

## Forecasting Enrollment in Higher Education using SAS® Forecast Studio®

Paper Author 1 Mr. Erik Bowe, Kennesaw State University; Paper Author 2 Mr. Steven Merritt, Kennesaw State University

### ABSTRACT

Forecasting enrollment has become a critical part of institutional planning processes in higher education. In order to plan effectively at a comprehensive university in Georgia with approximately 25,000 students, Kennesaw State University developed two forecast models in the SAS® BI Platform. The forecast model is a ratio-based short-term, semester-by-semester model constructed in SAS® Enterprise Guide®, and the other a time series-based long-range, five-year model built in SAS® Forecast Studio®.

### INTRODUCTION

Kennesaw State University (KSU) is the third largest institution of the thirty-one member University System of Georgia, and like other institutions of higher education that are funded by state appropriations, forecasting enrollment has become a critical part of the institutional planning process. As a comprehensive university with approximately 25,000 students, KSU has constructed both ratio-based and time series models in SAS® Enterprise Guide® and SAS® Forecast Studio® respectively in order to predict future enrollment with a high degree of accuracy. The ratio-based models are used for short-term, semester-based forecasts, while the time series-based models are utilized for long-range (i.e., five or more years), annual forecasts.

### RATIO-BASED ENROLLMENT FORECAST MODEL

As part of the institutional planning process at KSU, like most institutions of higher education, forecasting of enrollment needs to occur on a moving semester-by-semester basis utilizing past enrollment trends. If business practices and other key conditions remain relatively stable from year-to-year, a short-term, highly reliable ratio-based forecast is possible. Since KSU is able to benefit from stable business practices, the institution is able to utilize this approach to forecasting its enrollment each semester.

For KSU, official census-based enrollments for a given semester (after all fee payments are reconciled) occur five to six weeks from the beginning of the semester. Typically, KSU drops several hundred registered students from the earlier totals of registrations because of failure to pay their fees. Consequently, the daily rolling counts of registrations captured in KSU's data warehouse are somewhat inflated with students who fail to complete the fee payment process. As a result, the official census enrollment for the semester will be lower despite the fact that some of those students will complete the fee payment process and have their courses reinstated. Nevertheless, utilization of registration-based totals can occur to make reasonably accurate ratio-based forecasts of the semester's census enrollment six to seven weeks in advance of the actual determination of the official census enrollment.

As indicated in Table 1 for KSU, the ratio of official census re-enrollments to the inflated total number of registrations at the end of late registration (drop/add) for Fall Semester has been very stable over the past eight years and can be used to forecast the final census total at a high level of accuracy.

Fall Semester	Fall Census	Total Registrations	Census/Registrations Ratio
2003	17,485	17,906	.976
2004	17,961	18,373	.977
2005	18,556	18,992	.977
2006	19,854	20,342	.976
2007	20,603	21,022	.980
2008	21,449	21,966	.976
2009	22,389	22,841	.980
2010	23,452	23,876	.982

**Table 1. Historical Ratios of Fall Census Enrollment/Fall Registrations at KSU**

Calculation of the projected census enrollment is the result of taking the product of the total number of registrations at the end of late registration multiplied by the average ratio of official census totals to total registrations over the past six years. As mentioned earlier, this projection method is highly reliable so long as business practices, other key conditions, and the annual ratios remain relatively stable from year-to-year. Although only Fall Semester is shown in Table 1, this ratio-based calculation can be applied to Spring and Summer Semesters as well, with the latter showing a bit more volatility in the average of the ratios.

Projecting percentage enrollment changes based on year-to-date comparisons of the daily rolling counts of registrations from KSU's data warehouse is also reasonably accurate by the end of the late registration period. However, caution in using such a ratio is advised when making earlier year-to-date comparisons during the registration periods. Such forecasts can be misleading, especially if registration schedules, business practices, economic conditions, or other key factors are not comparable across the two or more years. For example, the economic downturn in 2008 may have caused fewer KSU students to register early for the Fall during the previous Spring than in 2007, explaining the slight increase in registration totals from 2007 to 2008. In contrast, 6% more students had registered in 2009 by the end of pre-registration in July than was the case in 2008, however, that may have been due largely to early registration being expanded to two months in April 2009. Having fewer days of registration during the final and late registration periods in August 2009 contributed to the shrinkage of the daily year-to-year percentage change to 4.5% by the end of the drop/add period as compared to August 2008. Clearly, several key differences in the conditions affecting Fall registration totals can play havoc with comparative calculations of year-to-date changes early in the registration process.

In order to turn the ratio-based forecast into actionable intelligence, KSU surfaces the model explained above in SAS® Enterprise Guide®. Using output from ODS (Output Delivery System), a Microsoft Excel file is produced that is converted to Adobe® PDF and distributed via e-mail. Figure 1 shows the output from the Enterprise Guide® Project. The graph produced is a product of Microsoft Excel and not SAS® Graph, although KSU is considering a modification to the Project to incorporate SAS graphing capabilities. Implementation of the ratio-based forecast model itself occurred as a result of using PROC SQL to create a view directly against the KSU Data Warehouse, which contains the capture of daily registration counts as well as the official census enrollment for past semesters. PROC Report is utilized to produce the data summary table and notes shown in Figure 1.

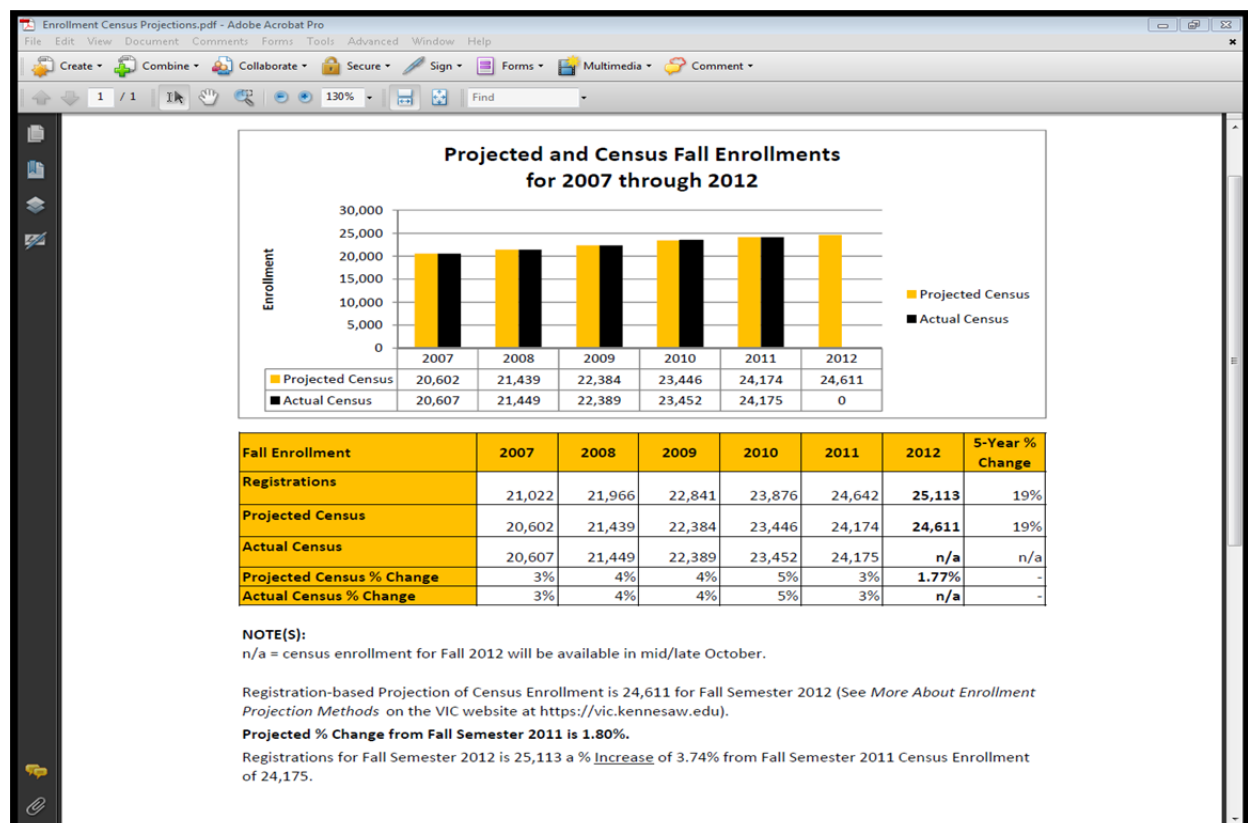


Figure 1. Sample Output of Ratio-Based Enrollment Forecast

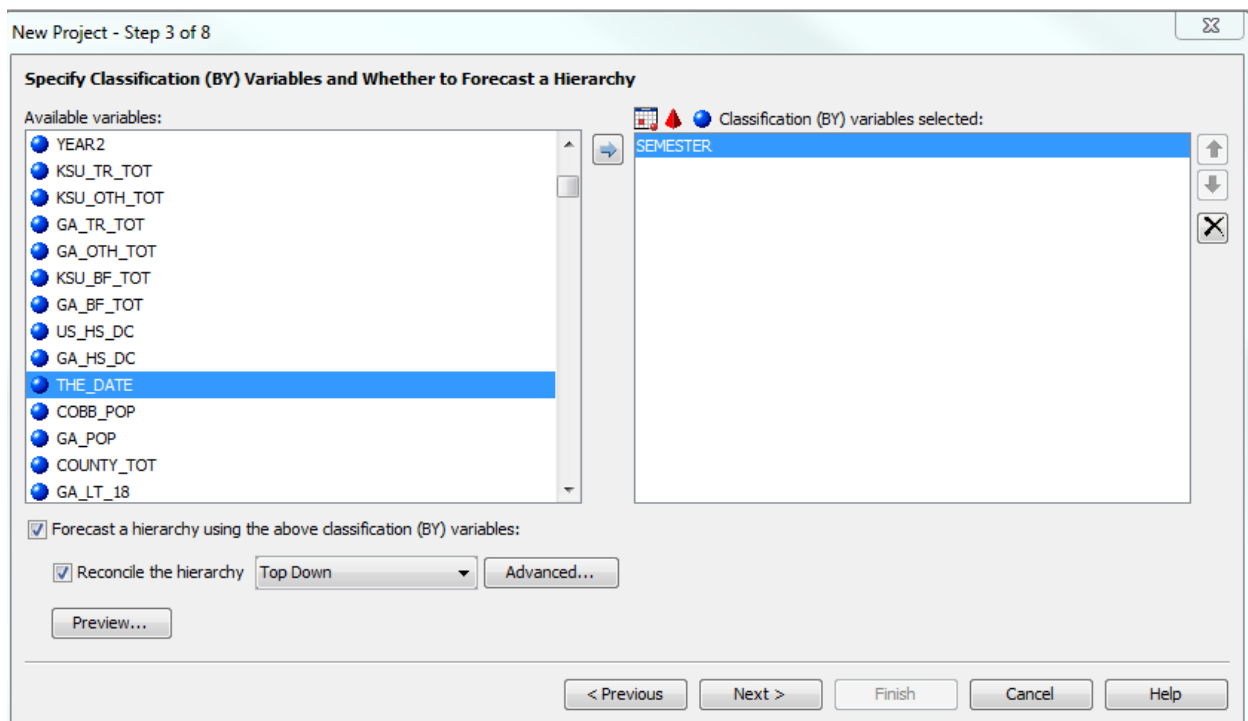
## TIME SERIES-BASED ENROLLMENT FORECAST MODEL

In order to construct an accurate long-range (i.e., five or more years) enrollment forecast, KSU examined several models and time series-based statistical methods. Given the limited number of predictor variables utilized, KSU selected a mixed ARIMA (autoregressive integrated moving average) model. This mixed form of the ARIMA model is a more general form of several different types of predictive models. Mixed ARIMA models can have autoregressive terms, non-seasonal differences and lagged forecast errors.

ARIMA models are a type of mathematical model that is used to make predictions based on historical data on what you are trying to predict as well as things that may influence what you are trying to predict. For instance, if you wanted to attempt to forecast future enrollment at a university, you would produce an ARIMA model based on past enrollment and influencers of that enrollment such as local population and economic conditions. To set up an ARIMA model in SAS® Forecast Studio®, you must collect the needed data including predictor variables and a date variable into a database/dataset that you can access within Forecast Studio®.

Once inside of Forecast Studio®, you create a new forecast by clicking on File->New Project. This opens up the new project wizard, which guides you through the process of creating a new project. First, enter the name of your project and a short description. Next, select the data set you will be working with to forecast.

In Step 3, you will have the option to include a “by” variable, as shown in Display 1. This is useful if you wish to create a hierarchy. For instance, if you wanted to predict future enrollment, you may decide to split out each year into semesters. You would then have a projection for each semester and another projection for the overall year. Naturally, when you have a projection for each semester and another projection for the overall year, the overall year projection is the same thing as the underlying semesters added together. This is where reconciliation comes in. SAS® Forecast Studio® can either change the semester projections to fit in line with the overall projection or change the overall year projection to fit the semester projections. Since the yearly projection is at the top of the hierarchy, a top down model will modify the semester projections to fit the yearly projection, whereas a bottom up projection would modify the yearly projection to fit the semester projections.



**Display 1. Step 3 of the SAS® Forecast Studio® Project Wizard**

After selecting your time variable in Step 4, you are taken to the variable selection screen. The variable selection screen in Step 5 is shown in Display 2. Here you will select the role of variables. Variables can take on the role of None, Dependent, Independent, or Report Only. Variables with a role of None or Report Only will not influence your

projections. Your dependent variable is what you are trying to forecast. If you have chosen to create a hierarchy, you can only have one dependent variable. Your independent variable(s) are the variables which you are using to predict future behavior for your dependent variable. Including a dependent variable here does not mean it will be used, it only means that it may be used.

**New Project - Step 5 of 8**

**Assign Roles to Variables in Your Data**

**You must specify a dependent variable.**

Variable	Role	Aggregation	Accumulation	Usage in System-Generated ...
TR_FR_TOT	None			
TR_SO_TOT	None			
TR_JR_TOT	None			
TR_SR_TOT	None			
TR_FR_SO_TOT	Dependent	Sum of values	Sum of values	
TR_JR_SR_TOT	Independent	Average of values	Average of values	Try to use
TR_SO_JR_TOT	None			
TOT_BIG3	Dependent			
TOT_CHATTAHOOCHEE_TECH	Independent			
TOT_GEORGIA_PERIMETER	Report Only			
TOT_GEORGIA_HIGHLANDS	None			
CHATTAHOOCHEE_TECH_FR...	None			
CHATTAHOOCHEE_TECH_SO...	None			

☒ Set accumulation to the value used for aggregation

Adjustments...

< Previous   Next >   Finish   Cancel   Help

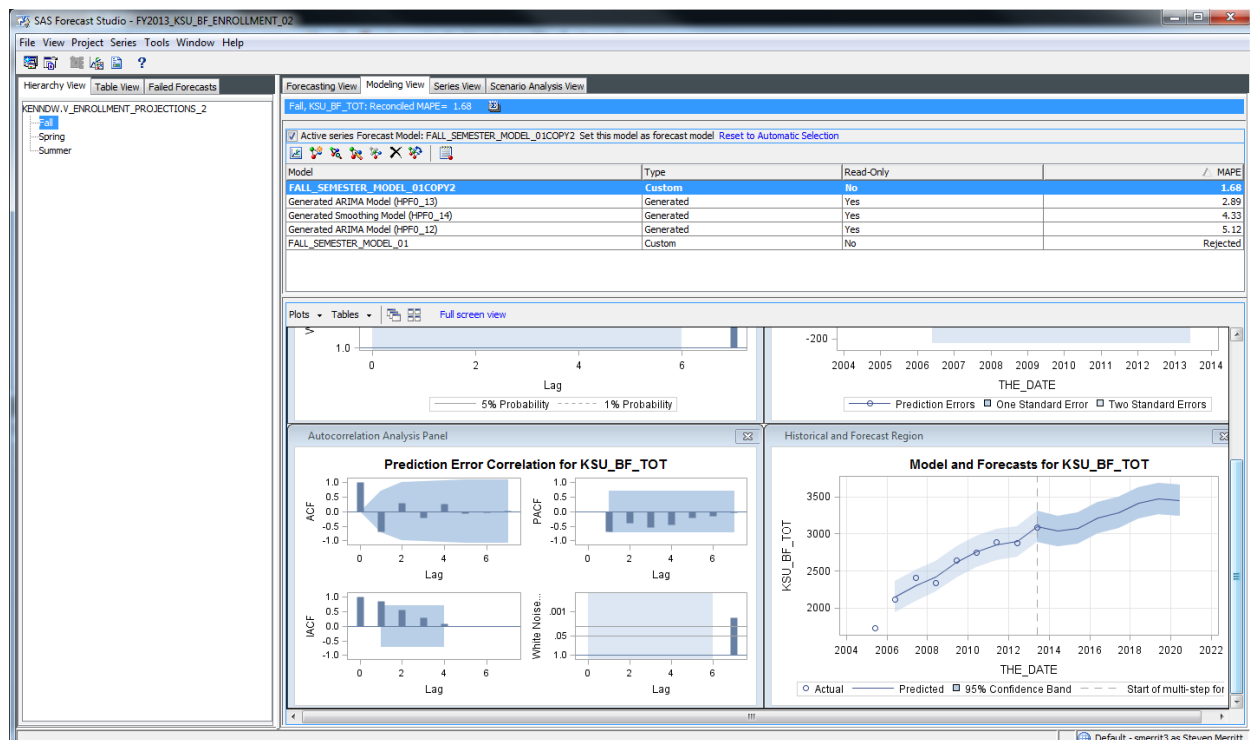
**Display 2. Step 4 of the SAS® Forecast Studio® Project Wizard**

Each variable also has an aggregation and accumulation method. Each variable has the same aggregation and accumulation method, which may differ from the methods of the other variables. We will look at two methods: sum of values and average of values. This characteristic only comes into play when you have a hierarchical model. In a hierarchical model, the method you choose will decide how projections lower in the hierarchy influence projections higher in the hierarchy. For an instance, if you were using population to predict future enrollment and chose sum of values, the population values in each semester would be added together and then used to project future enrollment.

In Step 6, you choose how to handle missing values and whether to ignore data starting before a particular date. In Step 7, you choose how many periods to forecast. This determines how far out you wish to forecast. For instance, if you have set your date variable so that it increments by one year and set the number of periods to forecast to seven, it will forecast seven years into the future.

Clicking the 'Finish' button will create a number of models as seen in Display 3. By default, you will get two ARIMA models and one general fit model. Models are sorted by MAPE, which is calculated based on previous performance of the model vs. actual historical values. The lower your MAPE value the better your model is considered to be. However, a lower MAPE does not always mean that your finished model is a good model; it only means that the historical predictions more closely match the actual historical values.

In some cases, the models created automatically are useful and may suffice for the purposes of your projection. However, in many cases you may wish to modify the model. Since Forecast Studio® is not able to understand what variables represent; it must rely on correlations rather than an actual understanding of the way things work in the real world. For instance, if you wanted to predict graduation rates you may choose to base it off enrollment. We know that it takes roughly four to six years for students to graduate with a bachelor's degree so it makes sense to put in a lag of five or six years when including the variable. Since Forecast Studio® does not understand the meaning of what enrollment or graduation is, much less the intricacies of how they relate to one another, it relies solely on correlation and therefore may make a model which does not take into account the nature of the relationship.



**Display 3. Automatically generated models in SAS® Forecast Studio®**

To edit the model, you can go to the modeling view and click on the 'Edit' button and this takes you to a screen where you can modify the model as shown in Display 3. Since all models generated automatically are read only, you will need to create a copy first or click 'Edit' and then 'Yes' when asked if you wish to make a copy.

In the ARIMA model editor you can set up transformations, decide whether to use an intercept and any autoregressive, differencing and moving average terms you may wish to include in your model within the specifications tab. Within the independent variables tab, you can set up your independent variables, choosing which terms to include and how much lag (if any) will be used as well as differencing order and numerator, denominator polynomials.

The Events tab will allow you to set up events. Events are special occurrences which do not follow normal patterns in the data. As an example, if you were in the retail industry, you may wish to include an event that occurs during the Christmas holiday season. In higher education, you may wish to include an event such as changes in grant money such as the PELL Grant.

Clicking on the 'Ok' button will implement any changes you have made. All changes are automatically saved to your project making it unnecessary to save your work. In fact, the only save feature is a 'Save As' feature which allows you to create a copy of the project you are currently working on. If you make changes to a projection that influences another projection through reconciliation, a 'Reconcile' button will appear. When this button is pressed, all projections will be reconciled.

## CONCLUSION

Kennesaw State University utilizes as part of its institutional planning processes both a short-term, ratio-based enrollment forecast as well as long-range, time series-based enrollment forecast. These forecast models are surfaced using both SAS® Enterprise Guide® and SAS® Forecast Studio®. In SAS® Enterprise Guide®, the ratio-based model is constructed using PROC SQL and traditional SAS® programming techniques. SAS® Forecast Studio® on the other hand allows robust models to be constructed and modified using a graphical user interface.