug., 1981), pp. 152-153. Published by: American Statistical Association

UNC Enrollment Change Funding Formula Needs Documentation and a Performance Component. Final Report to the Joint Legislative Program Evaluation Oversight Committee. Report Number 2010-05, November 17, 2010 by the Program Evaluation Division, North Carolina Assembly.

Tsui, P., Murdock, T., & Mayer, L. (1997). *Trend analysis and enrollment management*. ERIC Database: ED 410 887. Paper presented at the Annual Forum of the Association for Institutional Research (37th, Orlando, FL, May 18-21, 1997).

Vaughan, P. and Ormerod, S.J. (2005). *The Continuing Challenges of Testing Species Distribution Models*. Journal of Applied Ecology , Vol. 42, No. 4 (Aug., 2005), pp. 720-730. Published by: British Ecological Society.

APPENDIX A

Total SCH Calculations under Model-1

Total SCH Calculation under Model-1

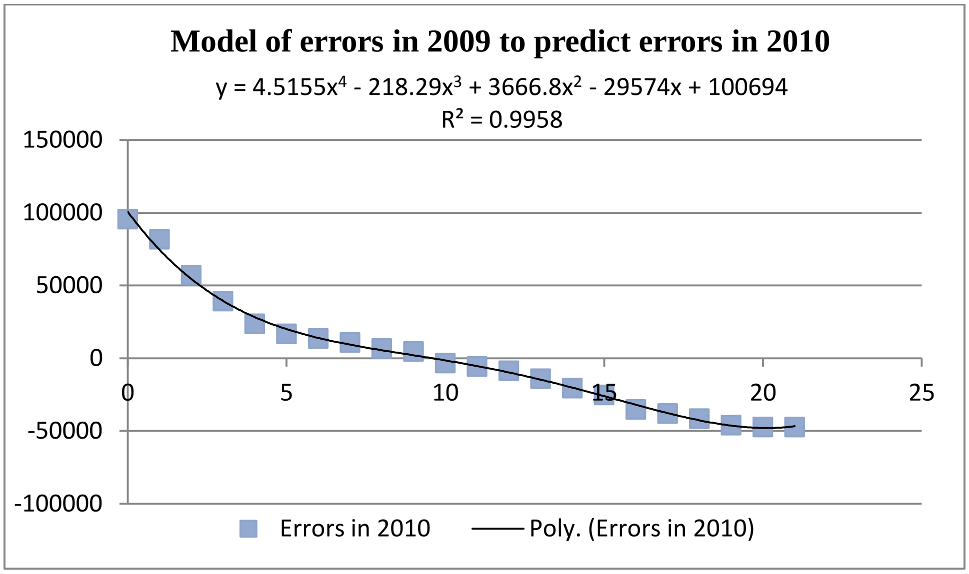

Figure A.1: Model of 2010 SCH for undergraduate students

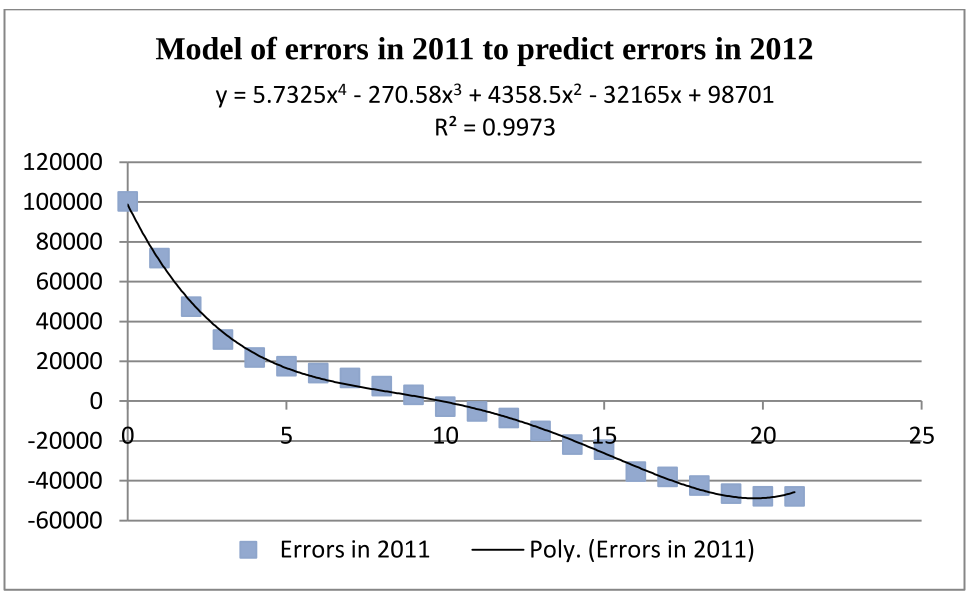

Figure A.2: Model of 2011 SCH errors for undergraduate students

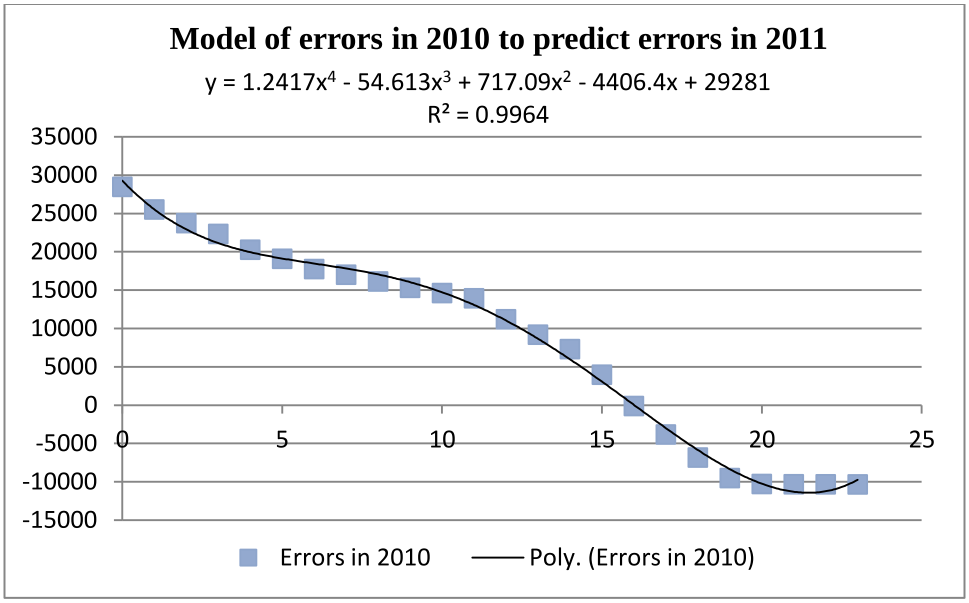

Figure A.3: Model of 2010 SCH errors for graduate students

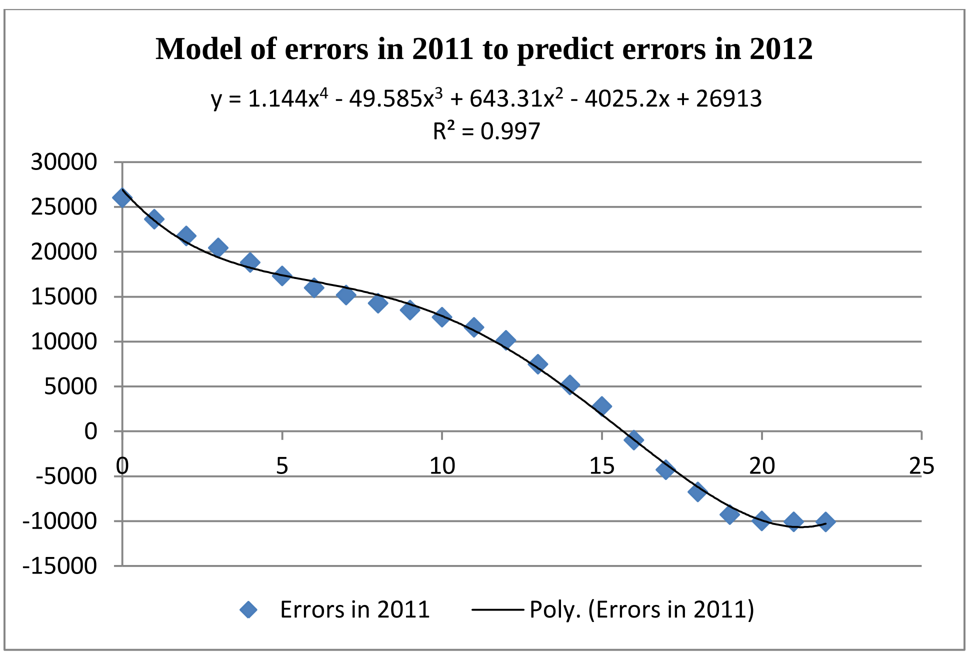

Figure A.4: Model of 2011 SCH errors for graduate students

Table A.1

Total SCH for Undergraduate Students for Each Week of 2011

Week	Sum T^{t_w}	Predicted Diff $\hat{T}^{t_w}{}'$	Predicted $\hat{T}$	Actual T	Off $\hat{T} - T$	off%
0	4558	100694.00	105252.00	104857	395.00	0%
1	33024	74573.03	107597.03	104857	2740.03	3%
2	57416	54539.13	111955.13	104857	7098.13	7%
3	73911	39445.13	113356.13	104857	8499.13	8%
4	82941	28252.21	111193.21	104857	6336.21	6%
5	87311	20029.94	107340.94	104857	2483.94	2%
6	90774	13956.25	104730.25	104857	-126.75	0%
7	93206	9317.45	102523.45	104857	-2333.55	-2%
8	97306	5508.21	102814.21	104857	-2042.79	-2%
9	101689	2031.59	103720.59	104857	-1136.41	-1%
10	107627	-1501.00	106126.00	104857	1269.00	1%
11	110031	-5369.75	104661.25	104857	-195.75	0%
12	113315	-9746.51	103568.49	104857	-1288.51	-1%
13	119789	-14694.73	105094.27	104857	237.27	0%
14	126669	-20169.51	106499.49	104857	1642.49	2%
15	129222	-26017.56	103204.44	104857	-1652.56	-2%
16	140233	-31977.23	108255.77	104857	3398.77	3%
17	142891	-37678.49	105212.51	104857	355.51	0%
18	147238	-42642.95	104595.05	104857	-261.95	0%
19	151300	-46283.83	105016.17	104857	159.17	0%
20	152632	-47906.00	104726.00	104857	-131.00	0%
21	152688	-46705.93	105982.07	104857	1125.07	1%

Table A.2

Total SCH for Undergraduate Students for Each Week of 2012

Week	Sum T^{t_w}	Predicted Diff $\hat{T}^{t_w}{}'$	Predicted $\hat{T}$	Actual T	Off $\hat{T} - T$	off%
0	5056	98701.00	103757.00	109487	-5730.00	-5%
1	19492	70629.65	90121.65	109487	-19365.35	-18%
2	44337	49732.08	94069.08	109487	-15417.92	-14%
3	61594	34591.17	96185.17	109487	-13301.83	-12%
4	81967	23927.40	105894.40	109487	-3592.60	-3%
5	87806	16598.81	104404.81	109487	-5082.19	-5%
6	91745	11601.04	103346.04	109487	-6140.96	-6%
7	94176.34	8067.29	102243.63	109487	-7243.37	-7%
8	99734.34	5268.36	105002.70	109487	-4484.30	-4%
9	106785.34	2612.61	109397.95	109487	-89.05	0%
10	113431.34	-354.00	113077.34	109487	3590.34	3%
11	115530.34	-3947.95	111582.39	109487	2095.39	2%
12	118190.34	-8348.12	109842.22	109487	355.22	0%
13	120925.34	-13595.83	107329.51	109487	-2157.49	-2%
14	131345.34	-19594.80	111750.54	109487	2263.54	2%
15	135426.34	-26111.19	109315.15	109487	-171.85	0%
16	142939.34	-32773.56	110165.78	109487	678.78	1%
17	146285.34	-39072.91	107212.43	109487	-2274.57	-2%
18	149979.34	-44362.64	105616.70	109487	-3870.30	-4%
19	154039.34	-47858.59	106180.75	109487	-3306.25	-3%
20	154525.34	-48639.00	105886.34	109487	-3600.66	-3%
21	154535.34	-45644.55	108890.79	109487	-596.21	-1%

Table A.3

Total SCH for Graduate Students for Each Week of 2011

Week	Sum T^{t_w}	Predicted Diff $\hat{T}^{t_w\prime}$	Predicted $\hat{T}$	Actual T	Off $\hat{T} - T$	off%
0	10274	36296	29281.00	39555.00	3259.00	9%
1	12666	36296	25538.32	38204.32	1908.32	5%
2	14534	36296	22919.52	37453.52	1157.52	3%
3	15864	36296	21141.64	37005.64	709.64	2%
4	17487	36296	19951.48	37438.48	1142.48	3%
5	18988	36296	19125.69	38113.69	1817.69	5%
6	20312	36296	18470.68	38782.68	2486.68	7%
7	21117	36296	17822.67	38939.67	2643.67	7%
8	22031	36296	17047.71	39078.71	2782.71	8%
9	22793	36296	16041.61	38834.61	2538.61	7%
10	23582	36296	14730.00	38312.00	2016.00	6%
11	24715	36296	13068.32	37783.32	1487.32	4%
12	26160	36296	11041.79	37201.79	905.79	2%
13	28823	36296	8665.44	37488.44	1192.44	3%
14	31127	36296	5984.12	37111.12	815.12	2%
15	33522	36296	3072.44	36594.44	298.44	1%
16	37273	36296	34.84	37307.84	1011.84	3%
17	40578	36296	-2994.43	37583.57	1287.57	4%
18	43049	36296	-5851.36	37197.64	901.64	2%
19	45577	36296	-8342.09	37234.91	938.91	3%
20	46277	36296	-10243.00	36034.00	-262.00	-1%
21	46379	36296	-11300.65	35078.35	-1217.65	-3%
22	46391	36296	-11231.79	35159.21	-1136.79	-3%

APPENDIX B

Details of the Criterion to Select a Model to Predict the Average SCH

Details of the Criterion to Select a Model to Predict the Average SCH

1. Undergraduate 2008-2010 pattern

Power model

$$\hat{\mu}(t_y) = \frac{11.784}{t_y^{0.016}}$$

$$\hat{\mu}'(t_y) = -\frac{0.188544}{t_y^{1.016}}, \hat{\mu}'(3) = -0.0618$$

Logarithmic model

$$\hat{\mu}(t_y) = -0.187\ln(t_y) + 11.784$$

$$\hat{\mu}'(t_y) = -\frac{0.187}{t_y}, \hat{\mu}'(3) = -0.062$$

Exponential model

$$\hat{\mu}(t_y) = 11.866e^{-0.008t_y}$$

$$\hat{\mu}'(t_y) = -0.094928e^{-0.008t_y}, \hat{\mu}'(3) = -0.093$$

2. Undergraduate 2009-2011 pattern

Power model

$$\hat{\mu}(t_y) = 11.601t_y^{0.0024}$$

$$\hat{\mu}'(t_y) = \frac{0.0278424}{t_y^{0.9976}}, \hat{\mu}'(3) = 0.00931$$

Logarithmic model

$$\hat{\mu}(t_y) = 0.0277\log(t_y) + 11.601$$

$$\hat{\mu}'(t_y) = \frac{0.0277}{t_y}, \hat{\mu}'(3) = 0.0092$$

Exponential model

$$\hat{\mu}(t_y) = 11.585e^{0.0014t_y}$$

$$\hat{\mu}'(t_y) = 0.016219e^{0.0014t_y}, \hat{\mu}'(3) = 0.016$$

3. Graduate 2008-2010 pattern

Power model

$$\hat{\mu}(t_y) = 6.9178 t_y^{-0.052}$$

$$\hat{\mu}'(t_y) = -\frac{0.359726}{t_y^{1.052}}, \hat{\mu}'(3) = -0.113$$

Logarithmic model

$$\hat{\mu}(t_y) = -0.339 \log(t_y) + 6.915$$

$$\hat{\mu}'(t_y) = -\frac{0.339}{t_y}, \hat{\mu}'(3) = -0.113$$

Exponential model

$$\hat{\mu}(t_y) = 7.1809 e^{-0.034 t_y}$$

$$\hat{\mu}'(t_y) = -0.244151 e^{-0.034 t_y}, \hat{\mu}'(3) = -0.220$$

4. Graduate 2009-2011 pattern

Power model

$$\hat{\mu}(t_y) = 6.9145 t_y^{-0.089}$$

$$\hat{\mu}'(t_y) = -\frac{0.615391}{t_y^{1.089}}, \hat{\mu}'(3) = -0.186$$

Logarithmic model

$$\hat{\mu}(t_y) = -0.593 \log(t_y) + 6.9169$$

$$\hat{\mu}'(t_y) = -\frac{0.593}{t_y}, \hat{\mu}'(3) = -0.198$$

Exponential model

$$\hat{\mu}(t_y) = 7.1882 e^{-0.046 t_y}$$

$$\hat{\mu}'(t_y) = -0.330657 e^{-0.046 t_y}, \hat{\mu}'(3) = -0.288$$

APPENDIX C

Total Headcount Projections for 2011 and 2012

Total Headcount Projections for 2011 and 2012

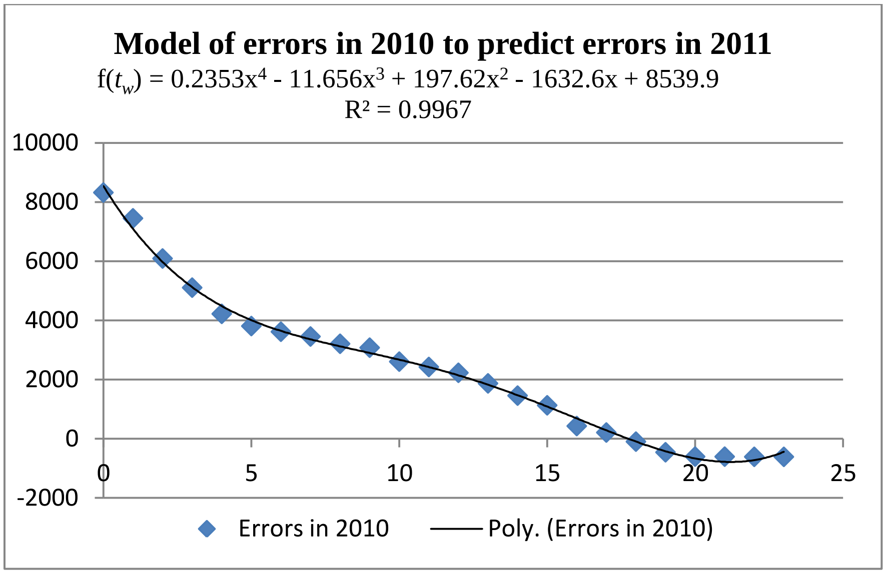

Figure C.1: Model of errors in 2010 to predict errors in 2011 for undergraduates

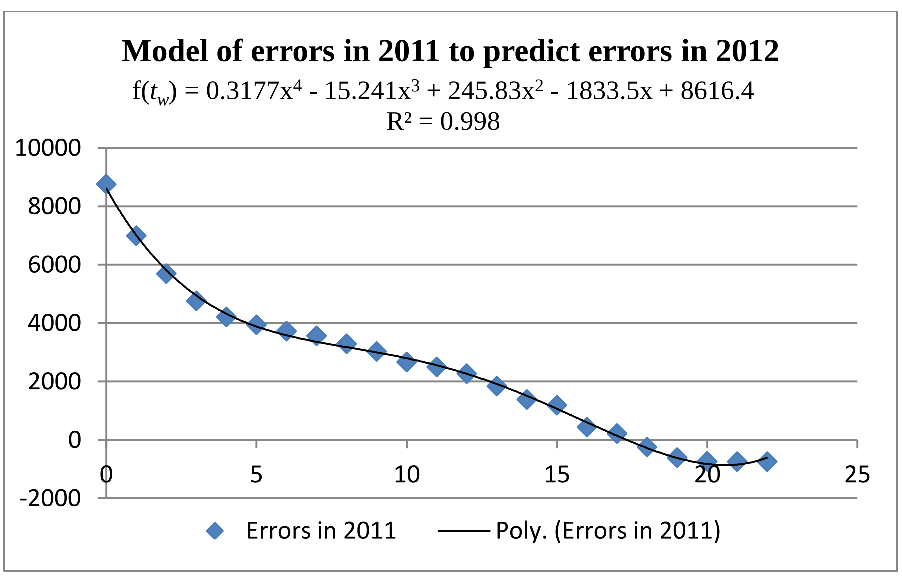

Figure C.2: Model of errors in 2011 to predict errors in 2012 for undergraduates

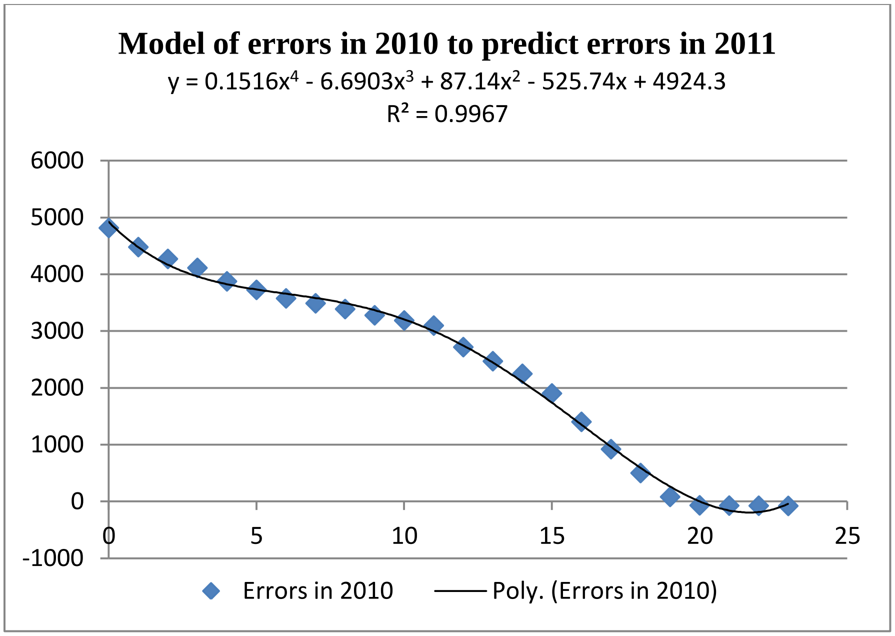

Figure C.3: Model of errors in 2010 to predict errors in 2011 for graduates

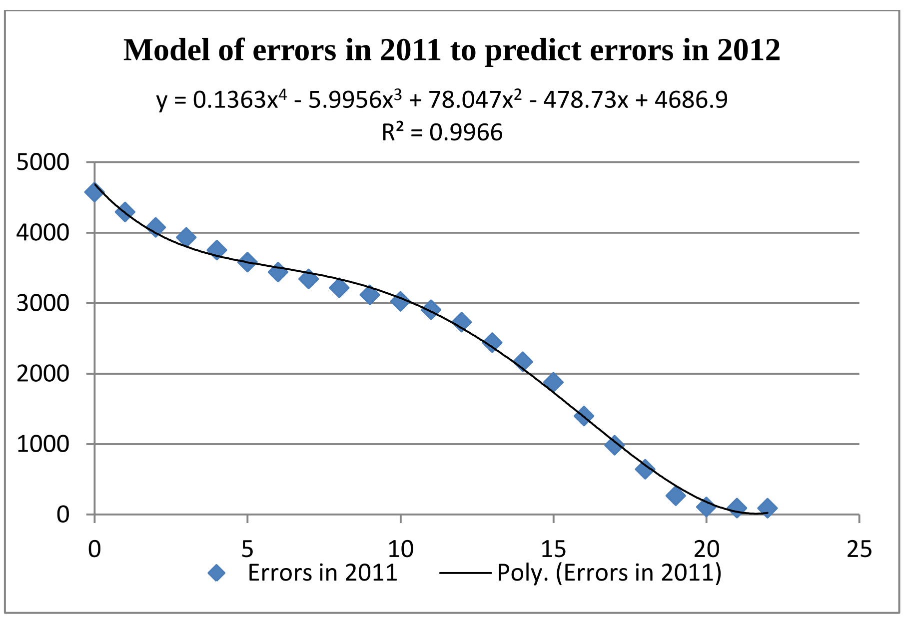

Figure C.4: Model of errors in 2011 to predict errors in 2012 for graduates

Table C.1

Headcount Projection for Fall 2011 Undergraduate Students

Week t_w	Sum n^{t_w}	Actual N	Predicted Difference $\hat{n}^{t_w'}$	Predicted Headcount $\hat{N}$	Off by $\hat{N} - N$	Off by %
0	251	9010	8539.90	8790.90	-219.10	-2%
1	2017	9010	7093.50	9110.50	100.50	1%
2	3316	9010	5975.70	9291.70	281.70	3%
3	4247	9010	5125.03	9372.03	362.03	4%
4	4798	9010	4485.67	9283.67	273.67	3%
5	5067	9010	4007.46	9074.46	64.46	1%
6	5284	9010	3645.87	8929.87	-80.13	-1%
7	5445	9010	3362.03	8807.03	-202.97	-2%
8	5718	9010	3122.70	8840.70	-169.30	-2%
9	5980	9010	2900.30	8880.30	-129.70	-1%
10	6341	9010	2672.90	9013.90	3.90	0%
11	6510	9010	2424.21	8934.21	-75.79	-1%
12	6740	9010	2143.59	8883.59	-126.41	-1%
13	7166	9010	1826.05	8992.05	-17.95	0%
14	7625	9010	1472.24	9097.24	87.24	1%
15	7820	9010	1088.46	8908.46	-101.54	-1%
16	8569	9010	686.66	9255.66	245.66	3%
17	8791	9010	284.44	9075.44	65.44	1%
18	9254	9010	-94.96	9159.04	149.04	2%
19	9616	9010	-422.65	9193.35	183.35	2%
20	9746	9010	-664.10	9081.90	71.90	1%
21	9753	9010	-779.12	8973.88	-36.12	0%
22	9754	9010	-721.87	9032.13	22.13	0%

Table C.2

Headcount Projection for Fall 2012 Undergraduate Students

Week t_w	Sum n^{t_w}	Actual N	Predicted Difference $\hat{n}^{t_w'}$	Predicted Headcount $\hat{N}$	Off by $\hat{N} - N$	Off by %
0	289	9443	8616.40	8905.40	-537.60	-6%
1	1279	9443	7013.81	8292.81	-1150.19	-12%
2	2725	9443	5815.88	8540.88	-902.12	-10%
3	3736	9443	4942.60	8678.60	-764.40	-8%
4	4920	9443	4321.59	9241.59	-201.41	-2%
5	5288	9443	3888.09	9176.09	-266.91	-3%
6	5543	9443	3584.96	9127.96	-315.04	-3%
7	5709	9443	3362.70	9071.70	-371.30	-4%
8	6068	9443	3179.43	9247.43	-195.57	-2%
9	6503	9443	3000.87	9503.87	60.87	1%
10	6938	9443	2800.40	9738.40	295.40	3%
11	7096	9443	2559.00	9655.00	212.00	2%
12	7284	9443	2265.30	9549.30	106.30	1%
13	7493	9443	1915.52	9408.52	-34.48	0%
14	8176	9443	1513.54	9689.54	246.54	3%
15	8476	9443	1070.84	9546.84	103.84	1%
16	9033	9443	606.53	9639.53	196.53	2%
17	9304	9443	147.36	9451.36	8.36	0%
18	9710	9443	-272.32	9437.68	-5.32	0%
19	10070	9443	-610.51	9459.49	16.49	0%
20	10126	9443	-817.60	9308.40	-134.60	-1%
21	10128	9443	-836.36	9291.64	-151.36	-2%
22	10128	9443	-601.92	9526.08	83.08	1%
23	10128	9443	-41.79	10086.21	643.21	7%

Table C.3

Headcount Projection for Fall 2011 Graduate Students

Week t_w	Sum n^{t_w}	Actual N	Predicted Difference $\hat{n}^{t_w'}$	Predicted Headcount $\hat{N}$	Off by $\hat{N} - N$	Off by %	Week t_w
0	1132	5708	4576	4924.30	6056.30	348.30	6%
1	1414	5708	4294	4479.16	5893.16	185.16	3%
2	1633	5708	4075	4170.28	5803.28	95.28	2%
3	1775	5708	3933	3962.98	5737.98	29.98	1%
4	1955	5708	3753	3826.21	5781.21	73.21	1%
5	2126	5708	3582	3732.56	5858.56	150.56	3%
6	2267	5708	3441	3658.27	5925.27	217.27	4%
7	2365	5708	3343	3583.20	5948.20	240.20	4%
8	2490	5708	3218	3490.86	5980.86	272.86	5%
9	2590	5708	3118	3368.40	5958.40	250.40	4%
10	2683	5708	3025	3206.60	5889.60	181.60	3%
11	2802	5708	2906	2999.89	5801.89	93.89	2%
12	2978	5708	2730	2746.32	5724.32	16.32	0%
13	3269	5708	2439	2447.60	5716.60	8.60	0%
14	3539	5708	2169	2109.06	5648.06	-59.94	-1%
15	3831	5708	1877	1739.69	5570.69	-137.31	-2%
16	4310	5708	1398	1352.09	5662.09	-45.91	-1%
17	4724	5708	984	962.52	5686.52	-21.48	0%
18	5065	5708	643	590.87	5655.87	-52.13	-1%
19	5441	5708	267	260.68	5701.68	-6.32	0%
20	5600	5708	108	-0.90	5599.10	-108.90	-2%
21	5616	5708	92	-163.05	5452.95	-255.05	-4%
22	5618	5708	90	-191.32	5426.68	-281.32	-5%

APPENDIX D

Total SCH Projections for 2011 and 2012
under Model-2

Total SCH Projections for 2011 and 2012 under Model-2

Table D.1

Total SCH for Undergraduate Students for each Preregistration Week of 2011

Week	$\hat{N}$	$\hat{\mu}$	$\hat{T}$	T	Off $\hat{T} - T$	Off %
0	8790.90	11.52476	101313.04	104857	3543.96	3%
1	9110.50	11.52476	104996.34	104857	-139.34	0%
2	9291.70	11.52476	107084.60	104857	-2227.60	-2%
3	9372.03	11.52476	108010.39	104857	-3153.39	-3%
4	9283.67	11.52476	106992.13	104857	-2135.13	-2%
5	9074.46	11.52476	104581.03	104857	275.97	0%
6	8929.87	11.52476	102914.67	104857	1942.33	2%
7	8807.03	11.52476	101498.90	104857	3358.10	3%
8	8840.70	11.52476	101886.93	104857	2970.07	3%
9	8880.30	11.52476	102343.34	104857	2513.66	2%
10	9013.90	11.52476	103883.06	104857	973.94	1%
11	8934.21	11.52476	102964.67	104857	1892.33	2%
12	8883.59	11.52476	102381.30	104857	2475.70	2%
13	8992.05	11.52476	103631.26	104857	1225.74	1%
14	9097.24	11.52476	104843.54	104857	13.46	0%
15	8908.46	11.52476	102667.92	104857	2189.08	2%
16	9255.66	11.52476	106669.34	104857	-1812.34	-2%
17	9075.44	11.52476	104592.33	104857	264.67	0%
18	9159.04	11.52476	105555.77	104857	-698.77	-1%
19	9193.35	11.52476	105951.15	104857	-1094.15	-1%
20	9081.90	11.52476	104666.74	104857	190.26	0%
21	8973.88	11.52476	103421.88	104857	1435.12	1%
22	9032.13	11.52476	104093.14	104857	763.86	1%

Table D.2

Total SCH for Undergraduate Students for each Preregistration Week of 2012

Week	$\hat{N}$	$\hat{\mu}$	$\hat{T}$	T	Off $\hat{T}-T$	Off %
0	8905.40	11.56260	102969.57	109487	6517.43	6%
1	8292.81	11.56260	95886.40	109487	13600.60	12%
2	8540.88	11.56260	98754.72	109487	10732.28	10%
3	8678.60	11.56260	100347.14	109487	9139.86	8%
4	9241.59	11.56260	106856.77	109487	2630.23	2%
5	9176.09	11.56260	106099.43	109487	3387.57	3%
6	9127.96	11.56260	105542.98	109487	3944.02	4%
7	9071.70	11.56260	104892.49	109487	4594.51	4%
8	9247.43	11.56260	106924.30	109487	2562.70	2%
9	9503.87	11.56260	109889.45	109487	-402.45	0%
10	9738.40	11.56260	112601.22	109487	-3114.22	-3%
11	9655.00	11.56260	111636.95	109487	-2149.95	-2%
12	9549.30	11.56260	110414.72	109487	-927.72	-1%
13	9408.52	11.56260	108786.98	109487	700.02	1%
14	9689.54	11.56260	112036.26	109487	-2549.26	-2%
15	9546.84	11.56260	110386.26	109487	-899.26	-1%
16	9639.53	11.56260	111458.04	109487	-1971.04	-2%
17	9451.36	11.56260	109282.28	109487	204.72	0%
18	9437.68	11.56260	109124.15	109487	362.85	0%
19	9459.49	11.56260	109376.33	109487	110.67	0%
20	9308.40	11.56260	107629.30	109487	1857.70	2%
21	9291.64	11.56260	107435.54	109487	2051.46	2%
22	9526.08	11.56260	110146.29	109487	-659.29	-1%
23	10086.21	11.56260	116622.79	109487	-7135.79	-7%

Table D.3

Total SCH for Graduate Students for each Preregistration Week of 2011

Week	$\hat{N}$	$\hat{\mu}$	$\hat{T}$	T	Off $\hat{T}-T$	Off %
0	6056.30	6.26780	37959.65	36296	-1663.65	-5%
1	5893.16	6.26780	36937.13	36296	-641.13	-2%
2	5803.28	6.26780	36373.79	36296	-77.79	0%
3	5737.98	6.26780	35964.50	36296	331.50	1%
4	5781.21	6.26780	36235.45	36296	60.55	0%
5	5858.56	6.26780	36720.27	36296	-424.27	-1%
6	5925.27	6.26780	37138.37	36296	-842.37	-2%
7	5948.20	6.26780	37282.09	36296	-986.09	-3%
8	5980.86	6.26780	37486.81	36296	-1190.81	-3%
9	5958.40	6.26780	37346.03	36296	-1050.03	-3%
10	5889.60	6.26780	36914.81	36296	-618.81	-2%
11	5801.89	6.26780	36365.04	36296	-69.04	0%
12	5724.32	6.26780	35878.86	36296	417.14	1%
13	5716.60	6.26780	35830.47	36296	465.53	1%
14	5648.06	6.26780	35400.90	36296	895.10	2%
15	5570.69	6.26780	34915.93	36296	1380.07	4%
16	5662.09	6.26780	35488.82	36296	807.18	2%
17	5686.52	6.26780	35641.94	36296	654.06	2%
18	5655.87	6.26780	35449.85	36296	846.15	2%
19	5701.68	6.26780	35736.94	36296	559.06	2%
20	5599.10	6.26780	35094.01	36296	1201.99	3%
21	5452.95	6.26780	34177.98	36296	2118.02	6%
22	5426.68	6.26780	34013.29	36296	2282.71	6%

APPENDIX E

Weight Calculations for *x* Number of SCH for Graduate Students

Weight Calculations for *x* Number of SCH for Graduate Students

Table E.1

Counts and Weights for Each Number of SCH (x) for Graduate Students

Number of SCH (x)	c_x for 2009 Graduates	d_x for 2009 Graduates	c_x for 2010 Graduates	d_x for 2010 Graduates
0	1	0.00018	5	0.00088
1	18	0.00327	38	0.00667
2	13	0.00236	17	0.00298
3	1367	0.24868	1951	0.34252
4	93	0.01692	86	0.01510
5	65	0.01182	84	0.01475
6	1563	0.28434	1547	0.27159
7	151	0.02747	221	0.03880
8	107	0.01947	126	0.02212
9	1253	0.22794	737	0.12939
10	136	0.02474	167	0.02932
11	89	0.01619	111	0.01949
12	278	0.05057	247	0.04336
13	43	0.00782	40	0.00702
14	171	0.03111	163	0.02862
15	126	0.02292	109	0.01914
16	6	0.00109	34	0.00597
17	13	0.00236	10	0.00176
18	2	0.00036	3	0.00053
19	0	0.00000	0	0.00000
20	2	0.00036	0	0.00000
21	0	0.00000	0	0.00000
22	0	0.00000	0	0.00000
23	0	0.00000	0	0.00000
Total	N_U =5497		N_U =5696	

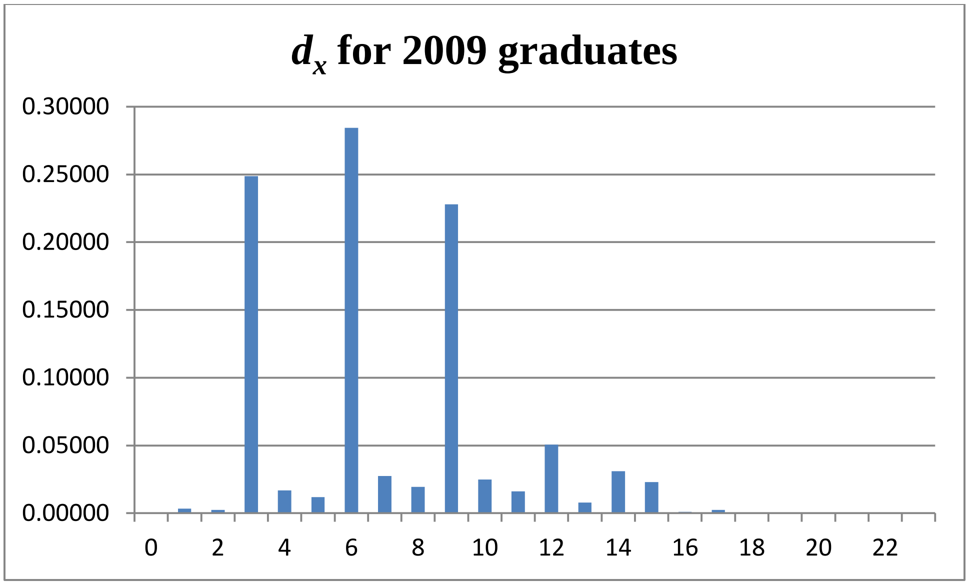

Figure E.1: Distribution of dx for 2009 graduates

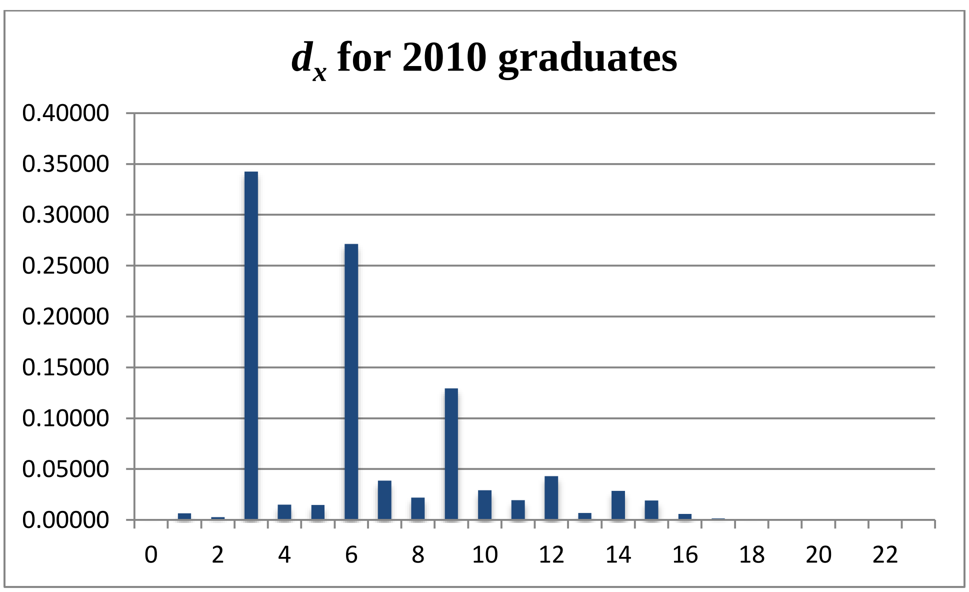

Figure E.2: Distribution of dx for 2010 graduates