Memorandum

TO: Thierry Prate, PB
Ridership and Revenue Peer Review Panel

FROM: Rachel Copperman
Jeff Buxbaum
Weimin Huang

DATE: January 8, 2014

RE: California High Speed Rail Ridership and Revenue Forecasts
Methods and Forecasts including Risk Analysis

CS employed a risk analysis approach to our ridership and revenue forecasts for the 2014 Business Plan. The risk analysis model includes a range of assumptions for the factors that we believe have the greatest impact on high-speed rail ridership and revenue. We ran the Version 2.0 Ridership and Revenue Model numerous times with varied input assumptions for each forecast year to develop the information necessary for the estimation of regression models of High Speed Rail (HSR) revenue. Working with assumptions regarding the probability distributions for each of the risk factors, we used the regression models in a Monte Carlo simulation to produce thousands of revenue forecasts for each of four forecast years corresponding to three potential phases of the project:

- Initial Operating Segment: Merced to San Fernando Valley: 2022;
- Bay-to-Basin: San Jose to San Fernando Valley, with a spur to Merced: 2027;
- Phase 1 Blended: San Francisco to Los Angeles, with a spur to Merced: 2029 and 2040

Ridership for each of the revenue forecasts was also estimated using the relationship between ridership and revenue.

This memo explains:

- The process for selecting the risk factors;
- The assumptions regarding the probability distributions of the risk factors;
- The fractional factorial design for running the full ridership and revenue model to obtain data points for development of the Risk Analysis Model regression equations;
- The estimated regression equations for each model year; and
- The forecast results from the Monte Carlo Simulation.

Selecting Risk Factors

HSR ridership and revenue forecasts are greatly impacted by assumptions. We started by compiling a comprehensive list of factors that could affect high-speed rail ridership and then selected the six factors we thought would have the greatest potential variability and/or influence on total high-speed rail ridership. Our rationale for selecting the factors is described below.

Potential Risk Factors

One of the most important considerations to remember about travel demand models is that they are based on cross-sectional snapshots of travel patterns and traveler behavior. Thus, an underlying assumption in travel demand models like the HSR Ridership and Revenue model is that travelers' responses to future travel options will be the same as they were when the snapshot was taken. While this may suffice for short-term forecasts, when considering long term forecasts this simple approach ignores the likelihood of significant changes in traveler behavior resulting from structural changes in society and the economy. Examples from the past of such changes include:

- Increasing workforce participation by women in the 60s, 70s and 80s of last century.
- Introduction of new work options such as telecommuting, shared jobs, and web-based conference calls (e.g. GoToMeeting™).
- Demographic shifts, such as the aging of the baby boomers, accompanied with increasing longevity and activity.
- Globalization of markets.
- Technological advances such as the internet, cell phones, and smart phones.

While we cannot know for certain what new changes will occur in society, we can speculate on what some of those changes might be and incorporate some of that speculation into our forecasts for the risk analysis.

There are two major risk categories:

- *Future Expectations Risks* – those that have to do with our expectations regarding the future (i.e. model inputs); and
- *Model Related Risks* – those that have to do with the inner workings of the model that reflect traveler behavior (i.e. model coefficients and constants)

While these two categories are not entirely independent, it is a useful way to think about the factors that influence potential long distance travel in California.

Appendix A of this memo has a long list of potential risk factors along with how we proposed addressing them in our analysis. As we considered more risks, we knew that the amount of computation time and analysis necessary to include those risks in the Monte Carlo simulation process would increase significantly. Therefore, we kept the risks to the handful that we judged would have the largest impacts on ridership and revenue. Some of the risks we believe to be so speculative or difficult to address that we did not include them but, rather, chose to address through discussion in our final report. There was also a middle ground of risks that we chose to address through sensitivity test model runs.

In all cases, we tried to select risk factors as independent of each other as possible to reduce the complications caused by correlation of the factors. A high level of correlation between the risk factors could lead to invalid results; many statistical software packages fail to execute when a high level of correlation between independent variables exists.

Selected Risk Factors

After compiling the comprehensive list of factors we thought could affect high-speed rail ridership, we narrowed the list to six factors that we thought would have the greatest impact on total high-speed rail ridership:

1. Total California population, households, and employment;
2. Spatial distribution of population and employment;
3. Auto operating cost;
4. Airline fares;
5. High speed rail main mode choice constants;
6. Trip frequency model constants;

Range of Risk Factor Values and Distributions

To conduct the risk analysis, each factor must be quantified so it can be treated as a continuous independent variable within a regression model represented as a distribution of values. The middle value often (but not always) has the greatest likelihood of occurring. The shape of the distribution can be triangular, normal, uniform, or another form. The shape of this distribution determines the likelihood of an independent variable's value under random sampling.

For each risk factor, we developed low, middle, and high values for each forecast year, and then developed a distribution around these values based on best available research and analysis (see Table 1). The distributions are described in more detail in the following sections.

Table 1 Risk Factor Values and Distributions

Risk Factor	Risk Factor Quantitative Value	Level[3]	Description	Regression Model Inputs				Distribution Description
				2022	2027	2029	2040	
Overall Population and Employment Growth	Ratio of future year households to observed year 2010 households	High	California Statewide Travel Demand Model Forecast - High household and employment growth rate	1.148	1.208	1.232	1.372	Correlated with Regional Spatial Distribution as shown in Table 4
		Mid	Mid-level household and employment growth rate	1.132	1.183	1.199	1.305	
		Low	Low household and employment growth rate	1.098	1.131	1.141	1.191	
Regional Spatial Distribution	Ratio of San Joaquin Valley population to rest of California	High	California Statewide Travel Demand Model Forecast - High growth rate in San Joaquin Valley	0.115	0.120	0.122	0.134	Correlated with Regional Spatial Distribution as shown in Table 3
		Mid	Mid-level growth rate in San Joaquin Valley	0.103	0.105	0.106	0.112	
		Low	Low growth rate in San Joaquin Valley	0.101	0.102	0.103	0.107	
Auto Operating Cost[1]	$/mile (2005$)	High	Based on high fuel forecasts and low fuel efficiency	$0.26	$0.24	$0.24	$0.24	Triangular, with Low set to 15 percent probability of occurrence and High at 85 percent probability of occurrence
		Mid	Reference/Base	$0.21	$0.20	$0.19	$0.20	
		Low	Based on low fuel forecasts and high fuel efficiency	$0.18	$0.17	$0.16	$0.15	
Airline Fares	Air fare skim factor	High	16 percent increase, as used in 2012 Business Plan airline competitive response scenario	1.16	1.16	1.16	1.16	Triangular, with Low set to 15 percent probability of occurrence and High at 85 percent probability of occurrence
		Mid	Base scenario, consistent with 2012 Business Plan runs	1.00	1.00	1.00	1.00	
		Low	9 percent reduction, as used in 2012 Business Plan airline competitive response scenario	0.91	0.91	0.91	0.91	
High Speed Rail Main Mode Choice Model Constants[2]	Change in HSR constant units from Base	High	Equivalent to 60 fewer minutes of IVTT for business/commercial (90 for recreation/other)	0.61	0.61	0.61	0.61	Normal distribution with Mean = 0 and Standard Deviation = 0.48
		Mid	Average of Offset Approach for CVR and Air Offset Method	0	0	0	0	
		Low	Equivalent to 60 more minutes of IVTT for business/commercial (90 for recreation/other)	-0.61	-0.61	-0.61	-0.61	
Trip Frequency Model Constants	Annual average roundtrips per capita	High	Increase from Mid scenario of 1.75 round trips per person	9.11	9.11	9.11	9.11	Truncated Normal distribution with Mean = 7.36 and Standard Deviation = 0.85
		Mid	Constants calibrated to CHTS trip rates that produce average of 7.36 round trips per person	7.36	7.36	7.36	7.36	
		Low	Decrease from Mid scenario of 1.75 round trips per person	5.61	5.61	5.61	5.61	

1 See memorandum, "*Revised forecasts of gasoline prices and fuel efficiency for use in 2014 Business Plan Model Runs and Forecasts*" dated September 30, 2013.
2 See memorandum, "*Version 2 Model High Speed Rail Alternative Specific Constants*" dated January 8, 2014.
3 High and low values for specific risk analysis experiments used to develop data points for risk analysis regressions – see Table 5.

Socio-economic Risk Factors (Overall Population and Employment Growth and Regional Spatial Distribution)

CS assembled county-level socioeconomic estimates and forecasts from many sources, including:

- California Statewide Travel Demand Model (CSTDM);
- U.S. Census Bureau;
- Moody's Analytics (Economy.com);
- Woods & Poole, Inc.;
- California Department of Finance (DOF);
- California Employment Development Department;
- California Economic Forecast Project (CEF);
- University of Southern California (Price School);
- UCLA (Anderson School);
- Center for Continuing Study of the California Economy; and
- MPOs: Metropolitan Transportation Commission (MTC); Sacramento Area Council of Governments (SACG); San Diego Association of Governments (SANDAG); Southern California Association of Governments (SCAG); and the San Joaquin Valley MPOs.

For most sources, we assembled and reviewed forecasts from multiple publication years beginning in the early 2000s (and as early as 1965 for one source). This history allowed us to assess each source's accuracy versus actual conditions over many years. Overall, we found that the U.S. Census Bureau's population and household projections were reasonably accurate. Other sources, mostly prepared by California-based organizations, tended to over-predict population, households and employment.

The CSTDM forecasts served as the starting point since they were recently updated to reflect adopted MPO forecasts (as of early summer 2013). The preponderance of information suggests that CSTDM forecast represents a likely high end of the future statewide socioeconomic growth. This forecast source assumes a statewide annual population growth rate of 1.01 percent, which is above the recent averages described in the prior paragraph and observed trends over the past several years. The CSTDM forecast also assumes an average population growth rate higher than the employment growth, which is counter to California's trends between World War II and the recent recession.

Beyond statewide trends, the CSTDM forecast incorporate very aggressive growth assumptions for the San Joaquin Valley[1]. These statewide and regional assumptions produce Valley-wide forecasts that are 10 percent to 20 percent larger than any other source. The CSTDM forecasts

[1] For this analysis, San Joaquin Valley includes San Joaquin, Stanislaus, Merced, Madera, Fresno, Tulare, Kings, and Kern counties.

are also at odds with recent growth trends and state growth policies that aim to reduce greenhouse gas emissions by directing new socioeconomic growth into currently developed areas.

Based on these analysis results, we incorporated two components of socioeconomic growth in the risk analysis and then combined them in a matrix of distributions.

- Statewide population, household, and employment forecasts (shown in Table 2 for each decade and the travel model years); and,
- Share of California population in San Joaquin Valley counties (Table 3).
  - Distribution 1 follows the CSTDM forecasts.
  - Distribution 2 follows the Valley-wide average distribution from recent statewide forecasts, with excess population, employment, and household-related employment shifted to the Bay Area, the Sacramento region, and Southern California.
  - Distribution 3 reflects a further shifting of population, household, and employment growth from the San Joaquin Valley to all other California regions. It assumes that the San Joaquin Valley will see 2010 to 2050 growth patterns that are closer to statewide averages (for population and households) and long-term historical patterns for jobs.

Table 2 Statewide Socioeconomic Forecasts for Ridership and Revenue Risk Analysis Model (in millions)

	High Range Forecast			Mid Range Forecasts			Low Range Forecast		
Year	Population	House-holds	Employ-ment	Population	House-holds	Employ-ment	Population	House-holds	Employ-ment
2010	37.309	12.587	16.052	37.309	12.607	16.078	37.309	12.606	16.078
2020	41.578	14.177	18.677	40.790	13.909	18.683	39.756	13.515	17.859
2022	42.436	14.454	19.018	41.889	14.268	18.773	40.583	13.839	18.188
2027	44.626	15.206	19.870	43.761	14.911	19.485	41.829	14.257	18.624
2029	45.503	15.506	20.211	44.359	15.116	19.703	42.218	14.386	18.752
2034	47.693	16.258	21.063	45.506	15.512	20.097	42.742	14.549	18.876
2040	50.357	17.272	22.198	47.951	16.447	21.138	44.111	15.016	19.445
2050	54.869	18.761	24.128	51.106	17.474	22.473	46.762	15.989	20.563

Note: Ridership and revenue model forecast years are indicated by bold font in the "year" column.

Table 3 Share of Statewide Socioeconomics in San Joaquin Valley Counties

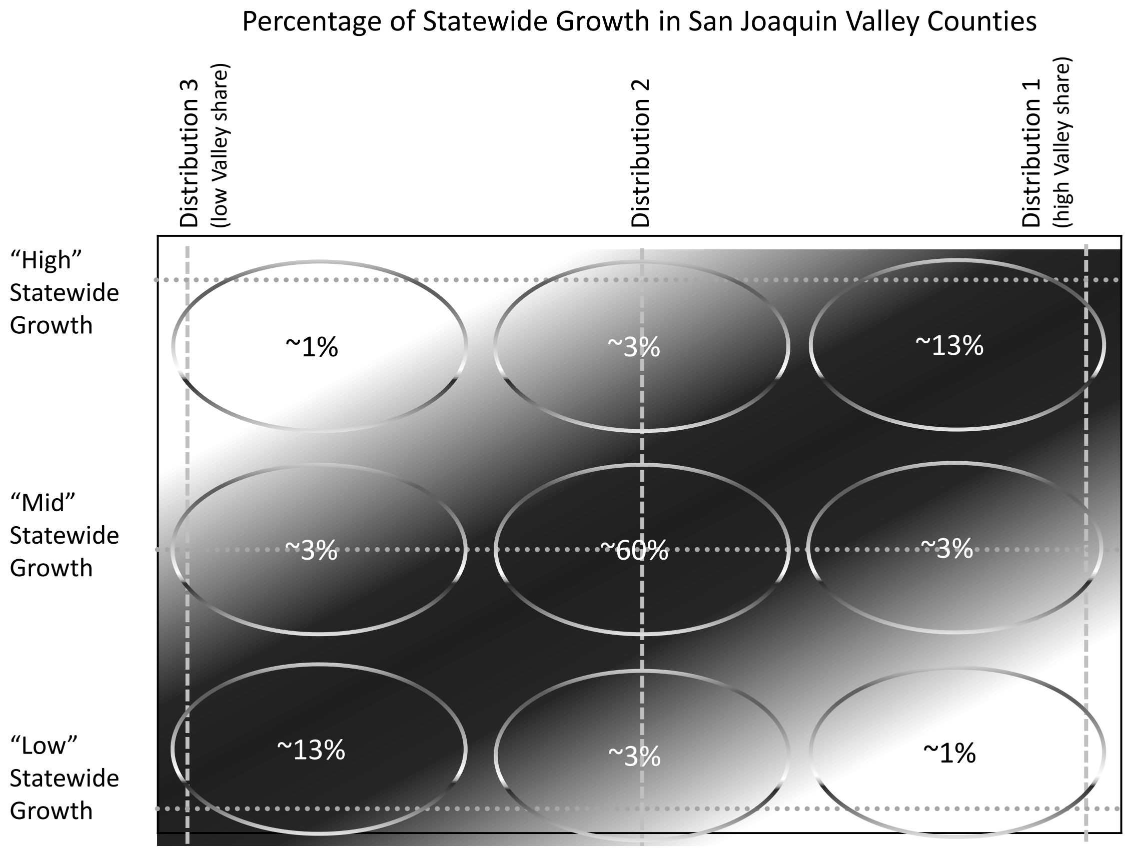
	Distribution 1 (CSTDM)			Distribution 2			Distribution 3		
Year	Population	House-holds	Employ-ment	Population	House-holds	Employ-ment	Population	House-holds	Employ-ment
2010	10.66%	9.66%	9.33%	10.66%	9.66%	9.33%	10.66%	9.66%	9.33%
2020	12.30%	11.34%	10.02%	11.11%	10.23%	9.57%	10.95%	10.08%	9.12%
2022	12.52%	11.53%	10.21%	11.20%	10.31%	9.62%	11.00%	10.13%	9.17%
2027	13.06%	12.00%	10.69%	11.42%	10.49%	9.96%	11.15%	10.24%	9.31%
2029	13.27%	12.19%	10.88%	11.51%	10.57%	10.09%	11.20%	10.29%	9.37%
2034	13.81%	12.66%	11.36%	11.73%	10.75%	10.43%	11.35%	10.40%	9.53%
2040	14.37%	13.38%	12.13%	12.00%	11.17%	11.07%	11.52%	10.72%	9.65%
2050	16.10%	15.24%	13.31%	12.45%	11.78%	11.67%	11.80%	11.18%	9.87%

Note: Ridership and revenue model forecast years are indicated by bold font in the "year" column.

We combined the information shown in Tables 2 and 3 and assigned probabilities to the outcomes (Table 4) based on the following rationale:

- We assigned the highest probability to the middle combination of mid-level statewide growth and "Distribution 2." This combination shows more modest statewide and San Joaquin Valley growth at rates that are consistent with the more recently-published third-party sources and are in line with historical trends.
- The highest probability is along the top left to bottom right diagonal because of how total growth and distribution match up. This diagonal reflects a general principal found in all of the third-party sources – namely, any departure from "average" statewide socioeconomic growth will depend on the fortunes of the San Joaquin Valley. If the growth levels assumed in the CSTDM were to occur, it is likely that the distribution associated with the CSTDM -- lots of growth in the San Joaquin Valley -- would occur. Similarly, lower statewide growth levels would more likely occur along with a distribution that has less relative growth in the San Joaquin Valley.
- The probability of a statewide growth and regional distribution combination decreases somewhat rapidly as we depart from the diagonal line mentioned above.

Table 4 Likelihood of Statewide and San Joaquin Valley Socioeconomic Growth Combinations



Auto Operating Cost

CS updated the range of gasoline prices and fuel efficiency forecasts in California with the latest U.S. Energy Information Administration (EIA) projections. This update is documented in our September 30th memo on auto operating costs.[2] The memo described low, mid, and high estimates of forecast year auto operating costs which were used as the respective values for the auto operating cost risk factor. The low and high values for auto operating costs were set as the 15th and 85th percentiles, respectively, in a triangular distribution. This means that 30 percent of the scenarios in the Monte Carlo simulation were likely to have values lower or higher than these levels—15 percent of the observations on either side. So, for 2022, the range of values used in the risk analysis was actually broader than $0.18 to $0.26 per mile values specified as the "low" and "high" values for auto operating cost. The highest probability of occurrence was at the mid value of $0.21/mile. Similar assumptions were made for other forecast years with the low and high values for the 15th and 85th percentiles, and the mid values as specified in Table 1.

Figure 1 shows the shape of the auto operating cost distribution for 2022. The x-axis shows the auto operating cost ($/mile) in 2005 dollars. The curve (i.e. triangle) shows the probability of occurrence for a specific auto operating cost value in the Monte Carlo simulation. Figure 2

[2] *Revised forecasts of gasoline prices and fuel efficiency for use in 2014 Business Plan Model Runs and Forecasts*

shows a histogram of the actual number of simulations by auto operating cost for 2022 for the 5,000 simulations performed juxtaposed on the specified curve (note that the other input variables also varied for the 5,000 simulations). As specified, 750 of the simulations, or 15 percent, had auto operating costs less than $0.18 per mile and 4,250 of the simulations, or 85 percent, had auto operating costs less than $0.26 per mile. The greatest numbers of simulations were performed with values around the mid value of $0.21 per mile. Since the triangular distribution was skewed to the right, only about 42.4 percent of the simulations had auto operating costs less than the mid value of $0.21 per mile.

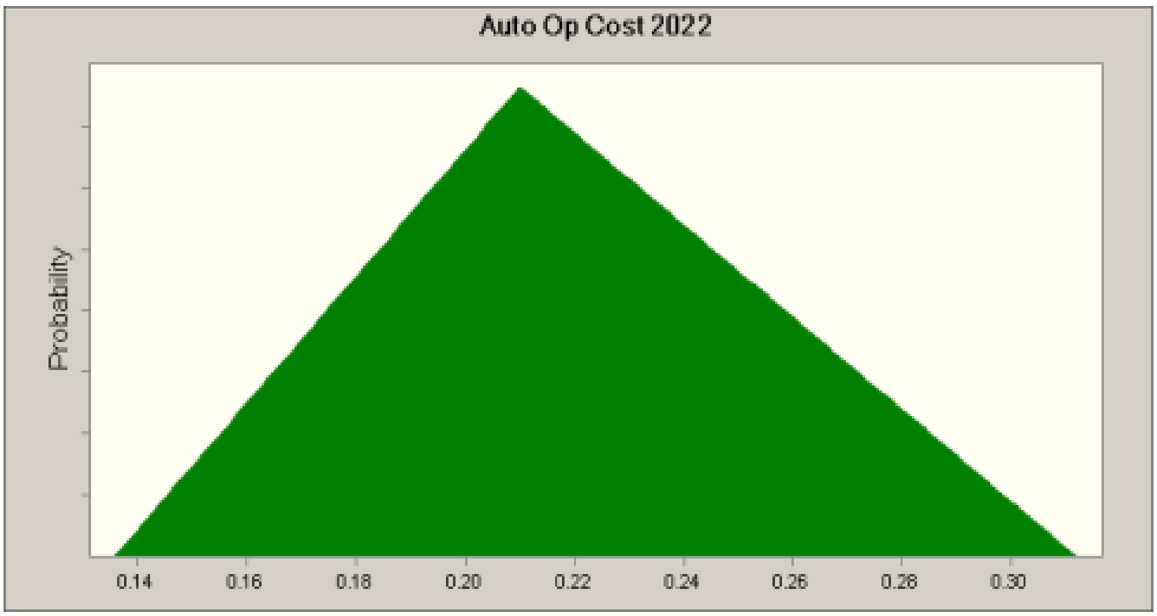


Figure 1 Year 2022 Auto Operating Cost Distribution

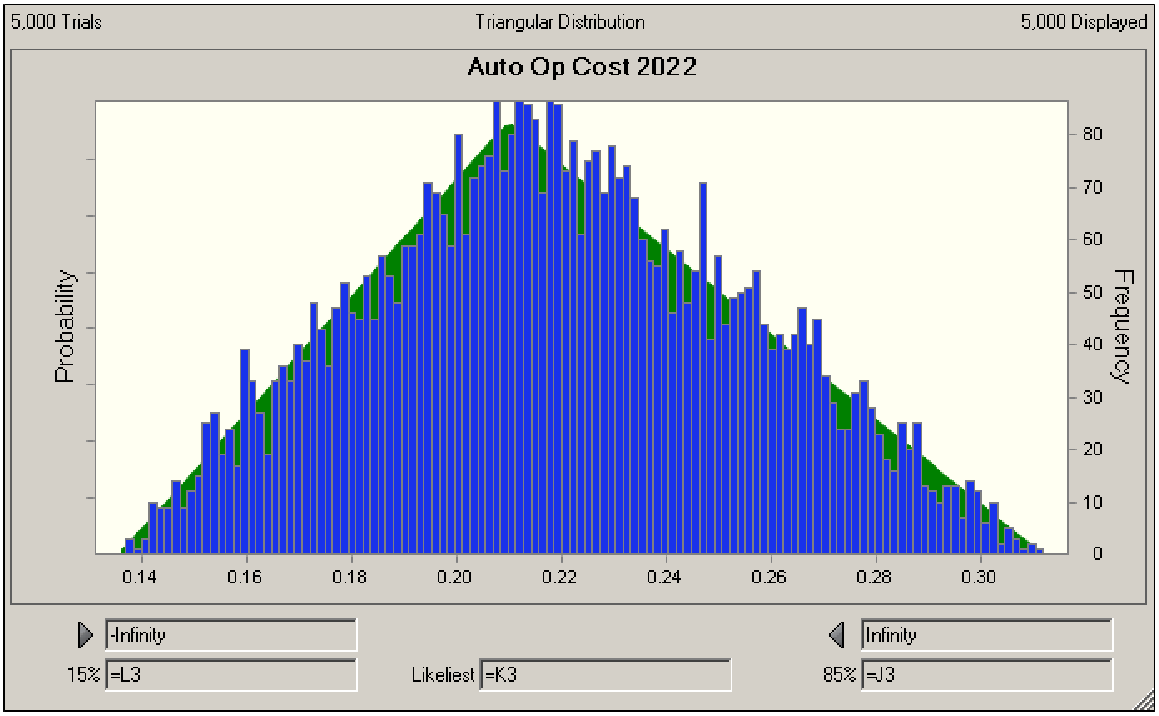


Figure 2 Year 2022 Auto Operating Cost Distribution – From Monte Carlo Simulation

Air Fares

Forecast year mid-level air fares remain consistent with the 2012 Business Plan, which were developed in 2011 by Cambridge Systematics and Aviation System Consulting. As part of sensitivity analysis for the 2012 Business Plan ridership and revenue forecasting, Cambridge Systematics, in partnership with Aviation System Consulting, developed airline competitive response scenarios. The low-fare scenario specified a 9 percent reduction in real fares from 2009 levels and the high-fare scenario increased real fares over 2009 levels by an average of 16 percent across all markets. The reasoning for the choice of these alternative scenarios is detailed in 2012 Business Plan Ridership and Revenue Forecasting Tech Memo.

We did not perform a new detailed analysis of the future of California airlines' reaction to the introduction of HSR; rather, we retained the 2012 Business Plan assumptions for the airline level-of-service and potential future variations in fares. A brief discussion with Aviation System Consulting confirmed that the analysis they performed in 2011 is still generally relevant and no significant changes had occurred since then.

It should be noted that the fares in the 2012 Business Plan high-fare scenario differed by market for an average of a 16 percent increase. However, varying the fares by air market segment would have significantly increased the effort needed to produce each full Monte Carlo simulation. Thus, for the 2014 Business Plan, the 16 percent increase in fares was uniformly assumed for all air markets.

A 9 percent reduction in fares was set as the low value at the 15th percentile, and a 16 percent increase in fares is set as the high value at the 85th percentile. Unlike the assumptions regarding auto operating cost, the same air fare distribution was applied for all forecast years. Figure 3 juxtaposes the histogram of simulations for 2022 the specified distribution. Histograms similar to that shown for 2022 could be produced for the other forecast years. For 2022, a total of 757 of the simulations, or 15.14 percent, had air fare factors less than 0.91 and 426 of the simulations, or 85.34 percent, had air fare factors less than 1.16. The greatest numbers of simulations were performed with values around the mid value of 1.0. Since the triangular distribution was skewed to the right, only about 41.7 percent of the simulations had air fare factors less than the mid value of 1.0.

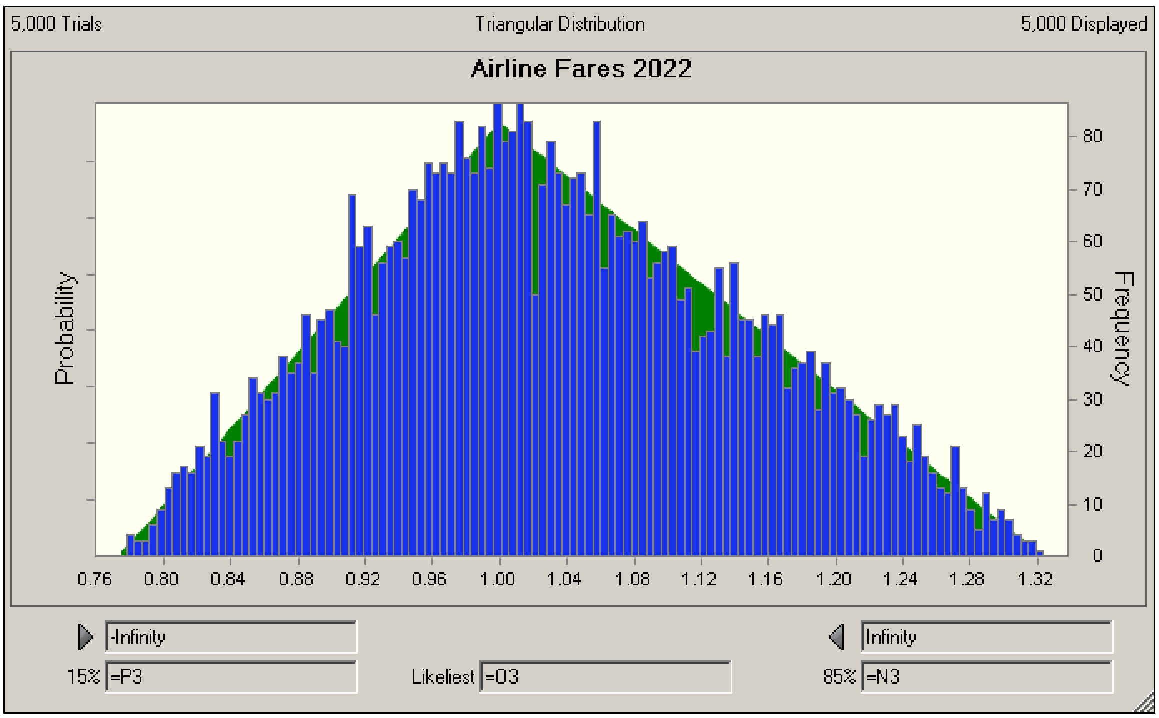


Figure 3 Year 2022 Distribution of Factors Applied to Air Fare Mid-level Matrix

High-Speed Rail Main Mode Choice Constants

An important part of any mode choice model is a modal constant that explains factors that are not quantifiable by the stated and revealed preference (RP) surveys. When dealing with existing modes such as auto, conventional rail (CVR), and air, we can calibrate this constant by comparing the model outcomes to observed behavior. With a new mode like HSR, this is impossible, so there is uncertainty in the specified constant.

Uncertainty in the HSR constants comes from the distributional assumptions of the model itself and the data used to estimate the model. The former is relatively straightforward, in that the logit model may not be an accurate representation of how individuals actually make mode choices. The latter refers to the uncertainties associated with how the stated preference (SP) data were collected, the survey instrument, respondents perceptions based on "public opinion" at time of the survey, and other related issues. This uncertainty is driven by the following:

1. HSR currently does not exist in California, and thus we are unable to calibrate the HSR constant to observed mode shares.

2. HSR does not exist in the United States. Americans have very little experience with HSR, so we can not use oberved data or experiences from other parts of the country to guide our knowledge in assessing Californian's willingness to use HSR. In addition, while we have gained some insights from SP surveys on the attractiveness of HSR between destinations within California, these results have a degree of error and uncertainty due to the lack of actual experience on HSR. In many travel-related SP surveys, individuals are asked to assess a new mode with which they are familiar, such as a new bus, toll road, or urban rail system, even if it does not provide service for the travel being considered.

3. Uncertainty exists in the HSR system itself. The HSR constant captures all unobserved attributes and variables that affect an individual's decision to use HSR that are not captured by other variables within the model. This includes wait and terminal times, the existence or non-existence of security checkpoints, attractiveness of the HSR stations, and amenities on trains such as food options, wireless internet, etc.

4. Uncertainty exists in the mode choice model and the methodology used to calculate the HSR constant. Inherent uncertainty exists in all parts of model estimation including, but not limted to, the estimated variables within the mode choice model, sampling error associated with the data summarized from the SP survey, sampling error with the observed data collected for calibration of the existing model, and the method used to specify the mid-level HSR constants.

Mid-level HSR constants were specified based on the relationships of the air, CVR, and HSR constants estimated using SP data, and the air and CVR constants after calibration to match observed 2010 travel[3]. The normal distribution was chosen to represent the uncertainty in the HSR constants. Further, to avoid overcomplicating the risk analysis model, the distribution of the HSR constant was not varied by trip purpose. The risk factor used in the risk analysis regression equation was the HSR constant unit change from the mid-level (specified) HSR constant. The mid-level risk factor value for the HSR constant is set to 0.0 (i.e. 0.0 change from the specified constant). A 0.1 unit change in the risk factor value would correspond to a 0.1 unit increase in the HSR constant for each purpose. Note that an increase in the constant means an increase in the desirability of the mode.

To develop the variance for the HSR constant distribution, we started by considering a value for the absolute minimum HSR constant. Since there was no apparent reason that any of the unobserved characteristics for the HSR mode should be any worse than those for the CVR mode, we thought the CVR constant should represent this minimum value for the HSR constant. As mentioned above, a single distribution was applied for all trip purposes due to the

[3] See memorandum, "*Version 2 Model High Speed Rail Alternative Specific Constants*".

constraints in our application of the risk analysis procedure. Because recreation/other was, by far, the most prevalent long-distance trip purpose (about 75 percent of all long distance trips), we focused on the relationship between the recreation/other mode choice model CVR and HSR constants. The CVR constant was -1.25 units lower than the HSR constant for the recreation/other trip purposes; thus, -1.25 was selected as the lower bound for the unit offset for the distribution. Since the normal distribution was used for the risk analysis, we chose the 0.5th percentile value of the distribution to correspond to the offset value of -1.25. Thus, 0.5 percent of the time (1 in 200), the HSR constant used in the risk analysis for recreation/other would be less than the CVR constant.

The above led to the specification that the deviation in the HSR constant used in the risk analysis would follow a normal distribution with mean zero and standard deviation 0.48. Figure 4 shows the distribution of HSR constant offsets used for 2022. 49.66 percent of the simulations used a HSR constant offset less than 0.0; 24 of the 5,000 simulations (0.48 percent) used offsets less than -1.25 (i.e. the net effect of producing the CVR constant as noted above) and 29 of the simulations (0.58 percent) used offsets more than 1.25.

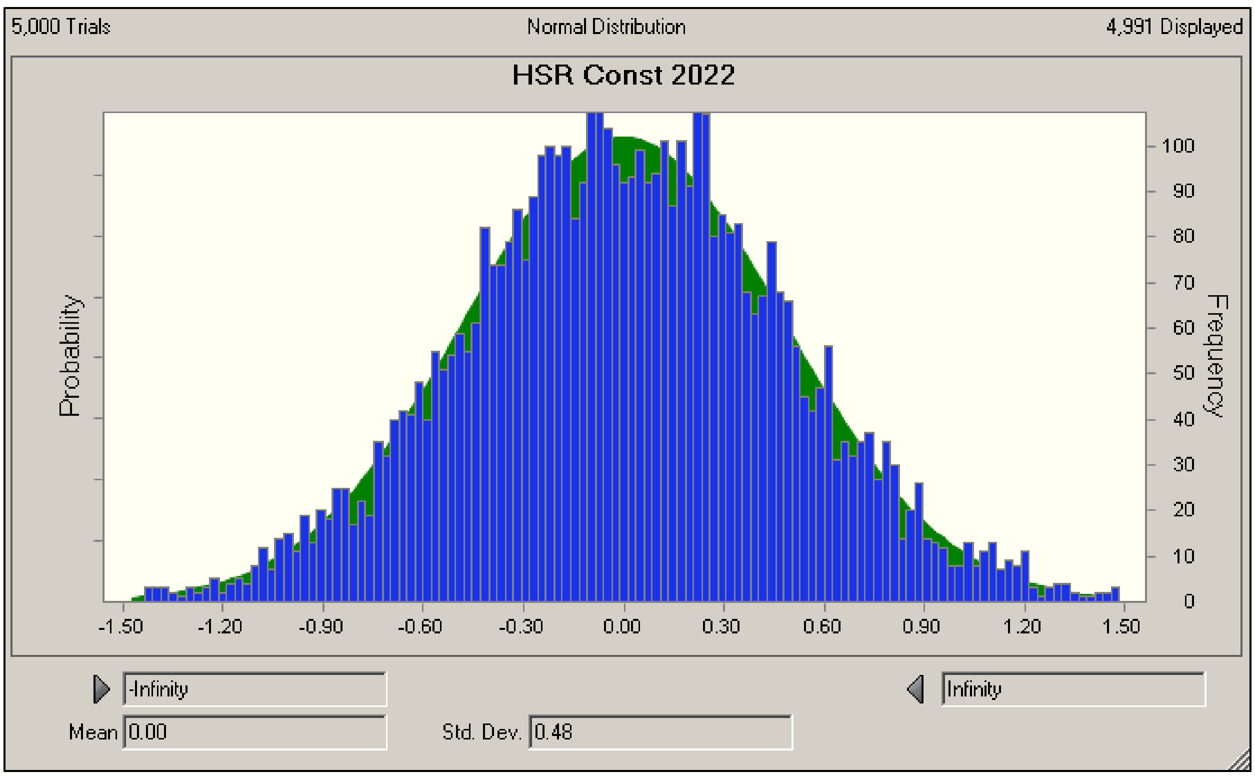


Figure 4 2022 Distribution of HSR Constant Units from Mid Scenario

Trip Frequency Model Constants

Similar to the uncertainty found in the HSR constants, uncertainty also exists in the constants calibrated for the trip frequency model. The trip frequency model estimates the total number long-distance trips (greater than or equal to a straight-line distance of 50 miles from the trip maker's home) made per person per day. The data used for the trip frequency model estimation was from the long-distance travel portion of the 2012-2013 California Household Travel Survey (CHTS). The data used for calibration was based on 2012-2013 CHTS data weighted (expanded) to match 2010 California population characteristics.[4]

Based on the weighted 2012-2013 CHTS data, California residents made an annual average of 8.2 intra-California long distance trips (50 miles or more) per person in 2010. For long distance trips over 100 miles in length, the overall average annual trips per capita estimated using the weighted 2012-2013 CHTS was close to the midpoint of national data collected in the 1995 American Travel Survey and the 2001 National Household Travel Survey. Thus, we were confident that the HSR ridership and revenue trip frequency model calibrated to match 2010 trip making estimated using the 2012-2013 CHTS data should be used to set the midpoint trip rates for the risk analysis.

The annual intra-California long distance trips per person estimated using data from the 2011 Harris panel long distance survey performed for the CAHSRA was 6.0 trips per person per year. We believed there was a very high probability that the true number of annual trips per person per year was above the reported Harris Survey number. Thus, we considered 2.2 annual trips below the annual average long distance trips per person forecast using the calibrated trip frequency model as the lower bound in the distribution.

The calibrated trip frequency model constants resulted in averages of 7.36 annual long-distance trips per person for each of the forecast years[5]. These values represented the mid-level values in the distribution for the risk analysis. We used a normal distribution, with a standard deviation of 0.85. This resulted in 2.2 annual trips less per person than the mid-level value to fall at the 0.5th percentile (i.e. 0.5 percent of the annual trips per person simulations were less than 5.16). Figure 5 shows the distribution for 2022.

[4] *Version 2.0 Model – Processing of California Household Travel Survey Data for Model Calibration and Validation*, September 2013.

[5] The difference in annual long-distance trips from the weighted CHTS is, in part, due to the elimination of long-distance bus trips from the dataset.

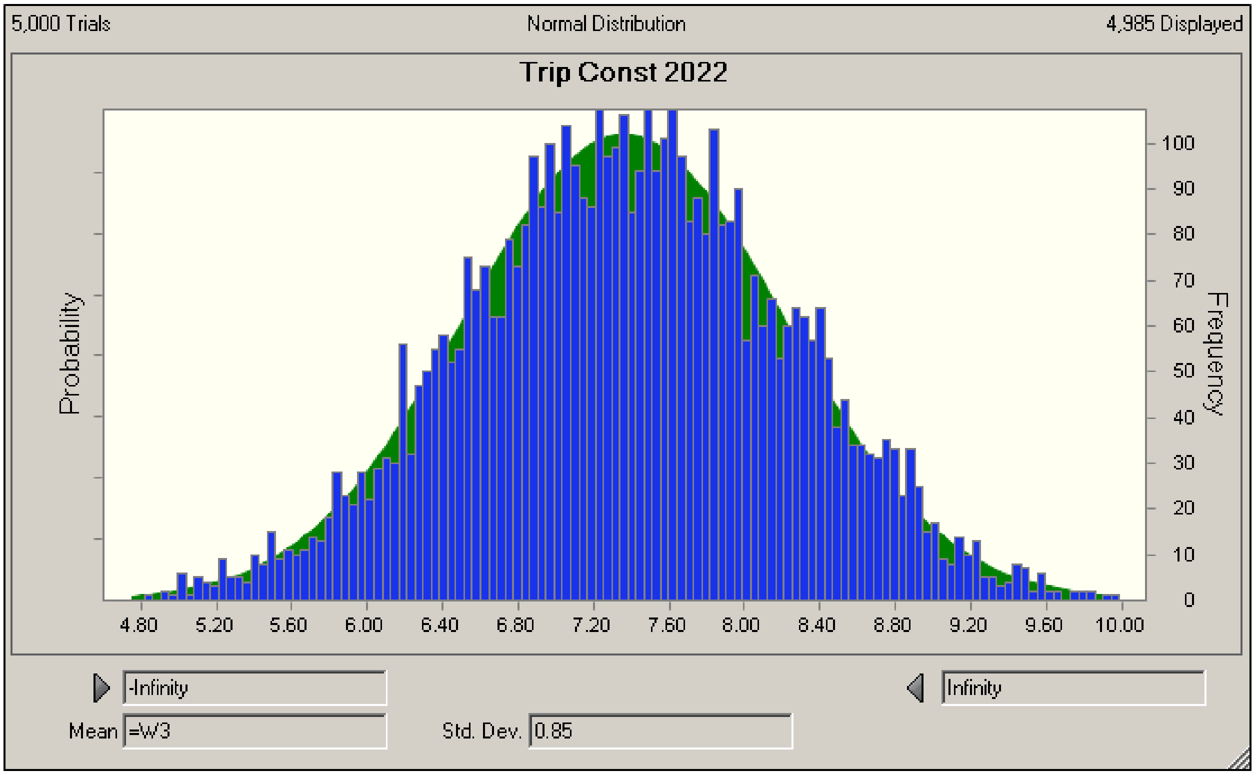


Figure 5 Distribution of Average Annual Roundtrips/person

The Risk Analysis Model

Ridership and Revenue Version 2.0 Model Runs

Once the risk factors and their distributions were defined, the full ridership and revenue model was run to obtain input into the Risk Analysis regression equations. We began by running a "mid-level" model run with all six factors set at the mid value (see Experiment Number 1 in Table 5) for each forecast year. To limit the number of model runs to a reasonable level we pursued a fractional 2-level factorial design for running the full model using high and low values for input variables as shown in Table 1. Thirty-two runs (Experiment Numbers 2-33 in Table 5) were used to estimate all the main effects and two-factor interactions resulting from varying the input data. This design, which was only one-half of the 64 runs required for a full 2-level factorial design for 6 factors, saved run time but could not be used to estimate interactions between three or more factors. However, we do not have reason to believe there would be large high-order interactions among the risk factors we selected. Additional runs with data points between the mid level and low level, and between the mid-level and high-level values of each risk factor distribution were added to provide information regarding the non-linearity of the forecast distributions and to ensure that the regression models represented the middle values within the distributions, and not just the extremes (Experiment Numbers 34-47 in Table 5). These additional runs were important since the regression models, discussed in the next section, were exponential rather than linear.

Risk Analysis Regression Models for Ridership and Revenue

Ridership vs. Revenue

We began the analysis by testing the relationship between ridership and revenue resulting from the Version 2 Model runs. Revenue and ridership were closely correlated with a R^2 of more than 0.999 for each year. The relationship between ridership and revenue for each forecast year was as follows:

*Year 2022 Revenue = 55.147 * Ridership*

*Year 2027 Revenue = 55.401 * Ridership*

*Year 2029 Revenue = 47.467 * Ridership*

*Year 2040 Revenue = 47.049 * Ridership*

For all the 47 runs in each model year, the predicted revenues from the above equations were compared with the actual revenues, and the results show the differences between predicted revenue based on ridership versus actual revenue was between -9 percent and 5 percent.

Since revenue and ridership were highly correlated, we developed regression equations for revenue only and used the above relationships between revenue and ridership to calculate the corresponding ridership forecasts for the risk analysis.

Table 5 Ridership and Revenue Version 2.0 Model Run Experiments for each Forecast Year[1]

Experiment Number	Overall Growth	Regional Spatial Distribution	Auto operating cost	Airline fares	HSR Mode Choice Constant	Trip Frequency Constant
1	Mid	Mid	Mid	Mid	Mid	Mid
2	Low	Low	Low	Low	Low	Low
3	Low	Low	Low	Low	High	High
4	Low	Low	Low	High	Low	High
5	Low	Low	Low	High	High	Low
6	Low	Low	High	Low	Low	High
7	Low	Low	High	Low	High	Low
8	Low	Low	High	High	Low	Low
9	Low	Low	High	High	High	High
10	Low	High	Low	Low	Low	High
11	Low	High	Low	Low	High	Low
12	Low	High	Low	High	Low	Low
13	Low	High	Low	High	High	High
14	Low	High	High	Low	Low	Low
15	Low	High	High	Low	High	High
16	Low	High	High	High	Low	High
17	Low	High	High	High	High	Low
18	High	Low	Low	Low	Low	High
19	High	Low	Low	Low	High	Low
20	High	Low	Low	High	Low	Low
21	High	Low	Low	High	High	High
22	High	Low	High	Low	Low	Low
23	High	Low	High	Low	High	High
24	High	Low	High	High	Low	High
25	High	Low	High	High	High	Low
26	High	High	Low	Low	Low	Low
27	High	High	Low	Low	High	High
28	High	High	Low	High	Low	High
29	High	High	Low	High	High	Low
30	High	High	High	Low	Low	High
31	High	High	High	Low	High	Low
32	High	High	High	High	Low	Low
33	High	High	High	High	High	High
34	High	High	Mid	Mid	Mid	Mid
35	Mid	High	Mid	Mid	Mid	Mid
36	Mid	Mid	Mid	MidHigh	MidLow	MidLow
37	Mid	Mid	MidHigh	Mid	MidLow	MidLow
38	Mid	Mid	Mid	MidLow	MidLow	MidHigh
39	Mid	Mid	MidHigh	Mid	MidLow	MidHigh
40	Mid	Mid	MidLow	Mid	MidHigh	MidHigh
41	Mid	Mid	Mid	Mid	MidHigh	MidLow
42	Mid	Mid	MidLow	MidLow	MidHigh	MidHigh
43	Mid	Mid	Mid	MidHigh	MidHigh	MidLow
44	Mid	Mid	Mid	Mid	Low	Mid
45	Mid	Mid	Mid	Mid	MidLow	Mid
46	Mid	Mid	Mid	Mid	MidHigh	Mid
47	Mid	Mid	Mid	Mid	High	Mid

[1] Refer to Table 1 for high and low values used.

Revenue Regression Models

Using the results from the ridership and revenue forecasts from each of the 47 full model runs, we estimated relationships between the revenue forecasts and the input risk factor levels. The Monte Carlo method, described in the next section, made it feasible to quickly produce the thousands of revenue forecasts based on varying levels of the input risk factor variables that were necessary to estimate probabilities of specific outcomes. The revenue forecasts produced using the Monte Carlo method were predicated on deterministic equations (in our case, the regression models). Therefore, special attention was given to the construction of the deterministic equations. We analyzed both linear and non-linear transformations of model variables, and found that exponential relationship between revenue and risk factors resulted in the best model fits, with all forecast years having R^2 above 0.99. The differences between predicted revenues and estimated revenues from the full model runs was between ±5 percent. For each of the forecast years, the regression models had the following functional form:

*Revenue = exp(Intercept + a * Overall Growth + b * Regional Spatial Distribution + c * Auto operating cost + d * Airline fares + e * HSR Mode Choice Constant + f * Trip Frequency Constant)*

The coefficients and related statistical measures for each forecast year are shown in Tables 6 through 9. The standardized estimates show the estimated changes in revenue (in standard deviation units) when the specified input variable is increased by one standard deviation. For all years, the HSR mode choice constant has the highest standardized estimate, followed by the annual round trips per person and the auto operating cost.

Table 6: Regression Equation for Year 2022 Initial Operating Segment

	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	16.961	0.186	91.25	<.0001	0.000
Growth in Households	1.411	0.151	9.32	<.0001	0.059
Regional Spatial Distribution	2.491	0.506	4.93	<.0001	0.031
Auto Operating Cost	1.569	0.096	16.31	<.0001	0.102
Airline Fares	0.085	0.031	2.76	0.0088	0.017
HSR Mode Choice Constant	0.895	0.006	145.27	<.0001	0.912
Annual Round Trips/Person	0.137	0.002	62.88	<.0001	0.395
Adjusted R-square	0.998				

Table 7: Regression Equation for Year 2027 Bay-to-Basin

	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	17.580	0.152	115.75	<.0001	0.000
Growth in Households	1.343	0.111	12.04	<.0001	0.091
Regional Spatial Distribution	1.461	0.450	3.25	0.0024	0.024
Auto Operating Cost	1.692	0.109	15.59	<.0001	0.117
Airline Fares	0.098	0.034	2.87	0.0065	0.021
HSR Mode Choice Constant	0.827	0.007	120.50	<.0001	0.898
Annual Round Trips/Person	0.137	0.002	56.46	<.0001	0.421
Adjusted R-square	0.997				

Table 8: Regression Equation for Year 2029 Phase 1 Blended

	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	17.961	0.135	133.54	<.0001	0.000
Growth in Households	1.302	0.095	13.68	<.0001	0.107
Regional Spatial Distribution	0.876	0.413	2.12	0.0400	0.016
Auto Operating Cost	1.631	0.103	15.87	<.0001	0.124
Airline Fares	0.093	0.034	2.74	0.0091	0.021
HSR Mode Choice Constant	0.791	0.007	116.32	<.0001	0.891
Annual Round Trips/Person	0.136	0.002	56.56	<.0001	0.433
Adjusted R-square	0.997				

Table 9 Regression Equation for Year 2040 Phase 1 Blended

	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	18.010	0.090	200.32	<.0001	0.000
Growth in Households	1.232	0.052	23.78	<.0001	0.198
Regional Spatial Distribution	1.022	0.328	3.12	0.0034	0.026
Auto Operating Cost	1.767	0.106	16.71	<.0001	0.140
Airline Fares	0.106	0.039	2.74	0.0092	0.023
HSR Mode Choice Constant	0.785	0.008	103.79	<.0001	0.869
Annual Round Trips/Person	0.136	0.003	51.06	<.0001	0.426
Adjusted R-square	0.997				

Monte Carlo Simulation

Crystal Ball add-on software to Excel provided us the capability to run a randomized series of scenarios (Monte Carlo simulation). We defined the scenarios by varying the six risk factor values throughout their associated distributions for each forecast year. The revenue regression equation for each forecast year was used to estimate the associated revenue for each scenario and the relationship between revenue and ridership for each forecast year was used to estimate the ridership. Crystal Ball was used to automate the simulation process by selecting combinations of input values for the risk factors that were used to construct individual scenarios Crystal Ball automatically calculated and recorded the results of thousands of runs for the randomly selected input values. For each 2014 Business Plan forecast year, we ran a series of 5,000 Monte Carlo simulations using Crystal Ball to obtain revenue probability distributions. The results are presented in the next section.

Summary of Ridership and Revenue Forecasts

Range of Ridership and Revenue Forecasts

In Table 10, the ranges of forecast ridership and revenue along with the associated probabilities of achieving those levels are shown for each forecast year. Forecasts for confidence levels from 5 percent to 95 percent are shown. For example, the 15 percent confidence level means that there is a 15 percent chance that the revenue will be lower than the forecast value (or, conversely, an 85 percent chance that it will be higher).

The HSRA should choose which confidence level best fits its needs for the business decisions that it needs to make. The magnitude of the forecast range for both ridership and revenue increases as the forecast year extends farther out in time, as expected. The range in revenue for Year 2022 between the 5th and 95th percentiles is $1,041 million compared to $2,258 million in Year 2040.

Table 10 Range of Ridership and Revenue (2013$) by Forecast Year (millions)

Likelihood that Ridership/ Revenue will be less than Stated Value	Year 2022		Year 2027		Year 2029		Year 2040	
	Ridership	Revenue	Ridership	Revenue	Ridership	Revenue	Ridership	Revenue
5%	5.4	295.5	9.5	528.6	14.6	695.0	17.1	805.6
15%	7.2	399.4	12.4	689.6	19.0	901.7	21.9	1,030.6
25%	8.5	469.3	14.6	806.3	22.0	1,045.0	25.4	1,195.0
50%	11.4	626.5	19.0	1,053.4	29.1	1,379.0	33.3	1,567.9
75%	15.4	851.1	25.1	1,389.0	38.2	1,811.9	43.5	2,045.7
85%	18.2	1,002.9	29.5	1,632.2	43.7	2,074.6	49.9	2,349.8
95%	24.2	1,336.7	38.3	2,120.2	56.0	2,656.9	65.1	3,064.0

Figures 6 through 9 graphically display the cumulative probabilities of achieving specified revenue levels for the various forecast years. The distributions are skewed to the right, indicating that the values where there is 99 percent confidence that revenue will be lower than the specified values are further away from the median (or 50th percentile) than the revenues for the 1 percent confidence level. This is a result of the right skewed risk factor input distributions for auto operating cost and airfare (see Figure 1 and 3).

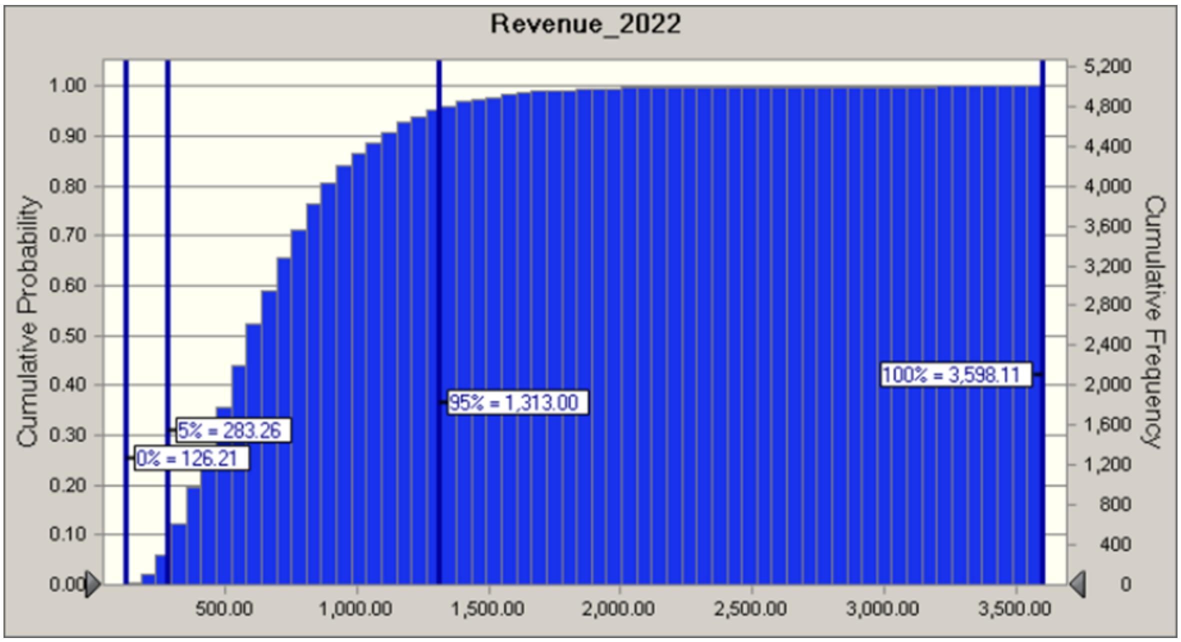


Figure 6 Range of Revenue for Year 2022 Initial Operating Segment HSR System

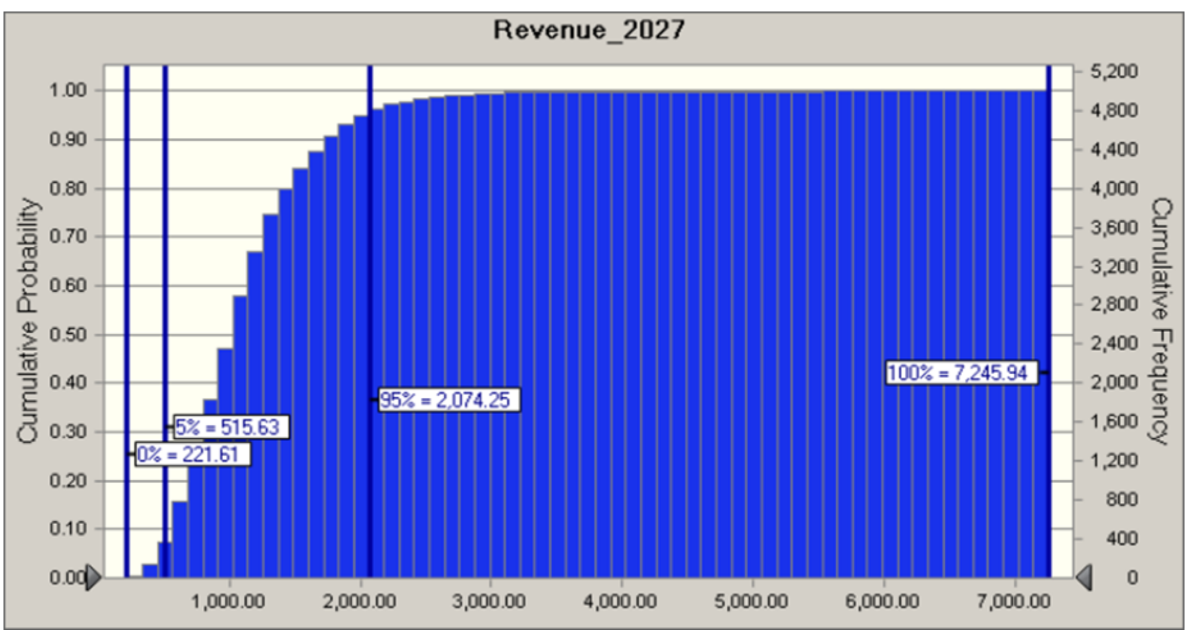


Figure 7 Range of Revenue for Year 2027 Bay-to-Basin HSR System

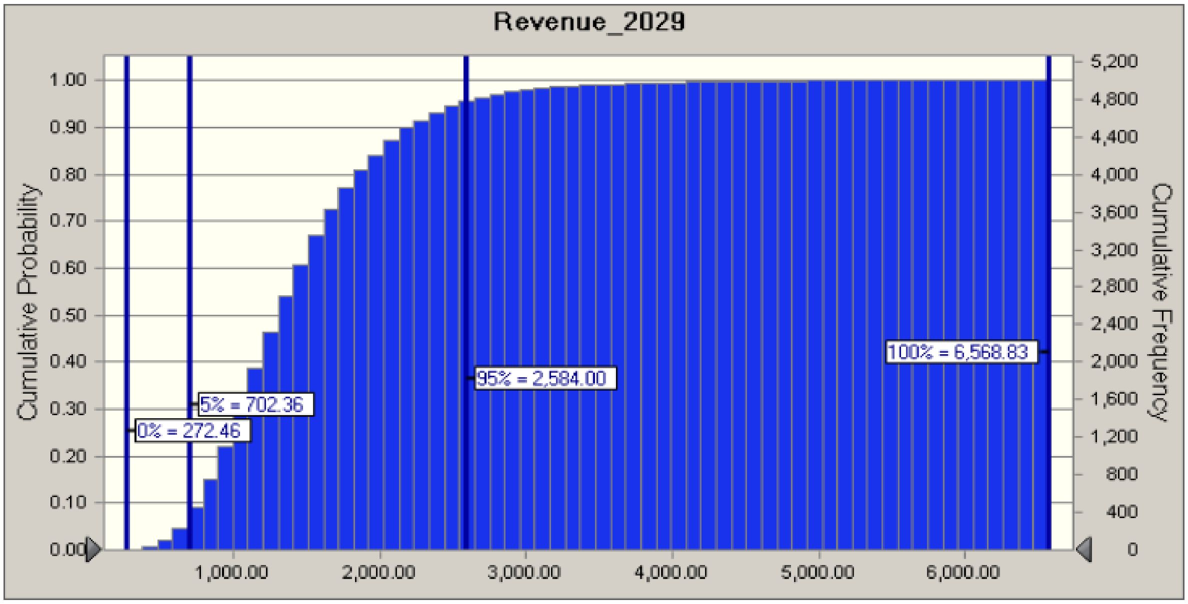


Figure 8 Range of Revenue for Year 2029 Phase 1 Blended HSR System

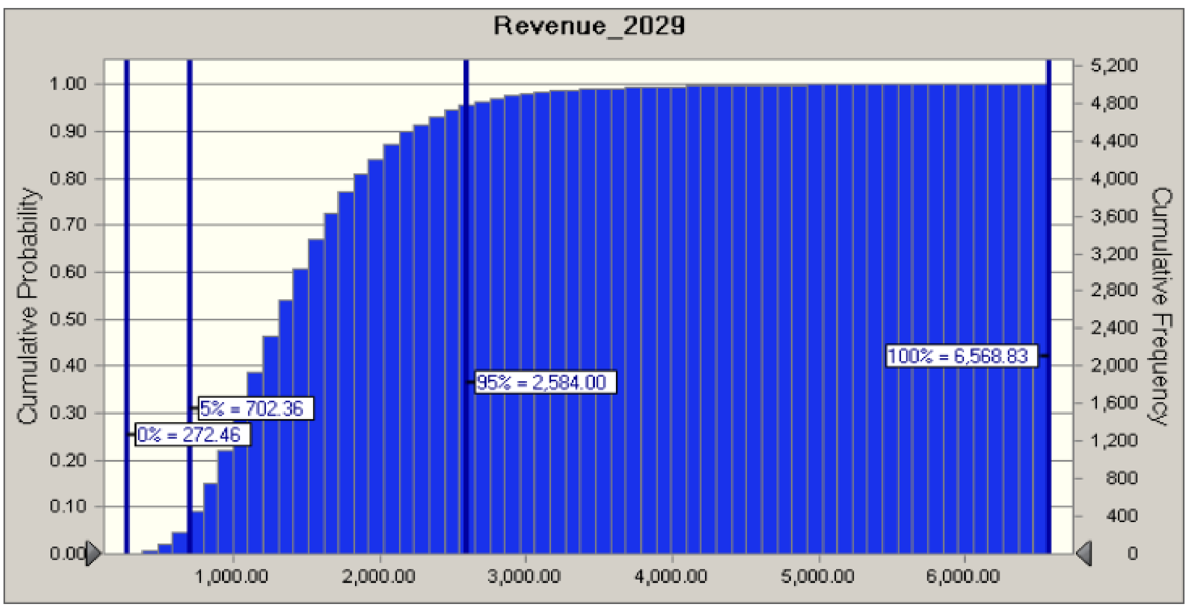


Figure 9 Range of Revenue for Year 2040 Phase 1 Blended HSR System

Contribution to Variance by each Risk Factor

Figures 10 through 13 show the contribution of each risk factors to the overall variance in revenue for each of the scenarios. The Crystal Ball software approximates the sensitivity test using the following approach:

1. Calculate the rank correlation coefficients between every risk assumption and revenue;
2. Calculate the "Contribution to Variance" by squaring the correlation coefficients and normalizing them to 100 percent[6].

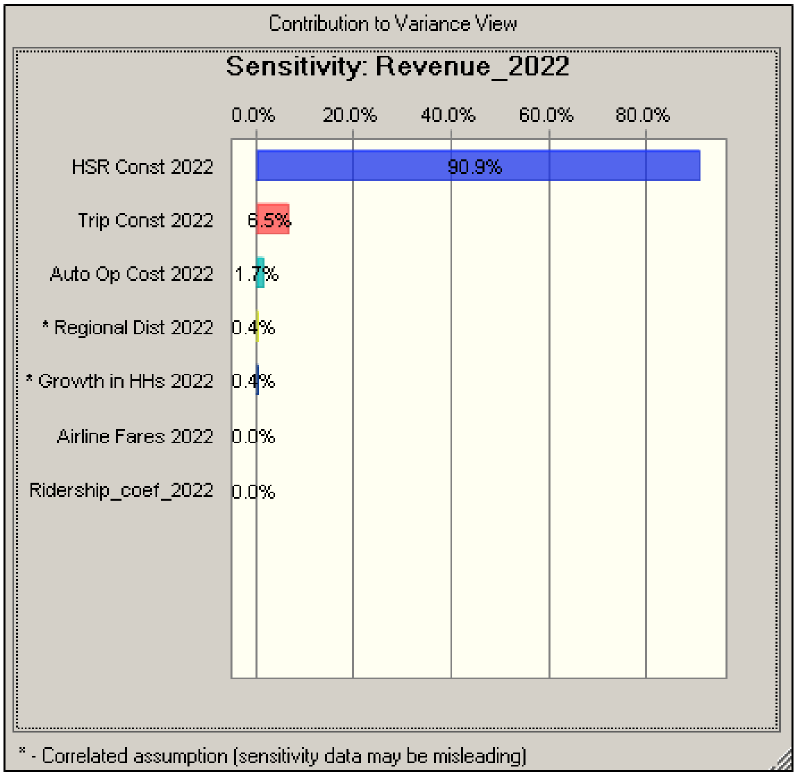


Figure 10 Percent Contribution to Variance of each Risk Factor for Year 2022 Initial Operating Segment

6 Crystal Ball's sensitivity analysis is only an approximation and is not precisely a variance decomposition, and is known to provide inaccurate results for correlated assumptions. The risk factors for growth in households and regional distribution of households are highly correlated (0.95 correlation coefficient) in the Monte Carlo simulation runs. However, because of small correlations of these two factors with revenue, they don't pose a significant issue in the sensitivity tests. Also, note that "Ridership_coeff_…" factor refers to the regression equation constants.

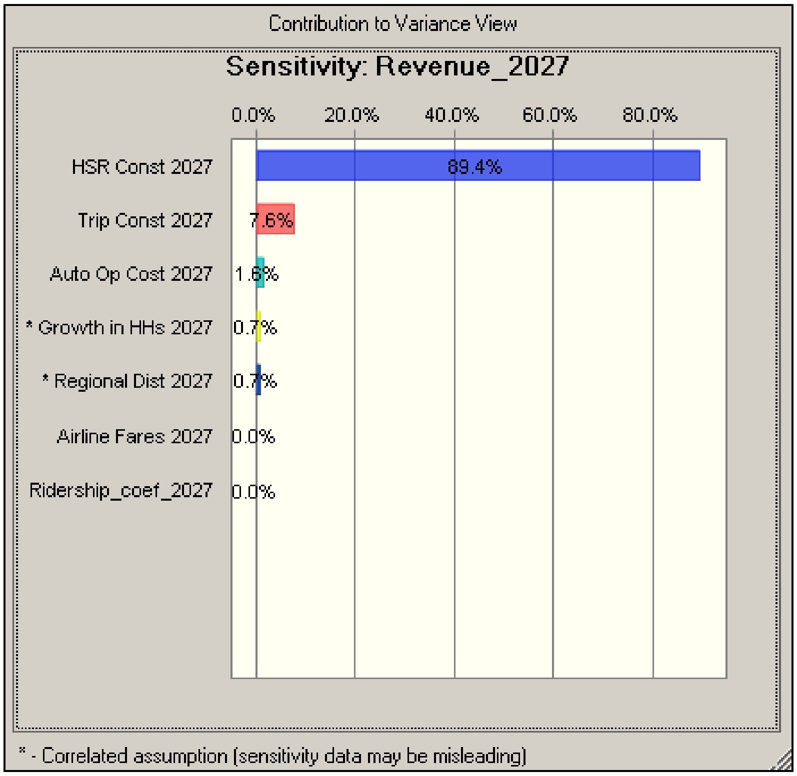


Figure 11 Percent Contribution to Variance of each Risk Factor for Year 2027 Bay-to-Basin

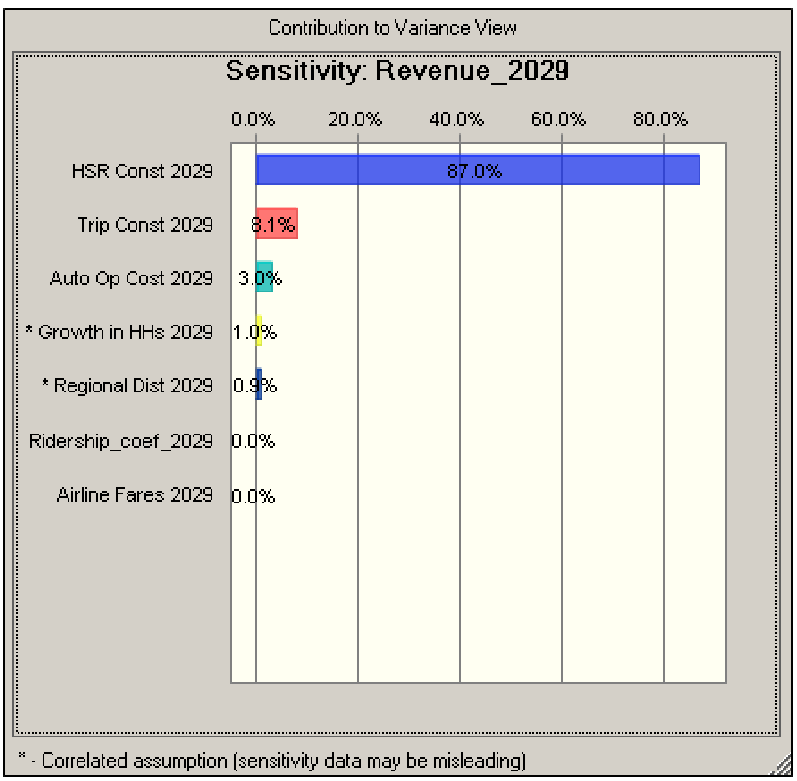


Figure 12 Percent Contribution to Variance of each Risk Factor for Year 2029 Phase 1 Blended

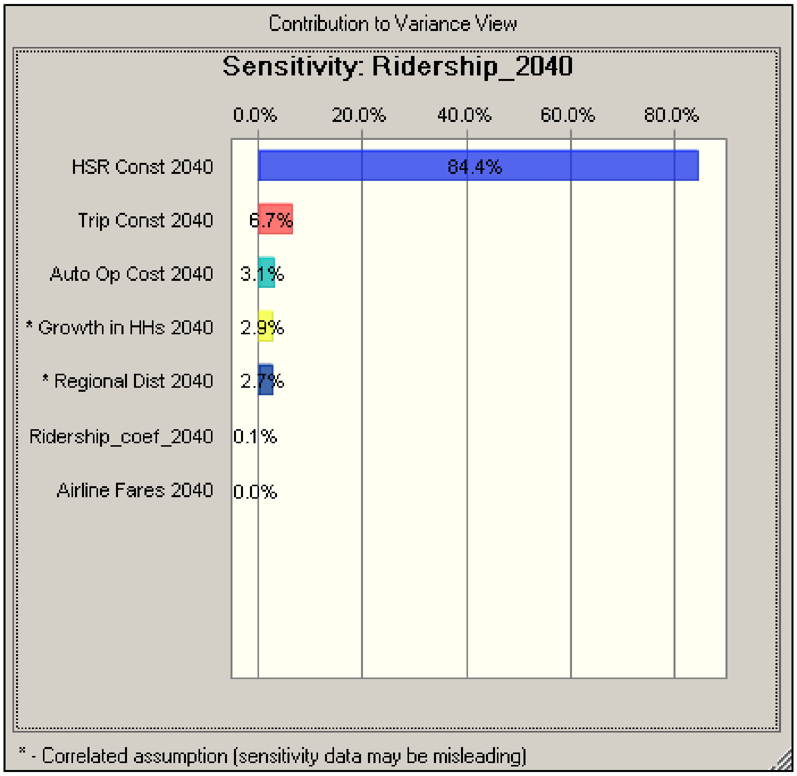


Figure 13 Percent Contribution to Variance of each Risk Factor for Year 2040 Phase 1 Blended

The HSR constant is the overriding risk factor that contributes to the variance in revenue, contributing between 84 and 91 percent of the variance, depending on the forecast year and extent of the HSR project. The large sensitivity of the HSR constant to overall revenue is a reflection the overall uncertainty of the attractiveness of HSR within the California transportation environment as shown by the high standardized estimate of the coefficient within the regression equations (shown in Table 6-9) and the estimated standard deviation for the HSR constant distribution.

Appendix A

California High Speed Rail Ridership and Revenue Forecasts for 2014 Business Plan
Potential Risk Factors and Implications for Forecasting

Risk Factor	Discussion
Future Expectations Risks	
State Growth and Fiscal Changes: **(relative to CSTDM projections):**	
Overall Growth: 1) Increase or decrease in overall expected level of households and/or employment; 2)Variation in growth rates over time.	CS has documented substantial variation in long-range population and employment forecasts over the last ten years. While current forecasts from different sources show similarities in the 2040/2050 timeframe, the sources differ as to growth rates in intervening years. This is a significant uncertainty, and included in the risk analysis.
Household Income: Changes in the number of high/middle/low income households throughout the State or in certain regions.	This is an important consideration since interregional trip rates vary by household income levels. Latest SANDAG forecasts show overall shift to poorer households. Other demographers are projecting increase in unskilled immigration combined with increased domestic out-migration of skilled labor. The current CSTDM socioeconomic projections may be the most optimistic scenario in terms of household income. Capturing the range of potential permutations here, especially at the regional level would be an enormous effort. We suggest that we keep this analysis at the state level, and evaluate uniform changes in a sensitivity evaluation.

Risk Factor	Discussion
Household Size: Changes in number of residents or workers per household.	This is an important consideration since long-distance trip rates vary between household size and workers/household. There is high uncertainty in household size characteristics given current state growth policies, aging population, and large reductions in fertility rates among immigrant populations. We also need to explore the extent to which household size is correlated with household income.We recommend developing sensitivity tests to evaluate the impacts of these factors.
Regional Spatial Distribution: Changes in planned development densities and/or housing types within major metropolitan areas.	California MPOs are projecting increased development density and "jobs-housing balance" as a result of SB-375. Much new growth is being assumed in areas well-served by transit and in proximity to HST stations. Both assumptions represent a departure from trend conditions. We suggest one scenario as a sensitivity test that assumes continuation of trend development patterns rather than increased development density and jobs/housing balance throughout State. This will provide a sense of scale of the impact on potential high-speed rail ridership.
Statewide Spatial Distribution: Different household and employment allocation between San Diego, SCAG, San Joaquin Valley, Bay Area, and Sacramento regions.	As noted above, State policy is trying to encourage more jobs-housing balance, particularly for the San Joaquin Valley and Inland Empire. This policy shift seems to be playing out in the MPO employment forecasts (instead of households). Since the HSR system will travel through the San Joaquin Valley, the risk analysis model will vary the ratio of households and employment within San Joaquin Valley to the rest of the State.
Job Types: Changes in job growth rates in key industries.	This is more subtle, and more difficult to evaluate in the risk analysis. If deemed important, we can handle this risk with sensitivity tests.

Risk Factor	Discussion
Changes in large California attractions: • Beaches wiped out due to climate change or manmade disaster (e.g., oil spill) • Yosemite and other natural parks eliminated (or less attractive) due to federal budget cuts or climate change • Disneyland closes • Googleland and Facebookland open to public in Silicon Valley	Over the course of a generation or two, it is reasonable to expect that people's tastes will change, and long-time popular attractions could go out of business or reduce in size. Witness the rise and fall and rise of Atlantic City, New Jersey. While not impossible, we believe these risks are so speculative that they can be ignored in the risk analysis, but can be suggested as considerations in our report.
Transportation System Changes:	
Automobile fuel cost	The cost of auto fuel is volatile both in short term and over the long term, subject to the uncertainties of geology, global economics and geopolitics, environmental concerns and others. Our previous analysis showed considerable sensitivity to this variable. Therefore, we suggest that this variable be incorporated directly into the risk analysis.

Risk Factor	Discussion
Highway capacity	Highway capacity assumptions in key urban and interregional corridors could be different than planned for any number of reasons, but in particular, more or less funding than implied by adopted plans, increased O&M costs which will leave less funding for new capacity, and policy shifts for or against highway expansion. Highway capacity assumptions affect peak and off-peak travel speeds, which in turn affects each model step. Alternate highway capacities should be evaluated as a sensitivity test. Changing the highway network is a labor intensive exercise that then requires re-running highway skims. In the past, we have tested the effect of changing highway travel times by factoring up or down the travel time skims. We will test the implications of differences in skims via sensitivity tests
Security/screening changes resulting in longer or shorter terminal times for air, high speed rail, or conventional rail.	Security screening on HST could increase terminal times. It could also change the mode specific constant (due to increased inconvenience in relation to air and conventional rail). Since HSR security screening is outside of the HSRA's control, we should include this in the risk analysis. We believe the most important risk relates to potential screening for high speed rail. Since the impact of terminal time is rolled into the constant, we will incorporate this risk analysis into the overall testing of the high speed rail constant described under model-related risks.

Risk Factor	Discussion
Airline ticket prices and frequency of service: • Increase or decrease in ticket prices due to factors such as fuel cost or competitive response. • Increase or decrease in frequency due to competitive response • Changes in pricing policies, such as elimination of baggage and other fees, or increases in such fees (relative to today's levels).	Since airlines compete directly with HSR service, and price is an important factor. . We will use the range of airline ticket prices developed for the last Business Plan by Geoff Gosling as a basis for this business plan. This variable is incorporated directly into the risk analysis. Similarly, airlines could choose to reduce or eliminate air service in certain markets in response to rail competition. Its also possible that non-stop service could be introduced between additional California city pairs as a competitive response. Non-price related airline level-of-service changes will be handled with sensitivity tests.
Changes to the automobile travel experience, such as: • Readily available real time traveler information • Driverless cars (e.g., Google Cars)	Real time traveler information has become common in the last few years. Although it will help people make choices about when they might drive, we do not expect it to be a big factor in choosing driving over traveling by rail. Driverless cars, on the other hand, would significantly change the driving experience, creating, in essence, a new travel mode. We have not included driverless cars in our stated preference surveying efforts, so incorporating this new mode into our analysis would not be possible for the 2014 Business Plan. However, we should point this out as a potential risk factor in our documentation (as we did for the 2012 Business Plan), and consider sensitivity tests that change the attractiveness of automobile travel.

Risk Factor	Discussion
Changes in HSR service characteristics, such as frequency, price, or travel time, or introduction of airport style security lines.	There could be a variety of reasons why the HSR service might not be delivered as proposed in the 2014 Business Plan. While these are real risks, our analysis will be cleaner and easier to understand if we assume the service levels proposed by the HSRA and handle any variations in these service levels as system alternatives that could be handled with sensitivity tests. However, some of the recent criticisms about the California High Speed Rail project focus on disbelief that the HSRA can achieve the service characteristics proposed. A separate analysis of the implications of less favorable characteristics would be reasonable.
Model Related Risks	
Overall amount of long distance travel. This aspect of model-related risk is related almost exclusively to the trip frequency model.	This could be reflected in the trip frequency values, and would be an appropriate value to test in the risk analysis. It could be reflected by a modification of the alternative specific constants for "make a trip."
Amount of travel by trip purpose. This aspect of model-related risk is also associated almost exclusively with the trip frequency model.	Home-based long distance travel is forecast for four different trip purposes: business, commute, recreation, and other. The variability in the percentage of trips for each trip purpose found by different surveys suggests that either the "true" distribution of trips by purpose are not adequately captured, or that the distribution of trips by purpose varies over time. Since we are accounting for the risk associated with the overall average number of annual long-distance trips made by an individual, we do not think it is necessary to add an additional risk factor varying the rate by trip purpose. This risk will be handled via sensitivity analysis.

Risk Factor	Discussion
Amount of travel induced by the introduction of HSR. This aspect of model-related risk is related to both the trip frequency and destination choice models	There are two components of induced travel on HSR: (1) new travel resulting from increased accessibility afforded by HSR (we'll call this) and (2) new travel on HSR resulting from changes in destination choice due to the increased accessibility afforded by HSR (we'll call this). We'll call the first type of induced travel "raw induced travel" and the second type "destination induced travel." We'll call the sum of the two, "total induced travel." While the amount of total induced travel can have high variability within the forecasts, we expect that raw induced travel will comprise a small percentage of overall HSR ridership. The impacts of raw induced travel can be accounted for in the total amount of long distance travel analyzed through changes in trip frequency. The destination induced travel impacts are probably greater. However, these impacts can probably be taken into account through the analysis of different land use patterns. Alternatively, they might be analyzed through varying the logsum coefficients in the destination choice models. A sensitivity test might be warranted.
Share of travel that can be captured by HSR. This aspect of model-related risk is related exclusively to the main mode choice model.	Since HSR does not exist in the United States, the only basis for estimating the relative attractiveness of HSR to other modes comes from the stated preference survey. We cannot calibrate the HSR constant to actual HSR service. Variation in the HSR constants would be appropriate in the risk analysis.