

## **Weekly Student Contact Hours Forecast Report**

Prepared for the California Community College Chancellor's Office

By the Research and Planning Group of California Community Colleges

June 27, 2011

### **WSCH Forecast Team**

#### **RP Group Members**

Barry Gribbons

Marc Beam

Craig Hayward

Terrence Willett

#### **CCCCO Members**

Fred Harris,

Assistant Vice Chancellor of

College Finance and Facilities Planning

Susan Yeagar,

Specialist, Facilities Planning and Development

**Table of Contents**

Introduction..... 2  
 Review of Forecast Methods..... 3  
 Assumptions ..... 4  
 Summary of Preparatory Analyses ..... 6  
 Analytic Approaches..... 7  
 Recommendations ..... 13  
 Appendix A: Zip Codes with Their Centroid within District Legal Boundary .....A-1  
 Appendix B: Diagnostics and Coefficients for the Ordinary Least Squares (OLS) Regressions.....B-1  
 Appendix C. Autoregressive Integrated Moving Average (ARIMA) Diagnostics and Coefficients.....C-1  
 Appendix D. Historical and Predicted WSCH levels with 90% Confidence Intervals for Selected Districts for the OLS2009 Model ..... D-1  
 Appendix E. Historical and Predicted WSCH Levels with 90% Confidence Intervals for Selected Districts for the ARIMA model ..... E-1  
 Appendix F. An Option for Enrollment Projections that Facilitate Capital Outlay Decisions..... F-1

**List of Figures and Tables**

Figures

Figure 1. Total Annual Credit and Non-Credit WSCH from 1973 to 2009 for the Five Pilot Colleges ..... 7

Tables

Table 1. Output of WSCH Forecasts Using Population Participation Rate (PPR), Ordinary Least Square (OLS) Regression and Autoregressive Integrated Moving Average (ARIMA) Methods ..... 11  
 Table 2. Average Forecasts of Three Methods: Three Year Average Population Participation Rate, Ordinary Least Squares Stepwise Hierarchical Regressions ..... 12

## Weekly Student Contact Hours Forecast Report

The California Community College Chancellor's Office (CCCCO) relies on annual forecasts of Weekly Student Contact Hours (WSCH) in order to estimate the future facilities needs of the 72 California Community College (CCC) districts. The CCCCCO determined that it would be beneficial at this juncture to explore new methods of arriving at WSCH forecasts. The Research and Planning Group of the California Community Colleges (RP Group) agreed to investigate methods for improving WSCH forecasts. The project comprised four distinct components:

- I. Review of current practices as well as practices in other post-secondary systems and in other states
- II. Exploration of data sources
- III. Model development
- IV. Refinement of forecasting procedures in order to create a sustainable process

### Review of forecasting practices

Historically, the Chancellor's Office has conducted enrollment and WSCH forecasts. The following is their description of their methods:

"The Chancellor's Office long-range enrollment and WSCH forecast model is an econometric regression model. The model uses the past actual fall student headcount enrollments for each community college district. In addition, (the model uses) independent variables such as the Department of Finance's county adult population forecasts, the estimated student cost of attendance based upon the Student Aid Commission's Student Expenses and Resources Survey (SEARS) and current mandated student enrollment fees, estimated district budgets, and a factor to indicate pre and post 1978 Proposition 13 cap funding years.

The adult population forecasts are derived from DOF "Race/Ethnic Population Projections with Age and Sex Detail, 1970-2040" based upon the 1990 U.S. census. The Student Aid Commission's SEARS study is usually updated every three years. The future district budgets are based upon a theoretical economic cycle of yearly rate of funding changes ranging from 0.01 to 0.03. The latest projected enrollment is derived from a fall enrollment of districts. This estimated fall enrollment becomes the base year for the projected annual changes from the regression. The last year of actual enrollment and actual WSCH data is used to compute the enrollment to WSCH ratio. This enrollment/WSCH ratio is assumed to be constant for future years and is applied to the projected headcount enrollments to calculate the future WSCH forecasts."

## Review of Forecast Methods

The research team reviewed five models of enrollment forecasts used in four states (CA, IA, MD, TX).

- Maryland Higher Education Commission (MHEC) uses a linear regression model based on the past relationship of enrollments to population using recent high school and community college enrollments, tuition increases, and per capita disposable income. For the five years where comparisons are possible, the MHEC model forecast of the entire Maryland public post-secondary headcount has on average been within 13% of actual. Variances at individual institutions may be considerably larger or smaller .
- California Postsecondary Education Commission (CPEC) uses regression analyses based on life tables of persistence and graduation rates to produce two system-level enrollment scenarios for each public post-secondary system in California: a baseline and mid-range forecast. CPEC uses CA Department of Finance (DOF) population forecasts by age group and ethnicity with participation rates for the same groups (age and ethnicity) to calculate a mean rate of change in participation rates by age. The projected participation rates are then applied to the demographic forecasts to arrive at an enrollment forecast.
- CA CSU Analytic Studies division uses regression analyses based on DOF forecasts of high schools graduates, historical college going rates, and population forecasts for college going rates of specific age cohorts. This approach to enrollment forecasting is sometimes referred to as a stock-and-flow model. CSU annually forecasts up to ten years of enrollment demand for first-time freshmen, undergrad transfers, and graduate students.
- Iowa Department of Education forecasts enrollment in their public four-year universities using local population data in a time series (ARIMA) model with four significant variables: prior year enrollment in community college; number of current high school students; number of high school grads; and a composite variable for the current economic condition based on the national gross domestic product and state unemployment.
- Texas Higher Education Coordinating Board uses multiple regression models to produce three alternative forecasts comparing past enrollments to population forecasts by location, age, and ethnicity by county. The resulting forecast is based on a fixed participation rate for the forecast period.

The research team considered the above models, variables and approaches as we explored options for forecasting enrollment in the 72 individual California Community College districts.

## Assumptions

In conducting analyses of factors and methods for improving Weekly Student Contact Hours (WSCH) forecast, several assumptions and delimitations have to be made. In the future, as new and improved data become available, some of these assumptions may need to be revisited. In this section, we present the assumptions, in part to confirm that the preferred assumptions are being made.

- Annual WSCH (or the annual change/difference in WSCH) is the preferred dependent variable. There are several alternatives that were considered:
  - Fall term WSCH. The major reason for potentially preferring WSCH for the fall term only is that in the summer, some WSCH can be counted in either fiscal year at the college's discretion. In other words, it can be "cashed in" in the current fiscal year or "pushed forward" in the next fiscal year if not needed to maximize apportionment in the current year. Changes to cashing in and pushing forward FTES can be sources of error in models. To compensate, one proposal is to merely use fall data with some annualization based on the ratio of fall term WSCH to annual WSCH. This, of course, has some downside as well in comparison to using actual annual data.
  - Headcount. As an alternative to WSCH, headcount could have been used as the dependent variable. The Chancellor's Office current model projects headcount and then uses that headcount to estimate WSCH. Since WSCH is sensitive to how many classes students are enrolling in, and thus how much time facilities are in use, we elected to forecast WSCH directly.
  - Using total WSCH, including both non-credit and credit WSCH, can have certain limitations and can be impacted by college's shifting enrollment management priorities. In effect, it gives preference to colleges that increase their non-credit offerings. Since a college can generate more non-credit WSCH than credit WSCH with the same capped dollars, the college can qualify for facilities based on WSCH that doesn't differentiate between credit and non-credit if they cut credit sections and replace with even more non-credit sections.
- WSCH can be projected based on a selected set of historical data. Often there are variables that have dramatic impacts on college districts' enrollments that are not foreseeable or that cannot practically or technically be included. For example, in 2006 it would have been difficult to foresee the magnitude of the state budget problems and the impact that they have had on enrollments at community colleges throughout the state.
- In general, factors that are shown to be related to WSCH should be included in forecast equations and factors not shown to be related to WSCH should not. Technical reasons, such as a lack of statistical power due to a small number of data points, could impact significance tests and thus inclusion of factors in models may be based on their theoretical importance as well as their statistical significance (i.e., non-significant factors may be included, as appropriate).
- Some variables have been excluded solely because historical data are not currently available. In the future, as additional data become available, we may be able to evaluate new variables and determine if they effectively predict WSCH.

- Factors that are independently forecast into the future (e.g., demographic forecasts of population numbers) are preferred as they provide a scaffold upon which to base expectations of future WSCH levels in community college districts.
- WSCH data and WSCH forecasts are for academic years (e.g., 2010-2011). However, for the sake of convenience we use the first year of the academic year when referring to WSCH data and forecasts (e.g., 2010 represents 2010-2011). By contrast, the demographic data cover calendar years. Typically, the 2010 calendar year demographic data is used to predict the WSCH level of the academic year that begins with the matching value (e.g., 2010 demographic variables predict 2010-2011 WSCH).

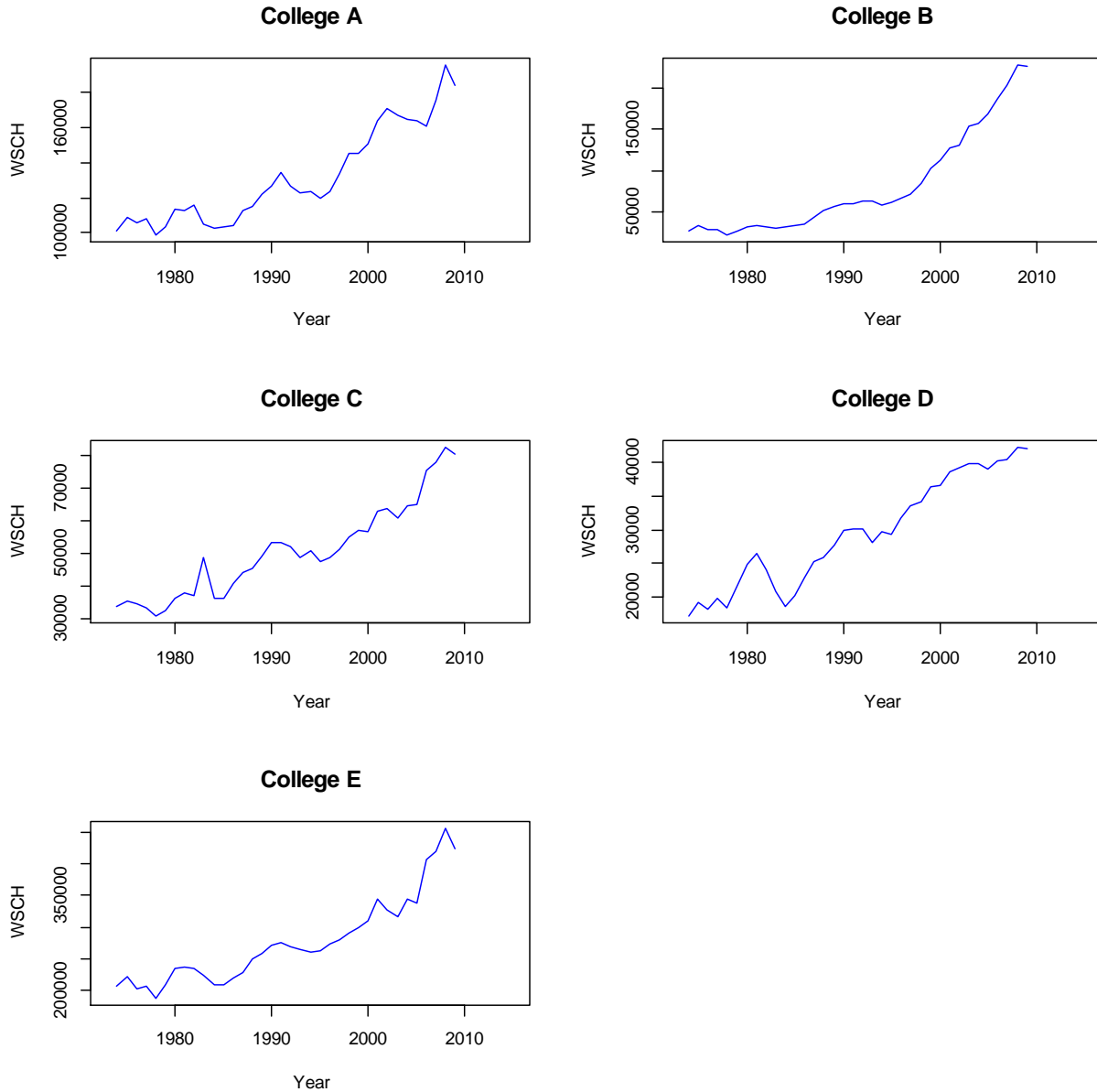
**Summary of Preparatory Analyses**

- For each of the five pilot districts, we have collected data on hundreds of variables, the majority of which were zip code level population counts disaggregated by ethnicity, age, and gender.
- Myriad exploratory analyses were performed to assess the degree to which general approaches (Autoregressive Integrated Moving Average (ARIMA); Ordinary Least Squares (OLS) regression; moving average; Prais-Winsten regression; OLS regression with differenced WSCH; Cochrane-Orcutt regression) and which predictor variables (e.g., zip code-level data disaggregated by ethnicity, age, and gender; interpolated data for missing years, historic budget data, etc.) yielded the best results.
- Several approaches were taken to control for autocorrelation (a serious issue with time series analyses that can lead to overly optimistic models). Using the annual change in WSCH (i.e., differenced WSCH) in an OLS regression provides good results for most scenarios.
- Analyses of Durbin-Watson statistics, the Box-Ljung Q test, Autocorrelation Functions (ACF), and Partial Autocorrelation Functions (PACF) were completed to assess the impact of autocorrelation.
- Eventually, attention focused on seventeen promising indicators or predictor variables from the following three categories: district level population counts segmented by age groups; district level population counts segmented by ethnicity; and a small set of economic/financial indicators.
- The annual change in WSCH (or differenced WSCH) was regressed on the seventeen most consistent predictor variables numerous times for each of the five pilot districts in order to assess the performance of different model selection techniques (e.g., forward entry, remove, backward, stepwise entry, etc.).

**Analytic Approaches**

Figure 1 shows the historical WSCH data for each of the five pilot colleges. Note that over the last three decades, the colleges have shown increases in WSCH over time with variability in growth rates from year to year.

**Figure 1. Total Annual Credit and Non-Credit WSCH from 1973 to 2009 for the Five Pilot Colleges**





To forecast these WSCH into future years, three main approaches were employed: Population Participation Rates, Ordinary Least Squares Regressions, and Autoregressive Integrated Moving Averages.

### **1. Population Participation Rate (PPR) Method**

The PPR method uses the proportion of residents in the district boundary who are students to predict the number of future students based on future population changes. The PPR is applied to forecasts for future population levels to create a predicted headcount. The headcount is converted to WSCH by multiplying an average WSCH to headcount (WSCH/HC) ratio. Headcounts are annual credit and non-credit students (STD07=A, B, C, or F). This method assumes constant PPR and WSCH to headcount ratios in the future. Three PPR models were created:

#### **1) Three year average PPR for district legal boundary population (3 Year Average PPR Legal)**

The 3 Year Average PPR Legal model was based on the average PPR from 2006 to 2008 in the district's legal boundary. The WSCH/headcount ratio used was the average WSCH/headcount from 2006 to 2008.

#### **2) Maximum PPR for district legal boundary based on 2000 to 2008 (Max PPR Legal)**

The Max PPR Legal model was based on the highest or maximum PPR from 2000 to 2008 in the district's legal boundary. The WSCH/headcount ratio used was the highest or maximum WSCH/headcount from 2000 to 2008.

#### **3) Three year average PPR for district service area population based on 2006 to 2008 (SA PPR)**

The SA PPR model was based on the average PPR from 2006 to 2008 in the district's service area. The WSCH/headcount ratio used was the average WSCH/headcount from 2006 to 2008.

### **2. Ordinary Least Squares (OLS) stepwise regression on differenced WSCH**

Claritas provided a host of population variables that can be independently and reliably forecast. These variables are extremely useful for WSCH forecasting. If a relationship can be established between WSCH and any of these demographic variables, then the WSCH forecast can piggyback on the demographic forecasts. Because demographic variables have established performance, theory and methods underlying their forecasts, they have the potential to provide a strong scaffold upon which to build WSCH forecasts. We evaluated four categories or blocks of annual demographic variables for their utility in forecasting WSCH:

1. Population counts divided by age groupings
2. K-12 enrollments
3. Population counts divided by ethnic group
4. Other demographic variables (unemployment, poverty levels, education levels, etc.)

The most consistently useful and reliable variables in the five pilot districts were found in the first block, population counts divided by age groupings. Simple correlations and other inspections identified candidate predictor variables for each district. Exploratory custom regressions were created for each

district to help determine a common method that could be applied to all districts. Presented here are regressions using a hierarchical stepwise method to predict annual changes in WSCH. The three blocks used in the regression were:

- A. Total population in district legal boundary by age category,
- B. Total population in district legal boundary by ethnic category,
- C. Economic variables: State budget for community colleges; the annual change in State budget for CCCs; and a binary recession indicator.

Due to the relatively small number of years for which there are relevant data points, the probability of significance required for a predictor to enter was set at 0.20 and the probability to remove was set at 0.25 rather than the customary 0.05 and 0.10. One set of regressions included data only up to 2008 to compare the predicted to actual 2009 WSCH. The other set includes data up to 2009 to show how the addition of the most recent data point can change the model's output.

### **3. Autoregressive Integrated Moving Average (ARIMA)**

A forecast method that takes advantage of all the available WSCH data back to 1974 is called Autoregressive Integrated Moving Average (ARIMA). This econometric time series analysis technique is useful for large sets of time series data. It has parameters that can be set to correct for both autocorrelation and seasonality. This method often requires implementing transformations on the source data in order to detrend the data series prior to modeling. To determine possible transformations, autocorrelation function (ACF) and partial autocorrelation function (PACF) charts of residuals were examined in addition to Ljung-Box statistics for various lags. It appeared that the simplest reasonable transformations was to use the differenced lagged WSCH with an autoregressive factor of order 1 ( $p=1, d=1, q=0$ ) with no adjustment for seasonality. While further customization by district was possible, it appeared that applying these transformations may generally lead to acceptable residual statistics. To help validate the ARIMA model, the last available year of WSCH (2009) was excluded so that the first year forecast could be compared to the actual value.

Table 1 shows the output of WSCH forecasts using PPR, OLS, and ARIMA methods. WSCH for 2009 was held back from use in all of the forecast models (except for OLS2009, which included WSCH for 2009) and then predicted and compared against actual WSCH for that year for a one year forecast (1 Year WSCH Absolute and Percent Difference Columns). Predicted WSCH for five years out is also shown along with the actual change in WSCH relative to the base year. For forecasts based on the 2008-2009 academic year, the 2013-2014 academic year represents the fifth year out and for the OLS2009 method the base year is 2009-2010 with the fifth year out of 2014-2015.

The 3 year average PPR legal was more likely to underestimate 2009 WSCH than OLS2008 but was closer to the true WSCH for 3 of the 5 pilot colleges. Using the maximum PPR legal resulted in forecasts 10% to 27% higher than the 3 year average PPR legal model. The service area PPR model was similar to and generally lower than the 3 year average PPR legal model. Service area PPR long term model forecasts ranged from 3% lower than to 1% higher than the 3 year average PPR legal model forecasts.

The OLS2008 model had higher one year forecasts than the 3 year average PPR legal model or ARIMA model in most cases. Longer term forecasts tended to retain much higher growth rates than any of the PPR models or the ARIMA model. The OLS2009 model showed some strong differences from the OLS2008 model with two of the forecasts becoming negative (College A and College D), one having much more subdued growth (College C), one remaining very similar (College E), and one showing a much higher growth rate (College B). Comparing OLS2008 and OLS2009 shows how a method can be sensitive to a single additional data point in dynamic times. OLS forecasts also provide a range of values where the true future values are expected to be within a specified level of confidence. OLS2009 forecasts with 90% confidence intervals are shown in Appendix D.

ARIMA forecasts for the first year out were higher than the 3 year average PPR legal model in four out of five cases and lower than the OLS2008 model in four out of five cases. The forecasts for subsequent years quickly became very flat suggesting the ARIMA is more sensitive to short term forecasting. ARIMA forecasts with 90% confidence intervals are shown in Appendix E.

Forecasts can be subject to biases due to the method or inputs used. It has been shown that averages of forecasts can help moderate biases and provide more accurate forecasts (Makridakis and Winkler 1983; Palm and Zellner 1992). Table 2 compares data from three promising approaches: a) the average of the Three Year Average Population Participation Rate and the Ordinary Least Squares Stepwise Hierarchical Regressions Including Data up to 2008, b) the average of the Maximum Population Participation Rate and the Ordinary Least Squares Stepwise Hierarchical Regressions Including Data up to 2008, and c) the Maximum Participation Rate.

**Table 1. Output of WSCH Forecasts Using Population Participation Rate (PPR), Ordinary Least Squares (OLS) Regression and Autoregressive Integrated Moving Average (ARIMA) Methods.**

Method	College Name	1 Year WSCH Forecast	1 Year Actual WSCH	1 Year WSCH Absolute Difference	1 Year WSCH Percent Difference	5 Year WSCH Forecast	5 Year Forecast Absolute Change	5 Year Forecast Percent Change
<b>3 Yr Avg PPR Legal</b>	A	179,781	183,746	-3,965	-2.2%	178,321	-5,425	-3.0%
<b>Max PPR Legal</b>	A	199,656	183,746	15,910	8.7%	198,039	14,293	7.8%
PPR SA	A	176,457	183,746	-7,289	-4.0%	175,053	-8,693	-4.7%
<b>OLS2008</b>	A	219,185	183,746	35,439	19.3%	258,924	75,178	40.9%
OLS2009	A	179,340	na	na	na	165,637	-13,703	-7.6%
ARIMA	A	200,564	183,746	16,818	9.2%	202,398	18,652	10.2%
<b>3 Yr Avg PPR Legal</b>	B	212,745	224,922	-12,177	-5.4%	228,278	3,356	1.5%
<b>Max PPR Legal</b>	B	270,878	224,922	45,956	20.4%	291,086	66,164	29.4%
PPR SA	B	210,550	224,922	-14,372	-6.4%	220,939	-3,983	-1.8%
<b>OLS2008</b>	B	247,727	224,922	22,805	10.1%	322,083	97,161	43.2%
OLS2009	B	196,797	na	na	na	391,066	194,269	98.7%
ARIMA	B	236,833	224,922	11,911	5.3%	243,011	18,089	8.0%
<b>3 Yr Avg PPR Legal</b>	C	82,263	80,432	1,831	2.3%	85,523	5,090	6.3%
<b>Max PPR Legal</b>	C	100,333	80,432	19,901	24.7%	104,269	23,836	29.6%
PPR SA	C	80,502	80,432	70	0.1%	83,583	3,151	3.9%
<b>OLS2008</b>	C	89,484	80,432	9,052	11.3%	145,684	65,252	81.1%
OLS2009	C	79,408	na	na	na	94,382	14,974	18.9%
ARIMA	C	81,552	80,432	1,120	1.4%	81,705	1,273	1.6%
<b>3 Yr Avg PPR Legal</b>	D	40,641	42,009	-1,367	-3.3%	40,273	-1,736	-4.1%
<b>Max PPR Legal</b>	D	50,576	42,009	8,567	20.4%	50,156	8,147	19.4%
PPR SA	D	39,379	42,009	-2,630	-6.3%	39,103	-2,906	-6.9%
<b>OLS2008</b>	D	44,770	42,009	2,761	6.6%	60,340	18,331	43.6%
OLS2009	D	42,247	na	na	na	41,841	-406	-1.0%
ARIMA	D	42,832	42,009	824	2.0%	43,000	991	2.4%
<b>3 Yr Avg PPR Legal</b>	E	431,303	423,225	8,079	1.9%	437,385	14,161	3.3%
<b>Max PPR Legal</b>	E	473,837	423,225	50,612	12.0%	480,539	57,314	13.5%
PPR SA	E	430,716	423,225	7,492	1.8%	441,383	18,158	4.3%
<b>OLS2008</b>	E	457,009	423,225	33,784	8.0%	521,942	98,717	23.3%
OLS2009	E	421,049	na	na	na	474,107	53,058	12.6%
ARIMA	E	459,429	423,225	36,205	8.6%	459,709	36,485	8.6%

Notes: Methods with names in **bold** are used in the averaged forecast shown in next table.

3 Yr Avg PPR = Average population participation rate from 2006 to 2008 using district legal boundaries. See Appendix A.

Max PPR = Highest population participation rate from 2002 to 2008 using district legal boundaries.

PPR SA = Average population participation rate from 2006 to 2008 using district service area (zip codes that contribute at least 1% to total college enrollment in any fall term).

OLS2008 = Ordinary Least Squares hierarchical stepwise regressions including data up to 2008. See Appendix B.

OLS2009 = Ordinary Least Squares hierarchical stepwise regressions including data up to 2009. See Appendix B.

ARIMA = Autoregressive Integrated Moving Average. See Appendix C.

**Table 2. Forecasts Using Three Methods: Three Year Average Population Participation Rate, Ordinary Least Squares Stepwise Hierarchical Regressions Including Data up to 2008, and Autoregressive Integrated Moving Average.**

Method	College Name	1 Year WSCH Forecast	1 Year Actual WSCH	1 Year WSCH Absolute Difference	1 Year WSCH Percent Difference	5 Years WSCH Forecast	5 Year Forecast Absolute Change	5 Year Forecast Percent Change
Average of 3yr PPR and OLS2008	A	199,483	183,746	15,737	8.60%	218,622	34,876	19.00%
Average of Max PPR and OLS2008	A	209,420	183,746	25,675	14.00%	228,481	44,735	24.30%
Max PPR Legal	A	199,656	183,746	15,910	8.70%	198,039	14,293	7.80%
Average of 3yr PPR and OLS2008	B	230,236	224,922	5,314	2.40%	275,181	50,259	22.30%
Average of Max PPR and OLS2008	B	259,303	224,922	34,381	15.30%	306,584	81,663	36.30%
Max PPR Legal	B	270,878	224,922	45,956	20.40%	291,086	66,164	29.40%
Average of 3yr PPR and OLS2008	C	85,874	80,432	5,441	6.80%	115,603	35,171	43.70%
Average of Max PPR and OLS2008	C	94,908	80,432	14,476	18.00%	124,976	44,544	55.40%
Max PPR Legal	C	100,333	80,432	19,901	24.70%	104,269	23,836	29.60%
Average of 3yr PPR and OLS2008	D	42,706	42,009	697	1.70%	50,306	8,298	19.80%
Average of Max PPR and OLS2008	D	47,673	42,009	5,664	13.50%	55,248	13,239	31.50%
Max PPR Legal	D	50,576	42,009	8,567	20.40%	50,156	8,147	19.40%
Average of 3yr PPR and OLS2008	E	444,156	423,225	20,932	4.90%	479,663	56,439	13.30%
Average of Max PPR and OLS2008	E	465,423	423,225	42,198	10.00%	501,240	78,016	18.40%
Max PPR Legal	E	473,837	423,225	50,612	12.00%	480,539	57,314	13.50%

**Recommendations**

1. Forecasts should be arrived at by averaging the results of the OLS regression method with one of the PPR methods if resources permit. These methods each have desirable characteristics that balance each other's weak points, for example:
  - a. PPR is a stable method that does not rely on advanced statistical techniques.
  - b. OLS is sensitive to changes from year to year, but is subject to significant variation from year to year as well. With the inclusion of a single additional year of data that reflect a significant change from prior years' data, the results could change the forecast significantly. This is especially true of apparent inflection point, such as the beginning or end of a recession.
  - c. OLS regressions have a tendency to reach extreme values when extrapolating far into the future.
  - d. Max PPR is an interesting approach in that it simulates what the enrollments might be if the state fully funded enrollments at community colleges. Therefore, it is more reflective of demand and less reflective of restrictions in historical funding when modeling future enrollments.
  - e. PPR tends to produce conservative estimates of growth which - though not always sensitive to year to year changes - provides a good balance to the more extreme forecasts obtained via OLS regression.
2. If the Chancellor's Office allows districts to propose alternative models to the one used by the Chancellor's Office, the district should be required to commit to using the same approach for five years in order to prevent districts from changing methods merely for a short term effort to drive up their projections. District personnel should also be required to submit complete documentation of their alternate method such that the method could be replicated.
3. Use aggregated zip codes that define a district's boundaries rather than county level data. Using the zip codes for all student residences, even those outside of the district boundaries, did not improve the enrollment forecast models for the five pilot districts.
4. Update the numbers annually as data are released.
5. Reevaluate the methods after five years. Additional data and other resources may become available in the future to make other models feasible. An example of one such model, which could be considered an extension of the population participation rate model, is described in Appendix F. Information included in Appendix F contains information that extended beyond the purpose of this report especially related to recommendations on how to implement the models. However, the information contained in Appendix F also serves as a good example of an alternative model that should be regularly evaluated in the future as data and resource availability change to ensure that the most effective and appropriate models are used.
6. Use data that match in time (all annual data or all fall data) and type of enrollments (e.g. credit/non-credit combinations or credit only) to minimize sources of error when possible.
7. Review forecasts annually to help detect unintended large shifts in forecasts and compare to previous forecasts to detect changes in forecast method assumptions such as WSCH to headcount ratios. An objective criteria for additional model review could be created such as

when more than half of the districts show a forecast that are more than 10% different than the previous forecast.

8. When feasible, confidence intervals should be produced even if point estimates are used. The confidence intervals can be useful in conveying a expected range in enrollments that are expected. Enrollments that fall outside of the expected range or confidence interval can be useful as a diagnostic tool in helping to determine when significant changes are affecting the model.
9. To arrive at a ten year projection, a linear extrapolation could be used since most population data that could serve as an input would be limited to 5 years.

In addition to these general recommendations, we provide two specific options. Option A may have more technically desirable properties, but also requires more resources than Option B. Option B is a reasonable alternative if resources are limited.

#### **Option A. Average of PPR and OLS Regression**

Option A is to average a population participation rate (PPR) method and an ordinary least square (OLS) regression approach.

1. For the PPR method, the projected population at five years would be multiplied by the average participation rate (the number per thousand residents who attended the college district) of the past three years. To convert to WSCH, the resulting number of projected students would be multiplied by the average WSCH per headcount for the past three years. These data would be extrapolated to a 10 year projection by doubling the five year increase.
2. For OLS, variables would be entered using three steps. The first steps would enter the number of people within the district's legal boundaries area by the following age groups:
  - Population 0 to 4
  - Population 5 to 9
  - Population 10 to 14
  - Population 15 to 17
  - Population 18 to 19
  - Population 20 to 24
  - Population 25 to 34
  - Population 35 to 44

The next block of variables includes population data by ethnicity:

- Population that is African American
- Population that is Latino
- Population that is Asian
- Population that is White

The last block of variables to enter is financial and economic variables, such as the following:

- Recession (0 if there was no official economic recession in that calendar year; 1 if there was a recession. Future years are forecast as no recession, but this could be adjusted as desired).
- State Budget (CCC system budget from the LAO)
- Differenced State budget (The annual change in historic and forecast state budget funding levels)

Using the stepwise approach, a separate regression equation is obtained for each district using the variable that is significant and the corresponding weights. The regression equation is used to obtain annual projections to year 10.

3. The projection from the OLS and PPR method are averaged.

**Option B. Max PPR.**

Option B uses a maximum population participation rate. For the MAX PPR method, the projected population at five years would be multiplied by the highest participation rate (the number per thousand residents who attended the college district) in the past five years. To convert to WSCH, the resulting number projected students would be multiplied by highest WSCH per headcount ratio from the past five years. These data would be extrapolated to a 10 year projection by doubling the five year increase.



## References

Makridakis, S. and Winkler, R. L. (1983). Averages of Forecasts: Some Empirical Results. *Management Science*. 29(9): 987-996.

Palm, F. C. and Zellner, A. (1992). To Combine or not to Combine? Issues of Combining Forecasts. *Journal of Forecasting*. 11: 687-701.

Pindyck, R. S. and Rubinfeld, D. L. (1981). *Econometric Models and Economic Forecasts*, 2<sup>nd</sup> Edition. San Francisco: McGraw Hill.

Sanchez, J. Introduction to Time Series: A Forecasting Model for Absorbent Paper Towels.  
<http://www.stat.ucla.edu/~jsanchez/teaching/course170/datasets/towelscasestudy.pdf> retrieved 3/1/2011.

Yin, P. and Fan. X. (2001). Estimating  $R^2$  Shrinkage in Multiple Regression: A Comparison of Different Analytical Methods. *The Journal of Experimental Education*. 69(2): 203-224.

## Appendix A.

### Zip Codes with Their Centroid within District Legal Boundary

A	B	C	D	E
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX	9XXXX	9XXXX	9XXXX	9XXXX
9XXXX		9XXXX	9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
9XXXX			9XXXX	9XXXX
			9XXXX	9XXXX
			9XXXX	9XXXX
				9XXXX

Population forecasts available are at the zip code level. To estimate future populations in the college district legal boundary, we created a set of zip codes for each district’s legal boundary. The zip codes listed in the table above have their centers within the legal boundary of each district. The total population living in each set of zip codes approximates the population within the district’s legal boundaries. Note that these zip codes are based upon boundaries in the year 2000. Zip code boundaries change over time and should be reviewed periodically.

## Appendix B.

### Diagnostics and Coefficients for the Ordinary Least Squares (OLS) Regressions

College Name	Last Year Included	R <sup>2</sup>	Adj. R <sup>2</sup> (Ezekial)	Adj. R <sup>2</sup> (Browne)*	Durbin-Watson	Equation
A	2008	0.920	0.872	<b>0.882</b>	2.34	WSCHdiff = 6.727*Age20to24Popn - 18.120*Age15to17Popn + 8409.519*Recession + 61,519.341
B	2008	0.240	0.132	<b>0.240</b>	3.21	WSCHdiff = 7.842*Age18to19Popn - 26,711.173
C	2008	0.274	0.170	<b>0.274</b>	3.03	WSCHdiff = -6.033*Age35to44Popn + 16,1393.853
D	2008	0.303	0.204	<b>0.303</b>	2.94	WSCHdiff = -11.144*Age18to19Popn + 26,572.072
A	2008	0.229	0.118	<b>0.229</b>	3.00	WSCHdiff = 12.315*AgeUnder5Popn - 769,953.542
A	2009	0.468	0.317	<b>0.404</b>	1.36	WSCHdiff = 14,006.65*RecessionFlag + 0.019664*DifferenceStateBudget - 5,969.830
B	2009	0.301	0.214	<b>0.301</b>	3.49	WSCHdiff = 21.524*Age5to9Popn - 290,925.012
C	2009	0.214	0.116	<b>0.214</b>	2.89	WSCHdiff = 10.700*Age10to14Popn - 141,884.000
D	2009	0.257	0.164	<b>0.257</b>	2.67	WSCHdiff = -9.565*NonHispBlackPopn + 7,669.187
E	2009	0.509	0.369	<b>0.447</b>	3.06	WSCHdiff = 28.672*Age0to4Popn - 19.335*Age15to17Popn - 1,000,048.000

\*Browne adjusted R-square =  $[(N-k-3)*R^2 + R^2]/[(N-2k-2)*R^2+k]$  (see Yin and Fan 2001).

## Appendix C.

### Autoregressive Integrated Moving Average (ARIMA) Diagnostics and Coefficients.

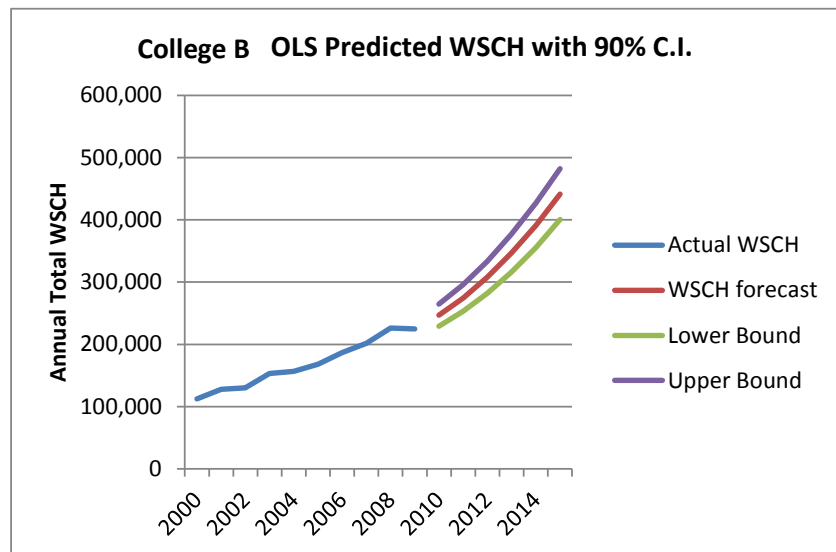
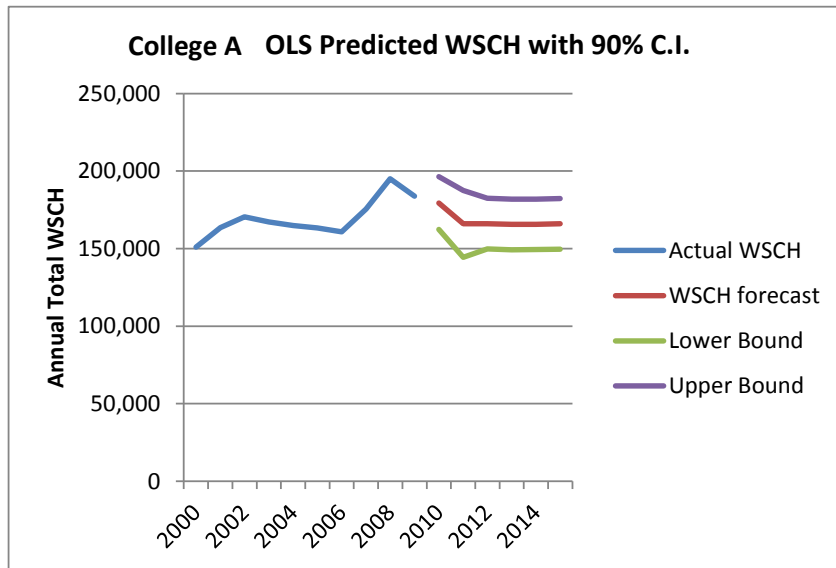
College Name	AIC	Autoregressive Coefficient	Autoregressive Standard Error	p-value
A	-99.84	0.27	0.18	0.034
B	-53.97	0.40	0.16	0.003
C	-62.12	-0.19	0.17	0.067
D	-77.24	0.26	0.17	0.030
E	-85.32	0.08	0.18	0.159

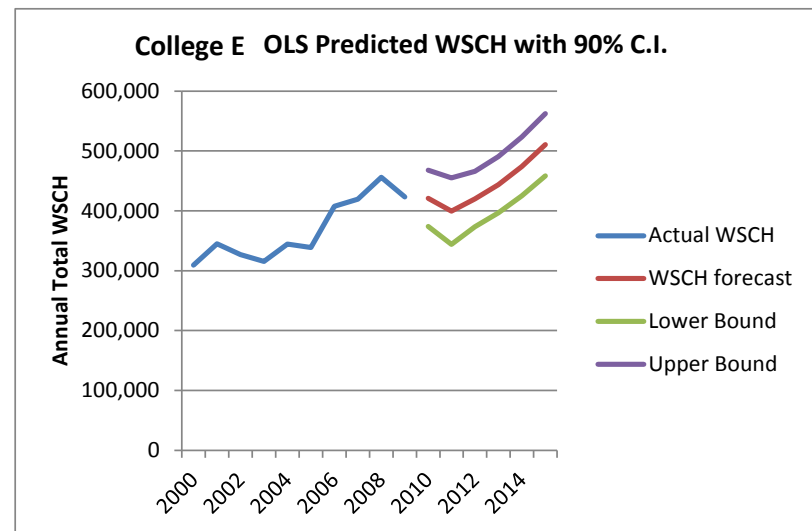
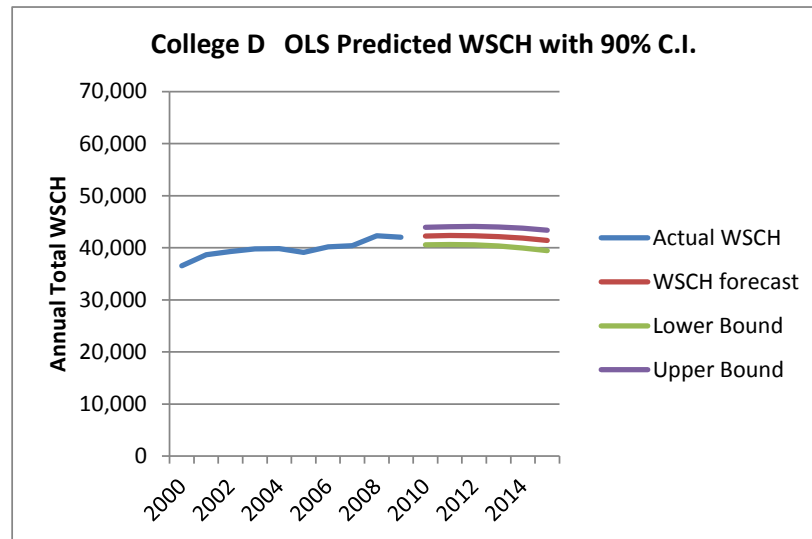
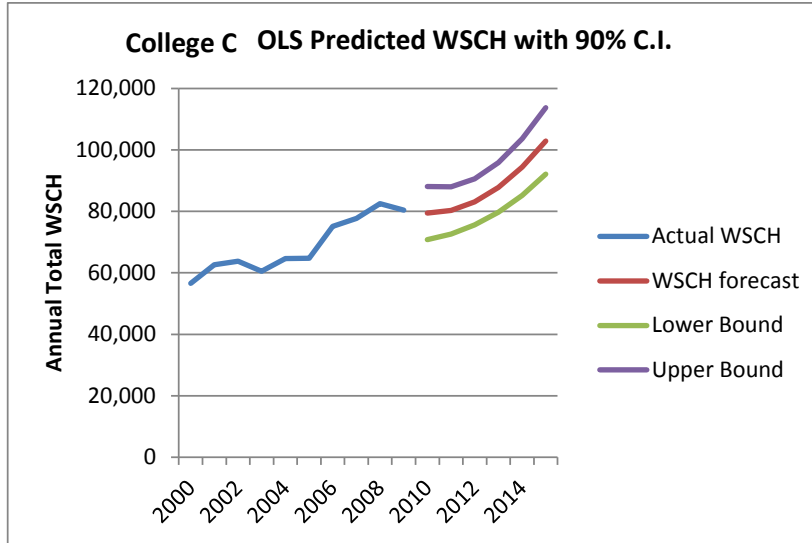
AIC = Akaike Information Criteria

The table above shows model fit parameters for the ARIMA model for each of the five pilot districts. The forecasts from this model are generally flat with very wide confidence intervals after the first year. It is suggested that only the first year forecast be considered as a useful forecast (see also Sanchez, J. and Pindyck and Rubinfeld 1981).

## Appendix D.

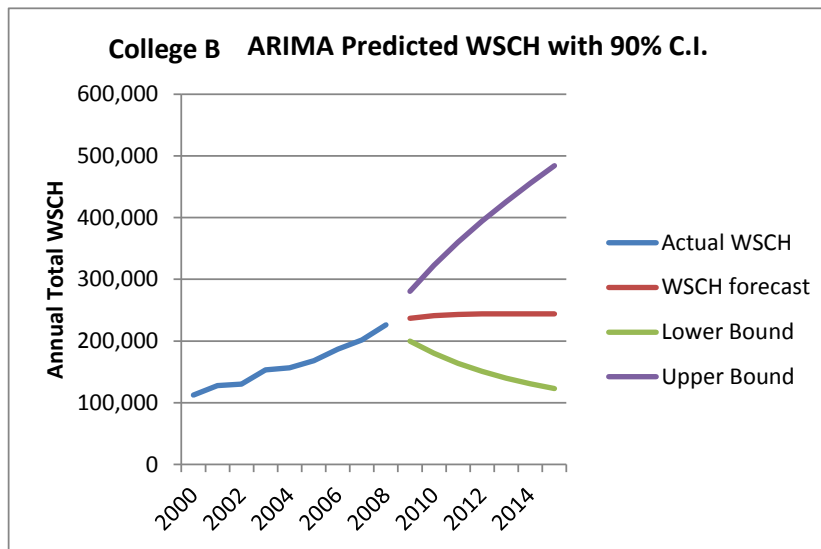
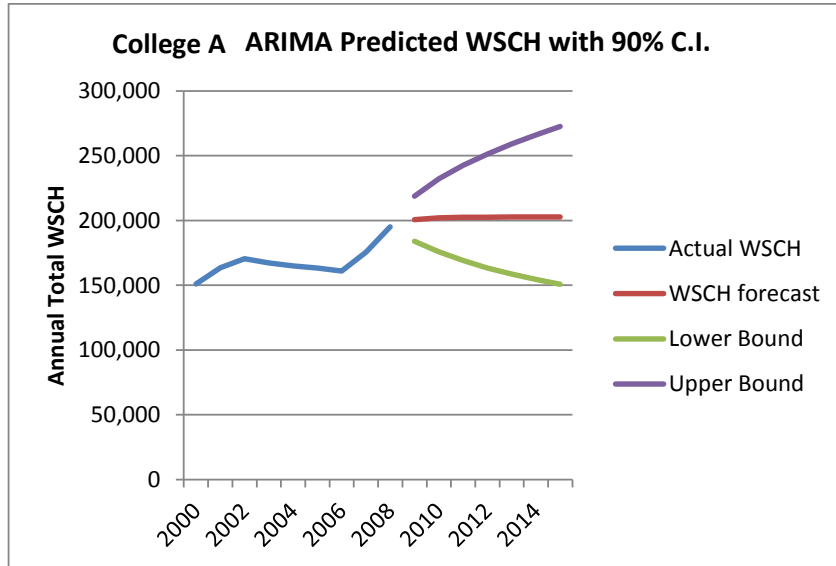
Historical and predicted WSCH levels with 90% confidence intervals for selected districts for the OLS2009 model (see Table 1).

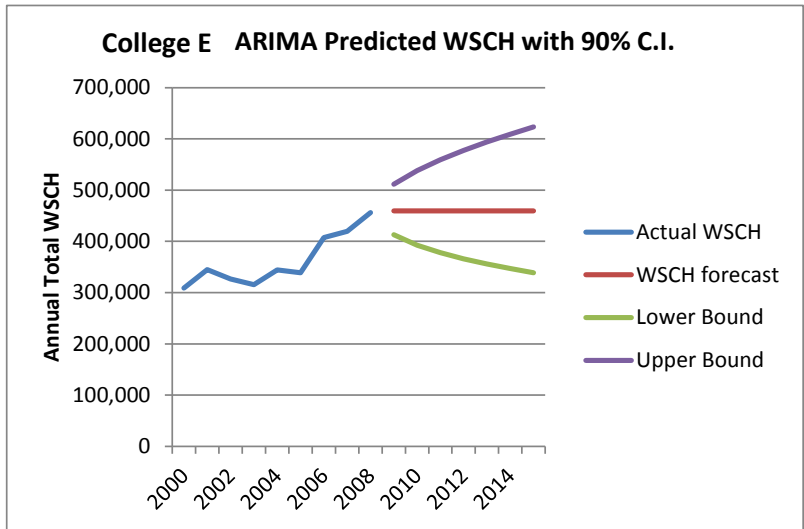
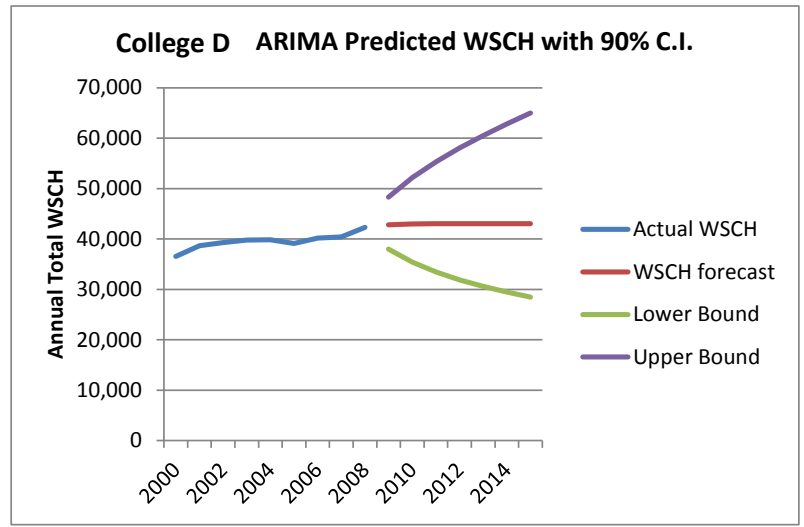
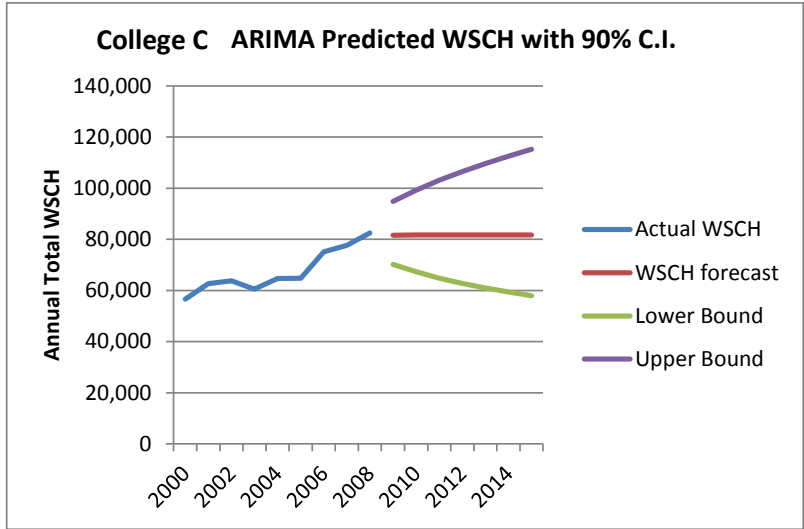




## Appendix E.

Historical and predicted WSCH levels with 90% confidence intervals for selected districts for the ARIMA model (see Table 1).







## **Appendix F.**

### **An Option for Enrollment Projections that Facilitate Capital Outlay Decisions**

Willard Hom, Director for Research, Analysis and Accountability

March 26, 2011

#### **Introduction**

Each community college (CC) district in California must submit to the Chancellor's Office documentation of expected growth in enrollment in order to obtain approval for any capital outlay plan and its corresponding budget support. Funding and oversight bodies (the California Department of Finance, the Legislative Analyst's Office, and the California Postsecondary Education Commission use documentation of projected enrollment growth to decide on whether they will approve or reject a specific capital outlay plan or request. The public oversight bodies ordinarily strive to assure that the tax dollars that eventually fund a capital outlay request constitute an efficient (i.e., not wasteful) use of public funds. Thus, these entities tend to use a fiscally conservative perspective (i.e., pessimistic about growth) in comparison to the CC districts which ordinarily tend to use a fiscally liberal perspective (i.e., optimistic about growth). Historically, the research unit in the Chancellor's Office was expected by the oversight bodies to function as a neutral party that would produce an objective estimate of the 5-year enrollment levels for each district in the state.

#### **Current Dilemma**

The authority for the research unit to produce objective estimates of 5-year enrollment levels per district has not been officially recognized, and no budget for staffing to perform these tasks has been allocated. Historically, the performance of the 5-year enrollment projections occurred under an informal understanding between districts and the oversight bodies. In reality, the system needs a neutral and expert party that can maintain the trust of both the oversight bodies and the districts by producing valid and credible estimates of 5-year enrollment levels. So it makes sense to authorize the research unit in the Chancellor's Office to use professionally accepted methods and data to develop and issue objective enrollment estimates that both districts and oversight bodies can accept upon review.

Now what are professionally accepted methods and data in this situation? In one respect, the historical approach that the research unit used (the approach begun by Dr. Chuck McIntyre) could be considered as professionally accepted in terms of method and data. However, that historical approach succeeded in an environment that differs radically from the environment of 2011. The primary environmental differences are the current cuts to funding for enrollment and for growth in capacity, and the widening disparities between enrollment patterns of different districts. The new environment requires a different

approach whereby the enrollment estimates (1) separate demand from supply, the budget, and (2) analyses account for district differences in the statistical models.

### **The Basic Option**

A basic option to the historical approach has the following steps:

1. Estimate a district's demand for CC education with an education market model.
2. Estimate a district's commitment for serving demand that arises from residents from another district.
3. Assume a set of budget (supply) scenarios that may occur for the 5-year horizon.
4. Evaluate the level of unmet education demand with each budget scenario given in the preceding step.
5. Obtain a decision from oversight/budgeting bodies as to the level of the demand that it will fund.
6. Given the preceding decision, a brief analysis should determine if a specific capital outlay plan is necessary in that the level of funded enrollment will exceed the existing capacity of a district's physical plant.

Step 1 is a major revision of the historical approach because it calls for the development of a model that analysts customize for each district. Past practice used a "one-size-fits-all" process that led to some forecasting anomalies. However, that one-size-fits-all process was necessary given the limited staff resources that the agency could allocate to this function. The increase in the labor requirement of Step 1 is only one hurdle for it. A second hurdle is the acceptance of different models by the different stakeholders.

Step 2 is a major revision of the historical approach to estimating total district "need" because it separates demand driven by residents of a district from demand driven by individuals who reside outside of a district. We can basically estimate the demand from district residents in Step 1 through demographic models of neighborhoods (or zip code tracking areas). On the other hand, we need to estimate the extra-district demand with some other data source. This alternate data source could use service area data to estimate the level of expected extra-district enrollment. One complication that this analysis for Step 2 will need to address is the potential effect of decisions by the neighboring community colleges upon extra-district enrollment. For example, if a neighboring district has historically "exported" 200 FTES to a district but that neighboring district will open a new satellite campus in a few years, then Step 2 should account for the expected effect of a future expansion in capacity in the neighboring district. That is, we should not continue to assume that the neighboring district will continue to "export" 200 FTES to a district.

Step 3 is new because the existing process has usually assumed a single scenario for budgets for the 5-year horizon. Note that we could simplify the work of Step 3 by choosing one possible budget scenario and using that selected scenario for every district estimate in the state. Analysts could use a more elaborate system that would measure levels of belief among stakeholders or experts for each considered budget scenario, and the resulting measurements would act as weights for each budget scenario in a final composite measure of an expected budget.

Step 4 may conceivably find that a district will have a drop in its population such that it either has no unmet demand or it has surplus capacity. In addition, other events or policy changes, such as large increases in the costs for enrollment, could also conceivably create a finding of no unmet demand or of excess capacity. On the other hand, seemingly excess capacity may be justified if historical participation rates are so low that the stimulation of demand through the provision of new supply (new facilities) would be in the public interest.

Steps 5 and 6 are administrative tasks rather than research and analysis tasks, and this option does not designate such tasks as work in the research unit. I show them as part of the plan in order to complete the picture of the proposed process. The value in delineating these two steps lies in our focus of attention on these two critical steps as judgment calls rather than as research.

### **To Infinity and Beyond**

The Basic Option given above is a major revision of the historical approach. But if stakeholders wanted a more comprehensive approach that would really aid strategic planning in the districts, then there are many enhancements that a research unit could make, given the proper resources and authority. These enhancements would primarily allow analysts to implement a market demand model for each district in much the same way that major retail businesses evaluate future investment opportunities. Examples of this kind of enhancement are swirl studies, modeling of spatial competition, and cost of enrollment analyses, among others. However, such enhancements would make sense only if stakeholders were to support at a minimum the Basic Option. So further consideration of enhancements should await acceptance of the Basic Option.

### **Acknowledgements**

The WSCH Forecast Task Force would like to thank the following people for their input and contributions:

Willard Hom, Director of Research, CCCCCO

Mei Cooc, Research Analyst, CCCCCO

Myrna Huffman, Director of Information Systems, CCCCCO

Patrick Perry, Vice Chancellor of Technology, Research, and Information, CCCCCO

John Roach, Director of Systems Analysis and Research, FCCC

Chris Yatooma, Administrator, State Apportionment Process & System Budget Development

Phil Garcia, Director of Research, CSU Analytics

Matt Wetstein, Dean of Planning, Research, and Institutional Effectiveness

Nathan Pellegrin, Statistician

Presidents/Superintendents of participating college districts