

IR Applications

Volume 15, January 18, 2008

Using Advanced Tools, Techniques, and Methodologies

An Integrated Enrollment Forecast Model

Dr. Chau-Kuang Chen,
Meharry Medical College

Association for
Institutional Research



Enhancing knowledge.
Expanding networks.

Abstract

Enrollment forecasting is the central component of effective budget and program planning. The integrated enrollment forecast model is developed to achieve a better understanding of the variables affecting student enrollment and, ultimately, to perform accurate forecasts. The transfer function model of the autoregressive integrated moving average (ARIMA) methodology and linear regression model are major forecasting techniques. The structural approach embedded in the models allows the researcher to construct candidate models, eliminate inappropriate ones, and retain the most suitable model. In addition, the expert system for the ARIMA model is a supplementary tool used to verify the resulting models in terms of model structure and forecasting accuracy.

The enrollment series of interest is the 1962 – 2004 student enrollment for Oklahoma State University (OSU). Fifteen independent variables are used in an attempt to increase explanatory power. These variables include demographics (Oklahoma high school graduates and competitor college enrollment from the University of Oklahoma), state funding, economic indicators, (e.g., state unemployment rate and gross national product), and

one-year lagged demographics and economic indicators.

The best ARIMA and linear regression models yield remarkably high R-squared values and exceptionally small mean absolute percentage errors (MAPEs), respectively. Moreover, they contain two identical demographics: Oklahoma high school graduates and one-year lagged OSU enrollment. Hence, the first-order autoregressive models appropriately depict the longitudinal and aggregated OSU enrollment series. An additional linear regression model shows that one-year lagged Oklahoma high school graduates and three economic indicators significantly contribute to OSU enrollment. This integrated enrollment forecast model has demonstrated its model validity and accuracy. Hence, it could be replicated for comparable universities elsewhere.

Introduction

Student enrollment translated into fiscal income is fundamentally important to budget, program, and personnel planning. Accurate enrollment forecasts are crucial for colleges and universities to remain competitive while inaccurate enrollment forecasts can lead to poor allocation of funds and resources. The upward

enrollment trend in the current decade coupled with unstable growth creates a critical demand for accurate enrollment forecasting. This presents challenges to researchers who must choose from a wide variety of possible influential factors and forecasting techniques.

Student enrollment is a key element in the determination of the funding levels and the capital outlay funds that the state legislature appropriates for institutions of higher education. Projected enrollments for upcoming fiscal years are used to calculate the magnitude of workloads and to estimate support funds required in the areas of teaching, administrative support, and facilities. Thus, accurate enrollment forecasting directly influences budget and program planning in colleges and universities.

Enrollment forecasts are difficult to make in periods of irregular enrollment patterns when turning points are unexpected. There are enormous challenges for researchers to determine which methods should be integrated and to develop an accurate forecasting model. Decision-makers need high quality enrollment forecasts to appropriately ascertain demands for programs and services.

Various factors ranging from demographics to economic climates have influenced higher education enrollment. It is impossible to forecast the changing figures of student enrollment accurately without prior knowledge of these influential factors. To build one or more forecast models for a particular institution, important factors such as demographics, funding policies, and economic indicators should be considered. Good forecasting generally calls for the use of integrated, logical, and analytical techniques.

In the following discussion, the linear regression technique and the transfer function model of the autoregressive integrated moving

average (ARIMA) methodology are used to compute enrollment forecasts for Oklahoma State University. Comparisons are made on the degree to which the two methods fit the data. Both techniques are indispensable because they are capable of explaining relationships among variables. In addition, the results from the two methods can be compared in terms of model structure and forecasting errors, and model accuracy. The ultimate goal of developing the integrated enrollment forecast model is to decipher the stories hidden in enrollment data. Such an understanding promotes the effectiveness of budget and program planning and the efficiency of resource allocation.

Possible Factors Affecting Student Enrollment

Researchers in higher education are unable to forecast the change in enrollment patterns unless their studies incorporate mechanisms that explain reasons for changes in enrollment (Mangelson, et. al., 1973). Hence, the ability to forecast enrollment accurately is quite dependent on the ability to select appropriate variables and to establish the relationships among these variables.

The number of high school graduates affects college student enrollment (Clagett, 1989; Gerald & Hussar, 2007; Lins, 1960; NCES, 1994; Song & Chissom 1994; Wing, 1974). Certain migration statistics, such as state and regional net in-migration and out-migration of students affect first-year college enrollment (Barbett 1996; Greiner & Girardi, 2006; Texas State Higher Education Coordinating Board, 2001). In fact, nearly 17% of 2.22 million first-time freshmen migrated between states according to the resident report of the 1996 Integrated Postsecondary Education Data System (IPEDS) (Barbett 1996).

Also, the post-war baby boom (those born between 1946 and 1964) may be another factor affecting higher education enrollment (Stapleton & Young, 1988; Wachter & Wascher 1984; Wagschall, 1983).

Factors which can affect enrollment patterns include the condition of the economy, the rate of increase in college tuition relative to growth in family income, trends in federal and state financial aid, and employment prospects for new graduates (Folger, 1974; Heller, 1999; McPherson & Schapiro, 1991). Economic indicators also include per capita personal income, unemployment rate, the competitor college's tuition and fees, and financial aid. (Brinkman & McIntyre, 1997; Stewart & Kate, 1978; Witkowski, 1974). Changes in disposable income per capita are indicators of the changing ability to pay for college.

Undoubtedly, forecasting enrollment is an even more difficult task because of the variety and uncertainty of external factors involved. For instance, domestic or international crises and changes in federal or state government policies can impact the enrollments of a given institution (Crossland, 1980). In addition, the more internal factors such as quality and diversity of programs, location, prestige, price relative to competitors, and recruitment policies will also affect student enrollment among the various institutions (Breneman, 1981). The historical data associated with the above factors are difficult to acquire.

A comprehensive list of factors has been compiled for studying the development of student enrollment forecast models (Breneman, 1984; Brinkman & McIntyre, 1997; Crossland, 1980; Folger, 1974; Lins, 1960; Maganell, 1980; McPherson & Schapiro, 1991; Stewart & Kate, 1978; Wagschall, 1983). These factors include: admission policy, programs, instructional facilities, high school graduates, post-baccalaureate

students, veteran enrollment, related economic structure, international situation, birth rates, mortality rates, migration, education benefits and costs, and scholarship programs. The feasibility of performing enrollment forecasts depends on the ability to identify and to measure the appropriate factors that influence that enrollment.

Techniques Used for Enrollment Forecasting

Forecasting techniques can be categorized into nine major methods: subjective judgment, ratio method, cohort survival study, Markov transition model, neural network model, simulation model, time series analysis, fuzzy time series analysis, and regression analysis. The choice of forecasting technique depends on the availability of data, user skills, appropriateness of method, cost and usability of the software packages. The following discusses these different methodologies.

Subjective Judgment

Although subjective judgment is not an analytical method, it can be used as a qualitative procedure for enrollment forecasting. The subjective estimates of influential factors can be implemented when objective measures or any mathematical models are not available (Brown, 1978; Jennings & Young, 1988; Wing, 1974). As another example, in the Delphi process enrollment management experts essentially debate how future events will affect enrollment forecasting (Brinkman & McIntyre, 1997; Faherty, 1997).

Ratio Method

To forecast future enrollment, one should consider the effect of high school graduates on previous enrollment. The ratio method computes the ratio of entering students to high school

graduates the preceding year (Wing, 1974). The proportion of college freshman enrollment from high school graduates is calculated from historical data. The projected freshman enrollment is then determined by multiplying the above proportion by the projected number of high school graduates.

Cohort Survival Study

In a cohort survival study, the number of students enrolled from a given cohort is estimated for the future by multiplying the survival rate of the cohort times the number in that cohort the prior year (Shaw, 1984). The technique is based on the assumption that survival ratio for a given cohort and the corresponding enrollment will be repeated in the future (Lyell & Toole, 1974). For example, if 83% of freshmen return, then the number returning can be estimated as 83% times the current number of freshmen. Of course there would also need to be an estimate of how many become sophomores and how many return as freshmen.

Markov Transition Model

The Markov transition model predicts the probabilities of future occurrence based on currently known probabilities (Render & Stair, Jr., 2000). It is a tracking technique that produces the transition matrix from one year to the next based on the probabilities in the state transition matrix (i.e., enrolled vs. not enrolled) (Anderson, et. al., 2000; Armacost & Wilson, 2002; Donhart, 1995). It can have various states of enrolling and for not enrolling. It does not have a multiple time period history, thus, the fraction of students in the current class depends only on that of the class of students in the immediately preceding time period. It is similar to the cohort survival methodology but can also include returning students who have the probability of continuing based on the level at which they returned.

Neural Network Model

The neural network model processes information in parallel and non-linear capabilities, which can be learned when it has been trained with some data based exemplars (Huang & Yu 2005). The weight of the connection from input neuron (state) to hidden neuron and the weight of hidden neuron to output neuron can be empirically adjusted based on a minimum of the mean square error (Skapura, 1995). Generally one runs the analysis on part of the data (training) and then uses the resulting model to estimate the likely outcomes for those observations not in the training computation. The quality of the model is then evaluated based on the ability to estimate the outcomes for the individuals in the holdout sample.

Simulation Method

The simulation method is typically a complex model with mathematical relationships for the key individual components. Inputs are then modified and the result is simulated based on the mathematical relationships. This method can be utilized to assess "what-if" scenarios such as changes in state or federal funding, financial aid amount, and tuition on student enrollment. It is a useful technique for goal-seeking that calculates a formula in reverse to evaluate independent variables to obtain the desired enrollment. For example, if the goal is to have 6,500 students, various inputs can be modified until this number is obtained. Often the method can be used to look at different strategies for achieving the goal. For this technique has been found to be useful in investigating the effect of a baby boom on student enrollment (Stapleton & Young, 1988; Wachter & Wascher, 1984).

Time Series Analysis

Time series is a collection of data points gathered sequentially through equally spaced time periods. Data points close together in time are usually expected to correlate with one another. The correlation from one period to another is employed to make reliable forecasts

(Diggle, 2004; Mabert, 1975; Vandaele, 1983). Thus, the assumption of time series forecasting is that the future depends upon the present while the present depends on the past (Brinkman & McIntyre, 1997; Jennings & Young, 1988; Vandaele, 1983). The problem with estimation of relationships comes from the fact that errors may not independent between the adjacent times since each error term may contain part of the errors made in the preceding times.

Not all lagged time series have autocorrelated errors but this needs to be empirically evaluated. Time series techniques include but are not limited to simple exponential smoothing, Holt's two-parameter exponential smoothing, and Box-Jenkins ARIMA methodology.

The simple exponential smoothing technique assumes that the most recent observations contain the most information about the level of what is likely to occur in the future (Makridakis & Wheelwright, 1989). Holt's two-parameter exponential smoothing model is similar in principle to the simple exponential smoothing model. Using another weighting factor, it smoothes the values of additional trends. It estimates both the level and the trend for future events. (Hanke & Reitsch, 1992; Makridakis & Wheelwright, 1989).

Box-Jenkins ARIMA model involves three basic parameters, p – the amount of autocorrelation, d – the level of systematic change over time (trend) and q – the component for including a moving average of the time based points. These parameters are estimated in an

iterative manner using three stages: (model identification, parameter estimation, and diagnostic checking) of modeling process until the most suitable model is found (Diggle, 2004; Jennings & Young, 1988; Mabert, 1975; Vandaele, 1983).

Where the assumptions of ARIMA are appropriate, the methodology is superior to other statistical techniques because researchers can apply a rational structure approach along with their own experience and judgment to determine a specific model. However, it requires longitudinal data with a minimum of forty-five or sixty data points to achieve highly accurate forecasting. With sufficiently long series, ARIMA methodology usually works well with discernable patterns of trend, seasonal, and cyclical components.

Fuzzy Time Series Analysis

The fuzzy time series model, a sophisticated merger of time series and Fuzzy Set Theory, can be constructed for a nonlinear pattern of enrollment forecasts in which the values of the time series are linguistic terms represented by fuzzy sets (Chen, 2002; Hwang, Chen, & Lee, 1998; Song & Chissom, 1993). These models, which are very mathematical, involve a modeling process based on fuzzification, fuzzy relationship, and defuzzification. Like ARIMA methodology, fuzzy time series initially adopts the method of differencing to remove the linear or curvilinear trend. In addition, it involves two basic parameters, m – the number of grades (fuzzy sets) describing the linguistic variation of enrollment, such as big increase, increase, no change, decrease, and big decrease for m of 5, and w – the window size describing the number of previous time periods to generate the prediction (Chen, 2002; Hwang, Chen, & Lee, 1998; Song & Chissom, 1993). These parameters are selected for the modeling process until the most accurate model is found based on a minimal

of the mean absolute percentage errors. Therefore, this technique is more of a data mining approach that is more frequently used to forecast enrollment rather than offers the explanation of enrollment changes.

Regression Analysis

Regression analysis is useful for predicting enrollment as soon as the key indicators and their lead times are determined. In other words, if enrollment lags purchasing power, how long is the time difference between when purchasing power changes and the enrollment changes. The concept of regression analysis originates from the straight line of least squares that is regressed about the mean of dependent variable. Both linear regression analysis and piecewise regression analysis fall into this category. Linear regression looks for the continuous impact of various factors while piecewise regression can have multiple break points in the relationship between the factor and the resulting enrollment. These techniques can be used to predict enrollment changes based on the change of an indicator and thereby not only model the enrollment but also allow for a discussion about the importance of specific factors in the shifts of enrollment (Lins, 1960; Marsh & Cormier, 2002; Pindyck & Rubinfeld, 1998). However, there are various assumptions that are included with interpreting the results of regression. Furthermore, if there is a limited number of observations, the estimated coefficients can be greatly influenced by one or two outliers.

Modeling Strategies for OSU Enrollment

There were three phases in developing an enrollment forecast model for OSU from Fall 1962 to Fall 2004. In the first phase, the three steps of the ARIMA methodology were iteratively applied: model identification, parameter estimation,

and diagnostic checking. This strategy allowed the researcher to generate the most suitable ARIMA model. In the second phase, a linear regression analysis was computed. Linear regression was used because it allowed the researcher to assess model validity and accuracy by making a head-to-head comparison with ARIMA. Both ARIMA and linear regression models (1) deal with longitudinal aggregated enrollment time series; (2) establish relationships among variables; and (3) perform enrollment forecasts using commonly available software. In the third phase, three model selection criteria were used to make judgments about the most suitable ARIMA and linear regression models: forecasting accuracy, model fitting, and model assumptions. These phases of the research address three research questions for the 1962 – 2004 OSU enrollment: (1) Is the time series of OSU enrollment attributable to the impact of demographics and economic indicators? (2) Does the integrated enrollment forecast model perform OSU enrollment forecasts accurately? and (3) Is the ARIMA forecast model more accurate than the simpler linear regression?

Phase I – ARIMA Methodology

While the ARIMA model has been part of the Time Series option in SPSS at least since version 11.0, The newly developed ARIMA module (SPSS Trends, Version 14.0) was used because it allowed the researcher to build the transfer function model that forecasts student enrollment based on some independent variables and the previous values (lagged) of student enrollment. This ARIMA model is called the transfer function model because it is capable of incorporating independent variables in the model. Note that the traditional ARIMA model involves only lagged enrollment series, lagged errors, or a combination of both as independent variables

(Diggle, 2004; Mabert, 1975; Vandaele, 1983). In this study, two different approaches were utilized to construct ARIMA models: (1) using the structured approach which required the researcher's intervention to identify the model structure; and (2) identifying the best model by the automated process of the Expert Modeler in the SPSS Trends which did not involve the use of the researcher's judgments.

Step 1 – Model Identification:

The first step of the ARIMA modeling process is to identify some candidate ARIMA (p, d, q) models based on the sequence plot of enrollment series and the plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residual series. The parameter values of $p, d,$ and q are denoted by the p th order of autoregressive effect, the d th order of differencing, and the q th order of the moving average.

Initially, the enrollment patterns need to be analyzed. If they are stationary (no linear or curvilinear trend), the assumptions of constant mean and homogeneity of variance are met. However, if the pattern presents a trend, the method of differencing advocated by Box-Jenkins can be used to remove the linear or curvilinear trend (Diggle, 2004; Mabert, 1975; Vandaele, 1983).

The first order of differencing ($d=1$) is designed to remove the linear trend while the second order of differencing ($d=2$) is used to remove the curvilinear trend (Diggle, 2004; Mabert, 1975; Vandaele, 1983). For the first differencing, a new enrollment series is created by subtracting the first-year in enrollment series from the second, the second from the third, etc. The mathematical notion of this first differencing can be illustrated as follows: $D_1 = Y_{t+1} - Y_t$, where $t = 1, 2, \dots, n$. If Y_t is a linear function of time t (i.e., $Y_t = \alpha + \beta t$), then $D_1 = [\alpha +$

$\beta(t+1)] - [\alpha + \beta t] = \beta$, a constant for all t (Diggle, 2004). There are two quick ways of determining the existence of the enrollment trend for the first differencing: (1) examining whether or not the overall enrollment trend consistently increases or decreases across time periods from the sequence plot; and (2) investigating if the autocorrelation function of the original enrollment series reveals positive values that tail off slowly. However, if the first differencing does not achieve stationarity, one of two approaches must be adopted: taking the second order of differencing or taking the natural logarithm of the enrollment series. It should be noted that the second order of differencing is obtained by taking the differences between adjacent first order differences and this removes the curvilinear trends.

After a method of differencing removes the systematic trends in the data, one needs to use the residual plots of ACF and PACF to identify the model structure, either AR(p) model, MA(q) model or mixed model (Armacost & Wilson 2002; Diggle, 2004; Jennings & Young, 1988; Mabert, 1975). The following are some guidelines. There is not a definitive set of rules. The ACF is the function of the serial correlation of error terms and the order of the ACF is the separation of the time periods. The PACF is the also the serial correlation of error terms but with lower order autocorrelations partialled out. After the basic trends in the time series are removed, the ARIMA (p, d, q) model can be generally identified according to two basic rules. First, if the higher order ACF dies down or takes zero value, and the PACF has a spike, it indicates that an autoregressive (AR) model may be appropriate. At this point, the PACF plot is a helpful tool for identifying the order of an AR model. If the enrollment series is actually AR(p), the values of PACF are zeroes after lag p . Therefore, if PACF cuts off after lag p , the AR(p) is the appropriate

model (Armocost & Wilson, 2002; Diggle, 2004; Jennings & Young, 1988; Mabert, 1975; Vandaele, 1983). Secondly, if PACF dies down rapidly without a spike or takes zero value, it indicates that a moving average (MA) model may be appropriate. At this moment, the ACF plot plays an important role in identifying the order of a MA model. If the enrollment series is actually MA(q), the values of ACF will tend to be zero after lag q. Therefore, if ACF cuts off after lag q, the MA(q) is the appropriate model (Armocost & Wilson, 2002; Diggle, 2004; Jennings & Young, 1988; Mabert, 1975; Vandaele, 1983).

Step 2 – Parameter Estimation:

Once the parameter values (p, d, and q) of the ARIMA model are identified, the next step in the ARIMA modeling process is to estimate the regression coefficients based on the least-squares method or the exact maximum likelihood method (i.e., the algorithm of Melard's parameter estimation) (Davis, 1989). Using Melard's parameter estimation for the nonlinear capacity, the regression coefficients are derived in a way that the estimated enrollment series come as close as possible to the actual enrollment series.

Step 3 - Diagnostic Checking:

The final step of the ARIMA modeling process involves examining the ACF of the residual series to be independently and randomly distributed around zero (Diggle, 2004; Mabert, 1975; Vandaele, 1983). To assist in the diagnostic checking process, a chi-square test is used to evaluate if the ACF of the residual series exhibits any systematic pattern (Diggle, 2004; Mabert, 1975; Vandaele, 1983). It is important to examine this step closely because if the model identified is not the suitable one, then it could result in larger forecasting errors.

Phase II – Linear Regression Analysis

The majority of independent variables cover state and national economic indicators, as well as, the lagged values of the same variables were used to explain the enrollment (See Table 1). In this circumstance, collinearity is likely to occur if an individual independent variable highly correlates with the others. The collinearity problem inflates the standard errors of the estimated parameters and leads to inaccurate results of the significance test and the R-squared value (Ott, 2000; SPSS, Inc., 2002). Hence, to restrict the impact of collinearity problems, the "modified" stepwise and backward procedures were implemented, which reduced the likelihood that highly correlated independent measures are included in the final equation and which also allowed the researcher to manually remove individual independent variables with high collinearity (tolerances less than 0.2), and re-fit the regression line to achieve a suitable model. Note that tolerance is defined as $(1-R_i^2)$, where R_i^2 is the *i*th independent variable regressed on the other independent measures. When the value of tolerance is less than 0.2, it means that the R_i^2 is greater than 0.8, indicating a potential problem with high collinearity.

Phase III – Model Comparison

The results of the two models were compared in terms of forecasting accuracy, the fit of the model to the data, and the assumptions involved.

I. Forecasting Accuracy:

The forecasting accuracy signifies the level of agreement between the actual values and the forecast values. The forecast errors (i.e., residuals) are the differences between the actual values and the forecast values. Small forecast errors are an indication of high accuracy

in forecasting. Forecast errors are commonly measured by (1) the mean absolute percentage error (MAPE) or the average of all ratios of the absolute forecast errors to the actual values; (2) the root mean square error (RMSE) or the square root of the mean squared error, where the mean squared error equals the average of the forecast errors; or (3) the mean absolute error (MAE) or the average of the absolute forecast errors (McClave & Benson, 1994; SAS Institute, Inc, 1986). Using a default, conservative, and often recommended choice of forecasting errors, the MAPE allows researchers to compare forecasting accuracy across various techniques (Chen, 1988; Guo, 2000).

II. Model Fit:

A measure of the model fit, R-squared value, is known as the coefficient of determination. This is the proportion of variation in the independent variable that is explained by the model. A higher R-squared value leads to better model fitting (Draper & Smith, 1981; Ott, 2000). Moreover, outliers (the absolute value of standardized residuals > 3) can affect model fitting. It is worth noting that R-squared value is strongly and inversely related to the value of RMSE among various models given equal sample size and an equal number of parameters. The R-squared value can also serve as a judgment for forecasting accuracy.

III. Model Assumptions:

Models are based on assumptions that can frequently be tested. For linear regression analysis, if a randomness pattern exists between the predicted values and residuals, it indicates that the linearity assumption is not violated (Draper & Smith, 1981; Ott, 2000). If the Durbin-Watson statistic is in a range of 1.5 and 2.5, it confirms the independence assumption is valid (McClave & Benson, 1994; Wikipedia

2006). If the scatter diagram of the residual series exhibits a normal distribution, it suggests that the normality assumption is met (Draper & Smith, 1981; Ott, 2000). If chi-square goodness of fit fails to reject the null hypothesis that the distribution is normally distributed with a mean of zero and constant variance, it indicates that the normality assumption is valid (Draper & Smith, 1981; Ott, 2000).

Study Variables

The enrollment series of interest is confined to the 1962-2004 OSU

student enrollments for the Stillwater main campus. The university is part of the land grant system established in 1890 under the Morrill Acts. The Stillwater campus is located in a small city, approximately 70 miles north of Oklahoma City, Oklahoma.

As depicted in Table 1, the 15 independent variables used were: demographics (Oklahoma high school graduates and competitor OU enrollment), state tax fund appropriations for Oklahoma higher education, and economic climate indicators (Oklahoma unemployment rate, Oklahoma per capita income,

the United States GNP, and the United States Consumer Price Index. To make the forecast models more logical and operational, one-year lagged demographics and related economic indicators are treated as independent variables. Lagged variables are not only necessary for forecasting but also supported by the general consistency of the economy. For instance, the national economy rarely changes its patterns drastically within one or two years unless an international crisis or political violence occurs.

Table 1.
Independent Variables for the 1962 – 2004 OSU Enrollment Forecasting Models

Independent Variables ^a	Variable Labels	Sources
OK_Unemp OKUneLG1	Oklahoma Unemployment Rate and One-year Lagged Variable	Business and Economic Research at Oklahoma State University
OkPerCap OKPCaLG1	Oklahoma Per Capita Income and One-year Lagged Variable	U.S. Department of Commerce: Bureau of Economic Analysis
OK_HGGra OKHGLG1	Oklahoma High School Graduates and One-year Lagged Variable	Oklahoma State Regents for Higher Education
US_CPI USCPILG1	U.S. Consumer Price Index and One-year Lagged Variable	National Aeronautics and Space Administration
US_GNP USGNPLG1	U.S. Gross National Product and One-year Lagged Variable	U.S. Department of Commerce: Bureau of Economic Analysis
AP_ST_TX AP_TX_LG1	State Tax Fund Appropriations for Oklahoma Higher Education Operating Expenses and One-year Lagged Variable	Edward R. Hines, Grapevine, March 1962-2004
OU OU_LG1	OU Enrollment (Norman Campus) and One-year Lagged Variable	The University of Oklahoma Website
OSU_LG1	One-year Lagged OSU Enrollment (Stillwater Campus)	Oklahoma State University Website

^a Seven independent variables (OK_Unemp, OkPerCap, OK_HGGra, US_CPI, US_GNP, AP_ST_TX, and OU) were initially used to build ARIMA models. The Expert Modeler in SPSS Trends also used the first and second lag of these seven variables as independent variables. Therefore, it is unnecessary to include the lagged variables in ARIMA models. However, all 15 independent variables in Column 1 were initially considered to build linear regression models.

Trend Analysis for OSU Enrollment Series

The enrollments for OSU and the University of Oklahoma, OSU's major competitor for student enrollment, are shown in Figure 1. As displayed in Figure 1, there was a sizeable increase in OSU students from 1962 to 1982. This record reflects the postwar enrollment expansion known as the "golden age" of higher education between the 1960s and 1970s. However, a decreasing trend is shown for years 1983 to 1994.

OSU student enrollment declined by 21% from 22,366 in 1983 to 17,784 in 1994. This retrenchment coincided with another national trend. The traditional college student cohort of 18 to 21 age groups decreased by 25% between the late 1970s and the early 1990s (WICHE, 1988). A steady increase is displayed in OSU student enrollment for years 1995 to 2004. The OSU enrollment trend parallels a national trend. There was an increase in the number of full-time students by 30% in the United States as well as an 8% increase in part-

time students from 1994 to 2004 (National Center for Educational Statistics, 2006).

ARIMA Modeling Processes and Results

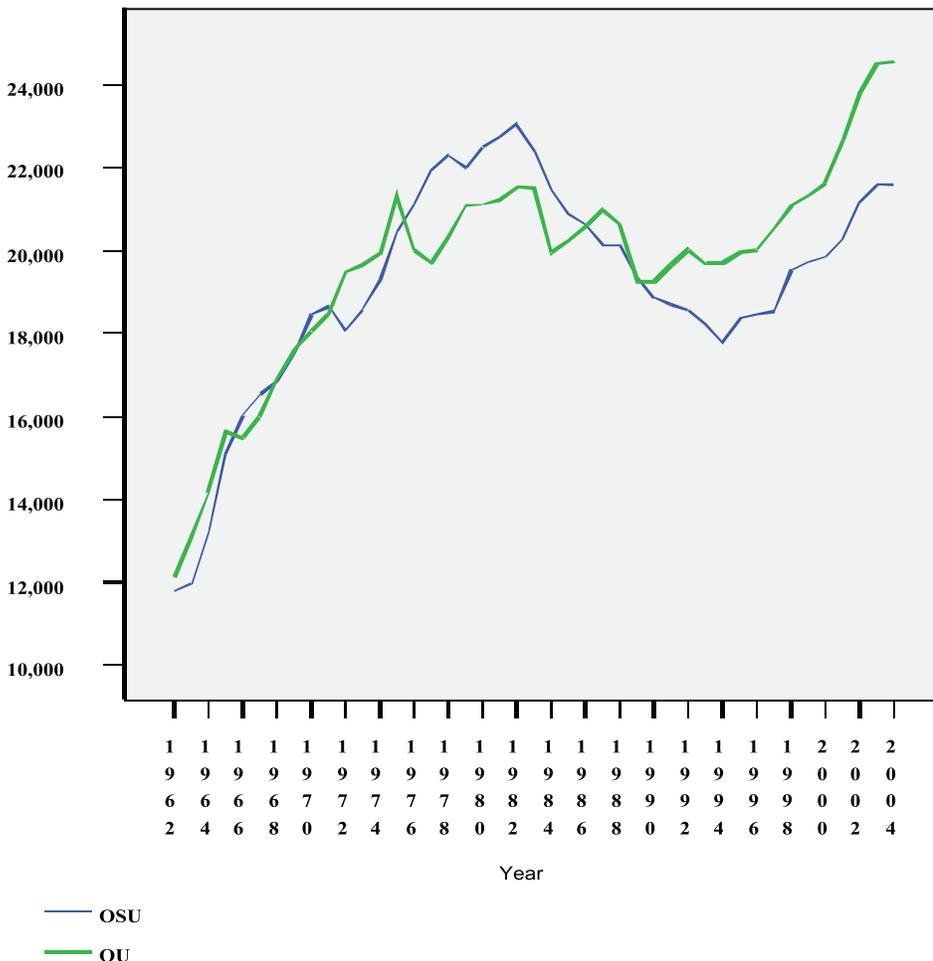
The ARIMA methodology starts with the baseline model ARIMA (0, 0, 0) in which the orders of autoregressive, differencing, and moving average are set to zero. The OSU enrollment series and all seven independent variables (Oklahoma high school graduates, state unemployment rate, state per capita income, state tax fund appropriations, U.S. CPI, U.S. GNP, and the OU competitor college enrollment) enter the model equations simultaneously. In the first trial run, the number of Oklahoma high school graduates was the only significant variable that contributes to OSU enrollment. Consequently, this variable remains active in the modeling process. The other six variables that show no significant association with OSU enrollment were excluded from the modeling process.

Step 1 – Model Identification:

Because the requirement for stationary data (i.e., de-trended) is a basic assumption for developing the ARIMA model, the first step was to investigate whether or not the enrollment series was stationary. From the plot in Figure 1 it appears that the 1962 to 2004 OSU enrollment series displays a non-stationarity pattern, i.e., a chain of rapid growth, steady decline, and uprising trends. Additional confirmation of this non-stationary pattern can be made in Figure 2, when ACF shows non-randomness (i.e., one or more of the autocorrelations are different from zero). Therefore, the first order of differencing, ARIMA (0, 1, 0) model, is implemented to achieve stationarity. In other words, the ARIMA model is transformed from the original

Figure 1.

OSU and OU Enrollment Trends, Fall 1962 – Fall 2004



enrollment series with ARIMA (0, 0, 0) model to the new enrollment series with the first order of differencing, ARIMA (0, 1, 0) model.

As a result of the first differencing, a much higher level of stationarity was accomplished as shown by ACF values of the residual series more frequently approaching zero (quick damping tendency) which can be seen when comparing Figure 3 and Figure 2. In Figure 3, the zero values of ACF are seen between lag 7 and lag 13 for the ARIMA (0, 1, 0) model. Conversely, the zero values of ACF in Figure 2 appear between lag 10 and lag 11 for the baseline model, ARIMA (0, 0, 0). Therefore, it was concluded that for our purposes there is no need to consider a higher order of differencing because the first order of differencing has achieved sufficient stationarity for the enrollment series. Strong evidence of adequate differencing can also be demonstrated based on the randomness of a standardized residual plot and the result of the Expert Modeler to be discussed shortly.

As shown on the residual ACF and PACF plots for ARIMA (0, 1, 0) in Figure 3, the ARIMA (1, 1, 0) model is suggested for the OSU enrollment series, which represents a first order of the autoregressive ($p = 1$) with a first degree of differencing ($d = 1$) and a zero order of moving average ($q = 0$). This is because ACF exhibits a pattern of exponential or sine-wave decay and PACF cuts off sharply at lag 1. This means that higher-order autocorrelations are effectively explained by lag 1. In other words, OSU enrollment is a function of one-year lagged OSU enrollment in addition to the number of Oklahoma high school graduates.

As displayed in Figure 3, the ARIMA (0, 1, 1) model may be considered as a tentative model because ACF of the residual series shows a large spike at lag 1. Note that identification of the MA model can be made directly without checking the PACF plot (Mabert, 1975). The ARIMA (0, 1, 1) model depicts the OSU enrollment

Figure 2.

ACF & PACF Plots for the 1962 – 2004 OSU Original Enrollment, ARIMA (0, 0, 0)

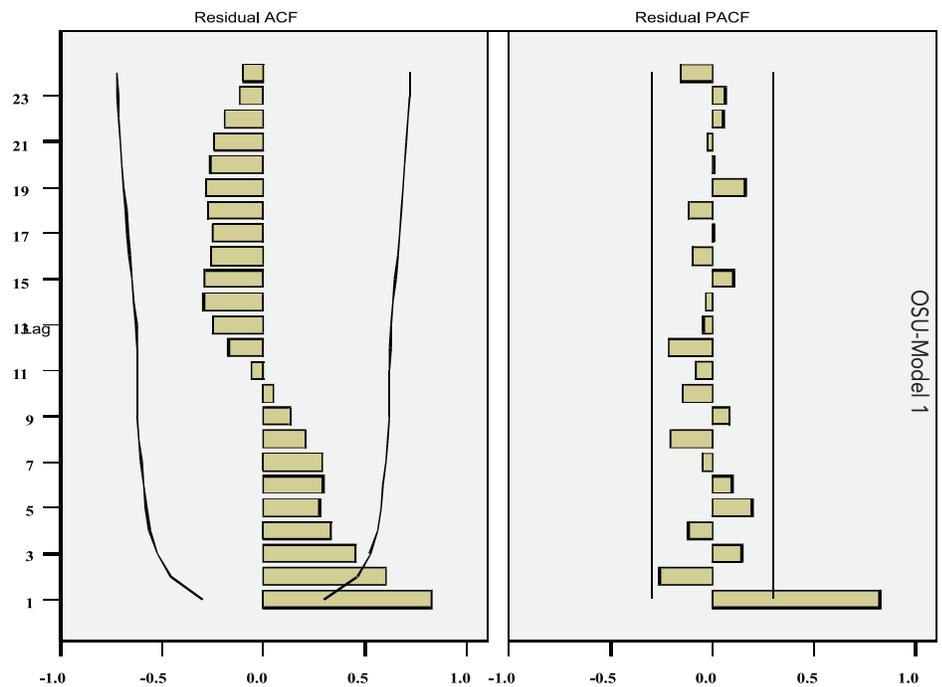
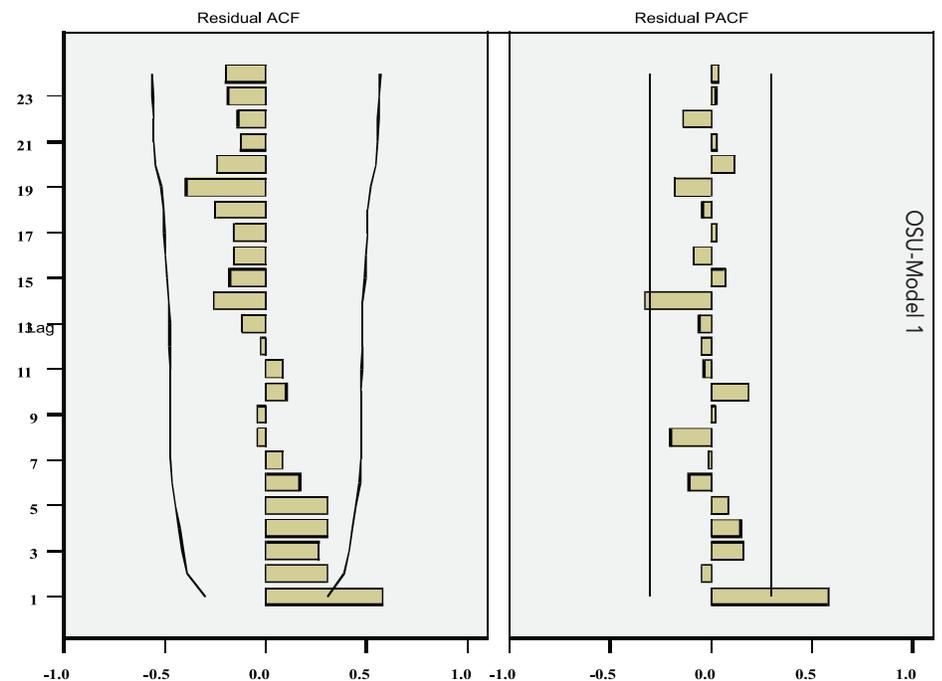


Figure 3.

ACF & PACF Plots for the 1962 – 2004 OSU Model, ARIMA (0, 1, 0)



series with a zero order of autoregressive ($p=0$), a first degree of differencing ($d=1$), and a first order of moving average ($q=1$). This indicates that OSU enrollment is a function of one-year lagged forecast errors and the number of Oklahoma high school graduates. Thus, at this point, two candidate models, ARIMA (1, 1, 0) and ARIMA (0, 1, 1), are identified as viable forecasting models.

Step 2 – Parameter Estimation:

As depicted in Table 2, the resulting model, ARIMA (1, 1, 0), contains significant variables: Oklahoma high school graduates ($\beta_1 = 0.175$ and $p < 0.01$) and AR (1), the one-year lagged OSU enrollment ($\phi_1 = 0.527$ and $p < 0.001$). The ARIMA (0, 1, 1) model covers significant

variables: Oklahoma high school graduates ($\beta_1 = 0.196$ and $p < 0.01$) and MA (1) and one-year lagged moving average (unevenly weighted) of a random shock (error) ($\theta_1 = -0.473$ and $p < 0.01$).

Step 3 – Diagnostic Checking:

In the final step of the ARIMA modeling process, the R-squared values, Ljung-Box Chi-square tests, and standardized residuals are used to assess the model appropriateness. As a result, the ARIMA (1, 1, 0) and ARIMA (0, 1, 1) models have an identical fitting statistic R-squared value of 0.96, which is quite high. In addition, the Ljung-Box Chi-square test indicates that both models are appropriate because the test fails to reject the null hypothesis that the ARIMA model tested is appropriate.

Moreover, no outliers (the absolute value of standardized residuals > 3.0) occur for either ARIMA model.

However, the ARIMA (1, 1, 0) model is determined to be the most promising ARIMA model because it yields higher accuracy in results than the ARIMA (0, 1, 1) model with smaller values of forecast errors: MAPE (2.11% vs. 2.21%), RMSE (440.89 vs. 524.67), and MAE (350.33 vs. 412.19). More importantly, the automated process of the Expert Modeler in the SPSS Trends Version 14.0 confirms that the ARIMA (1, 1, 0) model is the best of all possible ARIMA models.

Finally, the ACF and PACF plots for ARIMA (1, 1, 0) in Figure 4 provide the compelling evidence that the data have been sufficiently detrended because almost all

Table 2.

ARIMA Models for the 1962 – 2004 OSU Enrollment Forecasting

Model Name (Model Type)	Model-A: ARIMA (1, 1, 0) + (First Order of Autoregressive with First Differencing)	Model-B: ARIMA (0, 1, 1) + (First Order of Moving Average with First Differencing)
Independent variables entered the equation simultaneously	All 7 independent variables ^a	All 7 independent variables ^a
Ljung-Box test	$\chi^2 = 15.27$, $df=17$, $p=0.567$	$\chi^2 = 18.11$, $df=17$, $p=0.382$
Significant variables (coefficients) retained	OK_HGGra (0.175***) AR(1) (0.527***)	OK_HGGra (0.196***) MA(1) (-0.473***)
Model fitting R-squared	0.96	0.96
Forecasting accuracy MAPE	2.11%	2.21%
RMSE	440.89	524.67
MAE	350.33	412.19
Remarks	Better ARIMA Model	

⁺ Melard's algorithm for the parameter estimation

^a Seven independent variables (OK_Unemp, OkPerCap, OK_HGGra, US_CPI, US_GNP, AP_ST_TX, and OU) were used to build the ARIMA models. However, all variables except OK_HGGra were removed from the remaining process because they did not significantly contribute to OSU enrollment.

*** $p < 0.001$

autocorrelations approach zero within a 95% confidence interval. Moreover, the standardized residual plot for ARIMA (1, 1, 0) in Figure 5 shows that there is a random pattern, suggesting that the ARIMA (1, 1, 0) model is definitely appropriate. In essence, the ARIMA modeling process is complete and the ARIMA (1, 1, 0), the autoregressive model with the first order of differencing, is the best representation for the 1962 to 2004 OSU enrollment series.

Linear Regression Modeling Processes and Results

Table 3 presents two enrollment forecast models: (1) Model-I, a type of non-autoregressive model that excludes one-year lagged

Figure 4.

ACF & PACF Plots for the 1962 – 2004 OSU Model, ARIMA (1, 1, 0)

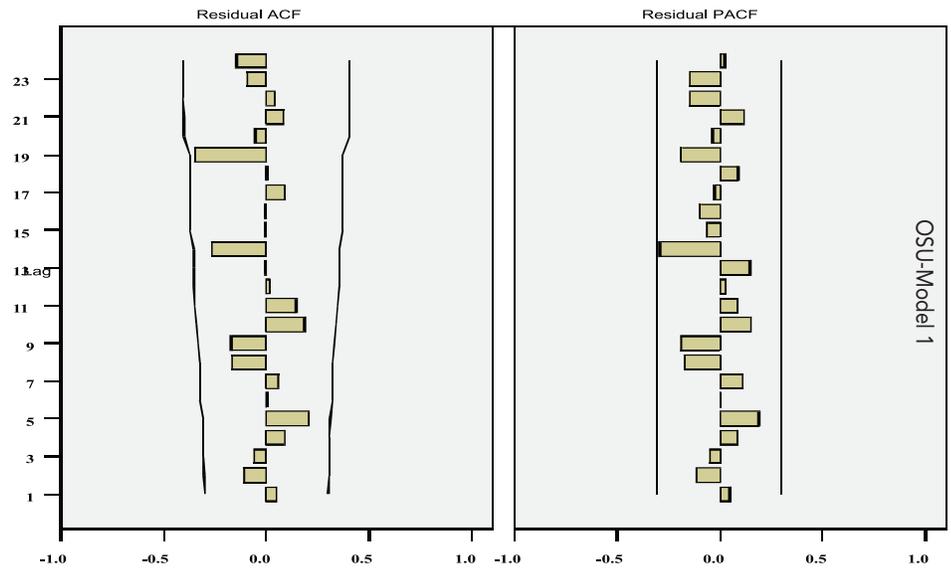


Table 3.

Linear Regression Models for the 1962 – 2004 OSU Enrollment Forecasting

Model Name (Model Type)	Model-I: Linear Regression (Non-Autoregressive -- Exclusive of One-year Lagged OSU Enrollment)	Model-II: Linear Regression (First Order of Autoregressive -- Inclusive of One-year Lagged OSU Enrollment)
Variables selection methods	Stepwise ^a and backward ^b	Stepwise ^c and backward ^d
Independent variables entered the equation simultaneously	All 15 independent variables except OSU_LG1 (See Table 1)	All 15 independent variables (See Table 1)
Significant variables (coefficients) retained	Constant (-7916.597***) OKHGLG1 (0.665***) OK_Unemp (319.802*) OKUneLG1 (281.326*) USGNP (0.380**) < Tolerances are between 0.72 and 0.98 >	Constant (-1519.660) OK_HGGra (0.170***) OSU_LG1 (0.773***) < Both tolerances are 0.53 >
Model fitting R-squared	0.84	0.97
Forecasting accuracy MAPE	3.90%	1.62%
RMSE	1029.52	426.53
MAE	740.00	313.76
Remarks		Best Regression Model

^a Stepwise procedure yields a model with the negative slope for US_GNP, which does not meet the commonsense expectation. Hence, the model is eliminated from the study.

^b Backward procedure requires the researcher to remove nine variables (OKPCaLG1, OKPerCap, AP_ST_TX, APTX_LG1, US_CPI, USCPILG1, OU, OU_LG1, and OK_HGGra) to avoid the collinearity.

^c Stepwise procedure does not require the researcher to remove any variable to avoid the collinearity because the resulting model has large values of tolerance statistics.

^d Backward procedure requires the researcher to remove four variables (US_GNP, OKPerCap, OKPCaLG1, and OU) to avoid the collinearity.

* p < 0.5; ** p < 0.01; *** p < 0.001

OSU enrollment from a pool of independent variables in Table 1; and (2) Model-II, a type of autoregressive model that includes one-year lagged OSU enrollment in the pool of independent variables in Table 1.

The Model-II is considered the best linear regression model with an R-squared value of 0.97 and a MAPE of 1.62, followed by the Model-I with an R-squared value of 0.84 and a MAPE of 3.90%. For the purpose of enrollment forecasting, one may initially draw a conclusion that it is better to adopt Model-II rather than Model-I; however, further observation may reveal that one model possesses useful information that the other does not hold. Model-II comprises only demographics, such as the number of Oklahoma high school graduates and one-year lagged OSU enrollment, instead of the economic indicators. Model-I

contains both demographics and economic indicators that include one-year lagged Oklahoma high school graduates, state unemployment rate, one-year lagged state unemployment rate, and U.S. GNP.

It is worth noting that the three basic assumptions of linearity, independence, and normality are not violated for the best linear regression model, Model-II. A randomness pattern between the predicted values and residuals seems to satisfy the linearity assumption. The Durbin-Watson statistic, 1.2, is almost in the range of 1.5 to 2.5, which may confirm the validity of the independence assumption. Additionally, the normality assumption is not violated because the Chi-square goodness of fit test ($\chi^2 = 1.519$, $df = 1$, and $p > 0.05$) fails to reject the null hypothesis of normality.

Comparison of the Best ARIMA and Linear Regression Models

Table 4 illustrates the comparison of the best ARIMA and linear regression models. The best ARIMA model, ARIMA (1, 1, 0), contains two significant variables: Oklahoma high school graduates ($\beta_1 = 0.175$ and $p < 0.001$) and one-year lagged OSU enrollment ($\phi_1 = 0.527$ and $p < 0.001$). Also, the best linear regression model consists of the same significant variables: Oklahoma high school graduates ($\beta_1 = 0.170$ and $p < 0.001$) and one-year lagged OSU enrollment ($\beta_2 = 0.773$ and $p < 0.001$). The structure of these two models is identical, the first order of the autoregressive model, suggesting that these models have probably demonstrated the model validity for OSU enrollment forecasting. Particularly, the research findings are supported by the literature; the portion of a university's student enrollees (freshmen) depends on the number of high school graduates within the state. Also, students enrolling each year are largely drawn from the same pool of eligible enrolled or returning students (one-year lagged OSU enrollment).

As shown in Table 4, both models fit the data exceptionally well based on their remarkably high R-squared values of 0.96 vs. 0.97. Also, they perform highly accurate enrollment forecasts with very small values of MAPE (2.11% vs. 1.62%), RMSE (440.89 vs. 426.53), and MAE (350.33 vs. 313.76), respectively. Obviously, these two models have demonstrated the forecasting accuracy for the OSU enrollment series. There is no significant difference in the absolute forecasting errors between these two models on the basis of a paired t-test (mean difference of the absolute percent errors = .0049; $t = 1.769$, and $p = 0.084$). From the viewpoint of practical difference, the MAPE difference, 0.49%, (subtracting

Table 4.
The Best ARIMA and Linear Regression Models for the 1962 – 2004 OSU Enrollment Forecasting

Model Name (Model Type)	Model-A: ARIMA (1,1,0)+ (First Order of Autoregressive with First Differencing)	Model-II: Linear Regression (First Order of Autoregressive --Inclusive of One-year Lagged OSU Enrollment)
Estimated parameters Constant	189.42	-1519.66
OK_HGGr	0.175***	0.170***
AR1 (or OSU_LG1)	0.527***	0.773***
Model fitting R-squared	0.96	0.97
Forecasting accuracy MAPE	2.11%	1.62%
RMSE	440.89	426.53
MAE	350.33	313.76
Remarks	Best ARIMA Model	Best Regression Model

+ Melard's algorithm for the parameter estimation
*** $p < 0.001$

1.62% from 2.11%) between these two models is quite small because it accounts for only a 100-student difference given that the average size of student enrollment (approximately 20,000) in recent years is known.

As illustrated in Figure 5, the best linear regression model performs more accurate forecasts (standardized residuals are closer to zero) than the best ARIMA model for two turning points in 1983 and 1995, where standardized residuals for the ARIMA model are close to 2. The best linear regression model is also more favorable to another turning point in 1964, which marks the second data point of the enrollment series. (Note: One-year lagged OSU enrollment in the model leads to the unavailability of the 1962 starting data point) The logical explanation for this outperforming is that the term "regression" literally means "movement backward". All points estimated in dependent variable are regressed about the mean of the dependent variable given that the values of independent variables are known.

Another interesting aspect of Figure 5 is that the best ARIMA model has demonstrated higher accurate forecasts than the best linear regression model in 1972 and 1989, where standardized residuals for linear regression analysis are also close to 2. These years marked the end of the Vietnam War and the closing stage of the postwar baby boom effect, respectively. The rational explanation for this surpassing is that the ARIMA methodology makes more reliable forecasts for time series (dependence and autocorrelation) than linear regression analysis given the finishing points of the Vietnam War (1965 – 1972) and the post-war baby boom (born 1946 - 1964) exhibit some form of successive or autoregressive effect.

As displayed in Figure 5, an outlier with a standardized residual of about less than -3.0 is occurred in 1972 for the linear regression model. It is a forecast error generated by overestimating the OSU enrollment at the end of the Vietnam era in 1972. The temporary removal of this

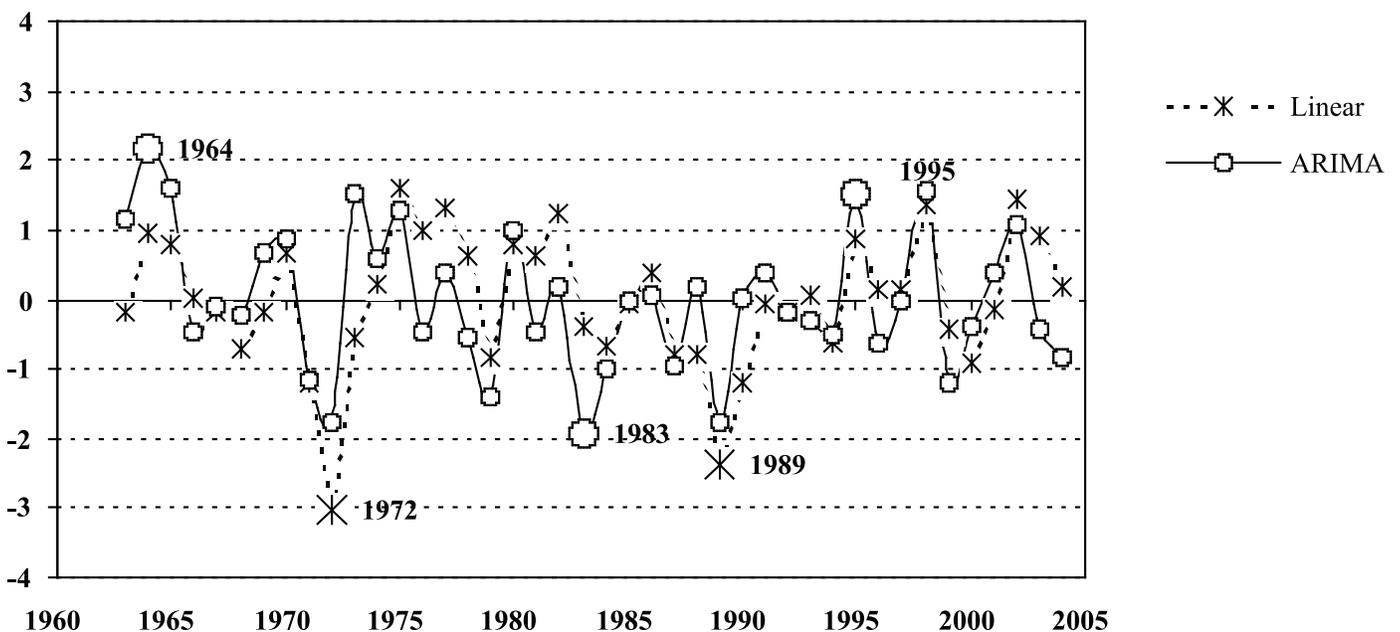
outlier does not change the model structure; instead, it produces a slight decline of the MAPE (0.16% or the difference of 1.62% and 1.46%), which is not significantly different from zero by using a t-test (mean difference = 0.0160, and $p=0.572$). The decision of not removing this outlier can be justified because the war effect on student enrollment may be repeated. For example, the current crisis of the Iraq War or the War on Terror could impact student enrollment in the future if the situation continues and as a result a policy of military drafting is implemented.

Summaries

This paper illustrates the development of the integrated enrollment forecast model for OSU enrollment series from Fall 1962 to Fall 2004. The two best models generated by ARIMA and linear regression methods fit the data exceptionally well with high R-squared values of 0.96 and 0.97, respectively. Both models also forecast highly accurate OSU

Figure 5.

Standardized Residuals Plot for the Best ARIMA and Linear Regression Models



enrollment with MAPE values of 2.11% and 1.62%, respectively.

The best linear regression model outperforms the best ARIMA model, ARIMA (1, 1, 0), for the turning points in 1983 and 1995. On the other hand, the best ARIMA model demonstrates more accurate forecasts than the best linear regression model in years 1972 and 1989, which mark the end of the Vietnam War and the closing stage of the post-war baby boom effect respectively. However, there is no significant or practical mean difference in the absolute percentage errors between the two models.

The resulting models generated by both ARIMA and linear regression methods indicate that OSU enrollment is primarily a function of two demographics: Oklahoma high school graduates and one-year lagged OSU enrollment. As a result, the first-order of autoregressive model in conjunction with High School graduations represent the 43-year OSU enrollment series adequately. Moreover, the structural approach has proven to be quite valuable in narrowing down the two significant variables, which accomplishes the principle of parsimony for enrollment forecasting.

It appears that the first order of autoregressive model becomes the best choice for OSU enrollment series within ARIMA methodology and linear regression analysis, respectively. This is because the current value of the enrollment series is expressed as a function of the previous value of the enrollment series. The integrated enrollment forecast model has demonstrated its model validity and forecasting accuracy. Hence, it can be replicated and may well be useful for estimating aggregated student enrollment in other similar institutions of higher education.

Strengths

There are major strengths of the integrated enrollment forecast

model. First, the structural approach is implemented to construct candidate models, eliminate inappropriate ones, and ultimately retain the most suitable model. This approach avoids the arbitrary choice of a specific model that may not be the best fit. In addition, a head-to-head comparison is made between both ARIMA and linear regression methods to demonstrate the model validity and forecasting accuracy.

Second, the modeling process enables the researcher to identify a subset of key independent variables from a pool of 15 variables. The best ARIMA and linear regression models yield only two identical demographics, one-year lagged college enrollment and high school graduates. These variables are significantly and positively associated with student enrollment, which are consistent with the research findings from the literature. The overall student enrollment closely tracks the variables of enrollment in the previous year and the number of high school graduates in the current year.

Third, linear regression analysis has the potential of being used to reinforce the ARIMA methodology. Collinearity problems can be managed by the examination of tolerance statistics and removing the more serious sources of Collinearity while preserving significant variables. In turn, the ARIMA methodology along with some help from the SPSS expert system validates the first order of the autoregressive model generated from linear regression analysis.

Limitations

One should be aware of the following limitations of both ARIMA and linear regression methods: (1) select independent variables for the historical data and future values that are available, (2) reconstruct and reevaluate the models as years progress because of the uncertainty that surrounds the enrollment forecasting, and (3)

forecast short-term enrollment under the assumption that independent variables remain fairly constant over time since changes in variables could be sensitive to student enrollment. This includes using lagged variables to help look for turning points and always involving expert judgment to understand major events. As stated by the OSU Office of Institutional Research in 2005, the following variables should remain constant when dealing with enrollment forecasting: national and state economy, federal and state financial aid programs, state funding for higher education, admission standards, graduation rates, and so forth.

Major Alternatives

There are four areas that are major alternatives. First, increasing female enrollment should be considered for future model construction. According to Postsecondary Education Opportunity (Number 109, July 2001), the proportion of women ages 18 – 24 enrolled in colleges doubled between 1967 and 2000, from 19.2 to 38.4 percent. Second, because Oklahoma's economic climate is frequently driven by oil production, oil drilling rig activity is another important state economic indicator that should be considered as a influential factor of student enrollment. Third, enrollment forecasting should be linked to student financial aid and tuition. Financial aid and tuition amounts may be used in the simulation process to maximize student enrollment and related revenues, given that other selected variables remain constant. Finally, the ARIMA methodology requires a longitudinal time series, at least 45 or 60 data points, to yield highly accurate forecasting as cited in the literature. This study needs to be extended in the future as more data becomes available.

Editor's Notes

Enrollment projections have long been one of the most likely and important activities of an IR function. These projections support financial estimates of both revenue and expenditure. These projections can support admissions decisions. These projections can support requirements for staffing and for facilities. Often, we have a methodology that is traditional for our specific institution, however, we neither publish that methodology nor do we seriously consider alternatives.

That is why it is delightful to share this IR Applications on the topic of enrollment projections by Chau-Kuang Chen. Several aspects of this article make it a very valuable and informative contribution to our discussions. First, there is a review of some of the major previous work in the area of projections. The article considers many of the demographic, political, and economic factors that have been associated with enrollment trends. These factors could be extended by using moderators such as population density.

The second, and perhaps one of the most valuable aspects of

this article, is the discussion of some of the major techniques used in projections. There are numerous alternatives ranging from Subjective Judgment to the very analytical Neural Network and Simulation Methods. Each of the various methodologies has its data requirements and its unique strengths and weaknesses. Unfortunately, space does not permit the full discussion of the strengths, weaknesses, and variations of these alternatives. The article does prompt the question: Which of these is most appropriate for the needs and characteristics of your institution?

Thirdly, this article gives a very practical step-by-step discussion of one of the more sophisticated techniques from econometrics, Time Series Analysis that considers autocorrelation and also moving averages. The technique, Autoregressive Integrated Moving Average (ARIMA) seems to be best suited for institutional-level projections and requires a large number of data points. It does a good job making the overall projections using reasonable variables of lagged enrollment and high school graduates. In addition to

the traditional measures of statistical quality, we are introduced to some of the standards from econometrics such as the Mean Absolute Percentage Error (MAPE). The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are used to determine the parameters of the ARIMA model. While there is a thorough discussion of this methodology, it would seem prudent to enlist the help of an econometrician to actually establish the parameters.

As a final contribution, Chen demonstrates how to compare the ARIMA methodology to the more traditional multiple regression. The use of the graph shows that both have points at which they track closely with the actual and points at which they track less well. This raises the question of how your assumptions of the future would affect the methodology that is selected.

In the context of IR, this article challenges us to look at our data and requirements for projections. It gives us a range of alternatives. What alternatives seem feasible for your institution? Are there several that seem to be useful? This issue broadens your options.

Bibliography

- Anderson, David R.; Sweeney, Dennis J.; and Williams, Thomas A. (2000). *An Introduction to Management Sciences: Quantitative Approaches to Decision Making*. 9th ed., Cincinnati South-Western College Publishing.
- Barbett, Samuel (1996). Residence and Migration of First-Time Freshmen Enrolled in Degree-Granting Institutions, ERIC ED 418651.
- Breneman, D. W. (1981). "Higher Education and the Economy," *Educational Record*, 62, 2, p. 18-21).
- Brinkman, Paul T. and McIntyre, Chuck (1997). Methods and Techniques of Enrollment Forecasting. Chapter 5 in D. T. Layzell (Ed.), *Forecasting and Managing Enrollment and Revenue. New Directions for Institutional Research*. No 93, Jossey-Bass Publishers. Pp. 67-80.
- Brown, D. J. (1978). "Forecasting University Enrollments by Ratio Smoothing," *Higher Education*, 7, p. 417-429.
- Campbell, R. and B. Siegel. (1967). "The Demand for Higher Education in the United States, 1919-1964," *American Economic Review*, p. 482-494.
- Carnegie Foundation for the Advancement of Learning. (1975) *More Than Survival: Prospects for Higher Education in a Period of Uncertainty*. San Francisco: Jossey-Bass.
- Chen, Chau-Kuang (1988). Enrollment Forecasts for Oklahoma State University, <http://e-archive.library.okstate.edu/dissertations/AAI18914991/>
- Chen, S. M. (1996). Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets and Systems*, 81: 311-319.
- Chen, S. M. (2002). Forecasting Enrollments Based on High-Order Fuzzy Time Series. *Cybernetics and Systems: An International Journal*, 33: 1-16.
- Clagett, Craig A. (1989). Credit Headcount Forecast for Fall 1989-90: Component Yield Method Projections. ED325162
- Crossland, F. E. (1980). Learning to Cope with a Downward Slope. *Change*, 12, 18, p. 20-25.
- Davis, William W. (1989). SPSS/PC+Trends. *The American Statistician*. Vol. 43, No. 1, pp 57-59
- Day, J.H (1997). Forecasting and Managing Enrollment and Revenue: An Overview of Current Trends, Issues, and Methods. *New Directions for Institutional Research*, No. 93, 51-65, San Francisco: Jossey-Bass
- Dckel, G.P. (1994). Projecting School Enrollment: A Solution to a Methodological Problem. *School Business Affairs*. 60(4), 32-34
- Diggle, Peter J. (2004). *Time Series: A Biostatistical Introduction*. Oxford Science Publications.
- Donhart, G.L. (1995). Tracking Student Enrollment Using the Markov Chain. *Journal of College Student Development* 36:5 p457-462
- Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*, New York: John Wiley & Sons, Inc.
- Fahert, V.E. (1997). Using Forecasting Models to Plan for Social Work Education in the Next Century. *Journal of Social Work Education*, 33(2), 403-411.
- Folger, J. K. (1974). "On Enrollment Projections." *Journal of Higher Education*, 45, p. 405-414.
- Frankel, M. M. and W. H. Forrest. (1977). *Projections of Educational Statistics to 1985-1986*. Ishington, D.C.: U.S. Government Printing Office.
- Freeman, J.A. and Skapura, D.M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley Publishing Company, Inc.
- Gardner, D. E. (1981). "Weight Factor Selection in Double Exponential Smoothing Enrollment Forecasts." *Research in Higher Education*, 14, 1, p. 49-56.
- Gerald, Debra E. and Hussar, William J. (2007). Projections of Education Statistics to 2009. Vol 1, Issue 4. *Education Statistics Quarterly – Crosscutting Statistics*.
- Glass, T.E. and Fulmer, C.L. (1991). Enrollment Projections – Factors and Methods. *CEFPI's Educational Facility Planner*. 29(6). 25-30
- Greiner, Keith and Girardi, Tony (2006). IOWA as a Destination for Out-of-State Students: State-to-State Migration of Students to IOWA, ERIC ED495022.
- Guo, Shuqin (2002). Three Enrollment Forecasting Models: Issues in Enrollment Projection for Community Colleges. Paper presented at the 40th RP Conference. Asilomar Conference Grounds, Pacific Grove, California. http://www.ocair.org/files/presentations/paper2002_03/rp_paper2002final.pdf
- Hanke, J.E. and Reitsch, A. G. (1992). *Business Forecasting* (4th Edition). Needham Heights, Mass Allyn & Bacon.
- Heller, D. E. (1999). The Effect of Tuition and State Financial Aid on Public College Enrollment. *The Review of Higher Education*, 23 (1), 65-89.
- Hoenack, S. A. and W. C. Weiler. (1979). "The Demand for Higher Education and Institutional Enrollment Forecasting." *Economic Inquiry*, XVII, p. 89-113.
- Huang, Kunhuang and Yu, Tiffany Hui-Kuang (2005). The Application of Neural Networks to Forecast Fuzzy Time Series. *ScienceDirect – Physica A: Statistical Mechanics and Its Applications*. Pages 1 – 14.
- Hwang, J. R., Chen, S. M., and Lee, C. H. (1998). Handling Forecasting Problems Using Fuzzy Time Series. *Fuzzy Sets and Systems* 100, pp 217-228.
- Jackson, G. A. and G. B. Weathersby. (1975.) "Individual Demand for Higher Education." *Journal of Higher Education*, 46, 6, p. 623-652.

- Jenning, Linda and Young, Dean M. (1988). *New Directions for Institutional Research* No. 58 v15 n2 p77-96
- Lins, L. J. (1960). *Methodology of Enrollment, Projections for Colleges and Universities*. Washington, D.C.: U.S. Department of Education.
- Liu, Richard (1998). "Short-term Enrollment Projection: An Example at State Level", *Higher Education*, Summer 1998
- Lyell, E. H. and P. Toole. (1974). "Student Flow Modeling and Enrollment Forecasting." *Planning for Higher Education*, 3, 6, p. 2-6.
- Mabert, V. A. (1975). *An Introduction to Short Term Forecasting Using the Box-Jenkins Methodology*. Norcross, Ga.: American Institute of Industrial Engineers, Inc.
- Maganell, J. (1980). "Despite Drop in Number of 18 Year-olds, College Rolls Could Rise During the 1980's." *The Chronicle of Higher Education*, 20, 1, 11.
- Makridakis, S. and S. C. Wheelwright. (1989). *Forecasting: Methods and Applications*. New York: John Wiley and Sons.
- Makridis, S. et. al. (1992). "The Accuracy of Extrapolation (Time-Series) Methods: Results from a Forecasting Competition", *Journal of Forecasting* (no. 1, pp 111-53)
- Mangelson, W. L., et. al. (1973). *Projecting College and University Enrollments: Analyzing the Past and Forecasting the Future*. Ann Arbor, Mich.: Center for the Study of Higher Education.
- Marsh, Lawrence C. and Cormier, David R. (2002). *Spline Regression Models*, Sage Publications
- McClave, James T. and Benson, P. George (1994). *Statistics for Business and Economics* (6th Edition). Prentice-Hall, Inc.
- McPherson, Michael S. and Schapiro, Morton (1991). Does Student Aid Affect College Enrollment? New Evidence on a Persistent Controversy. *The American Economic Review*, 81, 1.
- National Center for Educational Statistics (1978). *Projections of Educational Statistics, 1986-87*. Washington, D.C.: U.S. Government Printing Office.
- National Center for Educational Statistics (1994). *Projections of Education Statistics to 2004. Digest of Education Statistics 1994*.
- National Center for Educational Statistics (2006). *Digest of Education Statistics 2006*, <http://nces.ed.gov/programs/digest/d05/>; Pages 1 – 5
- OSU Office of Business and Economic Research (1987). *Economic Outlook*, Oklahoma State University document.
- Ott, Lyman (2000). *An Introduction to Statistical Methods and Data Analysis*, 4th ed. PWS-KENT Publishing Company
- Pindyck, R. S., and Rubinfeld, D. L. (1998). *Econometric Models and Economic Forecasts*, (4th ed.) New York: Irwin/McGraw-Hill
- Render, Barry and Stair, Jr., Ralph M. (2000). *Quantitative Analysis for Management*, 7th ed., Upper Saddle River: Prentice Hall
- SAS Institute, Inc. (1986). *SAS System for Forecasting Time Series*, Gary, N.C.
- Song, Q. and Chissom, B.S. (1993). Forecasting Enrollments with Fuzzy Time Series: Part I. *Fuzzy Set and Systems* 54: 1-9.
- Song, Q. and Chissom, B.S. (1994). Forecasting Enrollments with Fuzzy Time Series: Part II. *Fuzzy Set and Systems* 62: 1-8.
- SPSS, Inc (2002). *SPSS Advanced Model 10.0*, Chicago, IL.
- Stapleton, David C. and Young, Douglas J. (1988). Educational Attainment and Cohort Size. *Journal of Labor Economics*. Vol 6., No. 3, pp 330-61
- Stewart, C. T. and A. Kate (1978). "College Enrollment in Response to Fluctuation in Unemployment and Decline." *College and University*, p. 301-313.
- Shaw, Robert C. (1984). Enrollment Forecasting: What Methods Work Best? *NASSP Bulletin* v68 n468 p52-58.
- Skapura, David M. (1995). *Building Neural Networks*, Addison-Wesley Professional, New York, N.Y
- Texas State Higher Education Coordinating Board (2001). *Student Migration Report, 2000-2001*. Austin Division of Community and Technical Colleges. ERIC ED481778
- Tukey, D. (1991). Models for Student Retention and Migration. *Journal of the Freshman Year Experience*. P61-74
- Vandaele. Walter (1983). *Applied Time Series and Box-Jenkins Models*, Academic Press, Inc. San Diego, California.
- Wachter, Michael L. and Wascher, William L. (1984). Leveling the Peaks and Thoughts in the Demographic Cycle: An Application to School Enrollment Rates. *Review of Economics and Statistics*. Vol 66, No. 2, pp. 208-15.
- Wagschall, P. H. (1983). "Judgment Forecasting Techniques and Institutional Planning: An Example." *New Directions for Institutional Research* (No. 39). San Francisco: Jossey-Bass.
- Western Interstate Commission for Higher Education, et. al. (1988). *High School Graduates: Projections by State, 1986 to 2004*. Boulder, CO: Western Interstate Commission for Higher Education
- Wheelwright, S. C. and S. G. Makridakis (1977). *Forecasting Methods for Management* (2nd ed). New York: John Wiley and Sons.
- Wing, P. (1974). *Higher Education Enrollment Forecasting: A Manual for State Level Agencies*. Boulder, Colo.: National Center for Higher Education Management Systems.
- Wikipedia, the Free Encyclopedia (2006). http://en.wikipedia.org/wiki/Durbin-Watson_Statistics
- Witkowski, E. H. (1974). "The Economy and the University: Economic Aspects of Declining Enrollment." *Journal of Higher Education*, 45, p. 48-60.



IR Applications is an AIR refereed publication that publishes articles focused on the application of advanced and specialized methodologies. The articles address applying qualitative and quantitative techniques to the processes used to support higher education management.

EDITOR:

Dr. Gerald W. McLaughlin

Director of Planning and
Institutional Research
DePaul University
1 East Jackson, Suite 1501
Chicago, IL 60604-2216
Phone: 312-362-8403
Fax: 312-362-5918
gmclaugh@depaul.edu

ASSOCIATE EDITOR:

Ms. Deborah B. Dailey

Assistant Provost for Institutional
Effectiveness
Washington and Lee University
204 Early Fielding
Lexington, VA 24450-2116
Phone: 540-458-8316
Fax: 540-458-8397
ddailey@wlu.edu

MANAGING EDITOR:

Dr. Randy L. Swing

Executive Director
Association for Institutional Research
1435 E. Piedmont Drive
Suite 211
Tallahassee, FL 32308
Phone: 850-385-4155
Fax: 850-385-5180
air@airweb2.org

AIR IR APPLICATIONS EDITORIAL BOARD

Dr. Trudy H. Bers

Senior Director of
Research, Curriculum
and Planning
Oakton Community College
Des Plaines, IL

Dr. Philip Garcia

Director of
Analytical Studies
California State University-Long Beach
Long Beach, CA

Ms. Rebecca H. Brodigan

Director of
Institutional Research and Analysis
Middlebury College
Middlebury, VT

Dr. Glenn W. James

Director of
Institutional Research
Tennessee Technological University
Cookeville, TN

Dr. Harriott D. Calhoun

Director of
Institutional Research
Jefferson State Community College
Birmingham, AL

Dr. David Jamieson-Drake

Director of
Institutional Research
Duke University
Durham, NC

Dr. Stephen L. Chambers

Director of Institutional Research
and Assessment
Coconino Community College
Flagstaff, AZ

Dr. Anne Machung

Principal Policy Analyst
University of California
Oakland, CA

Dr. Anne Marie Delaney

Director of
Institutional Research
Babson College
Babson Park, MA

Dr. Jeffrey A. Seybert

Director of
Institutional Research
Johnson County Community College
Overland Park, KS

Dr. Paul B. DUBY

Associate Vice President of
Institutional Research
Northern Michigan University
Marquette, MI

Dr. Bruce Szelest

Associate Director of
Institutional Research
SUNY-Albany
Albany, NY

Authors can submit contributions from various sources such as a Forum presentation or an individual article. The articles should be 10-15 double-spaced pages, and include an abstract and references. Reviewers will rate the quality of an article as well as indicate the appropriateness for the alternatives. For articles accepted for *IR Applications*, the author and reviewers may be asked for comments and considerations on the application of the methodologies the articles discuss.

Articles accepted for *IR Applications* will be published on the AIR Web site and will be available for download by AIR members as a PDF document. Because of the characteristics of Web-publishing, articles will be published upon availability providing members timely access to the material.

Please send manuscripts and/or inquiries regarding *IR Applications* to Dr. Gerald McLaughlin.