CAMBRIDGE
SYSTEMATICS

California High-Speed Rail 2018 Business Plan

Ridership and Revenue Risk Analysis

technical
supporting document

prepared for

California High-Speed Rail Authority

prepared by

Cambridge Systematics, Inc.

June 1, 2018

www.camsys.com

technical supporting document

California High-Speed Rail 2018 Business Plan

prepared for

California High-Speed Rail Authority

prepared by

Cambridge Systematics, Inc.
10415 Morado Circle, Building II, Suite 340
Austin, TX 78759

date

June 1, 2018

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Forecasts of California High-Speed Rail (HSR) ridership and revenue are estimated using a travel demand model. The Business Plan Model – Version 3 (BPM-V3) travel demand model predicts, for a specified forecast year, the number of annual trips made by households in California, where these trips go within California, and the mode of transportation (i.e., auto, air, conventional rail, or HSR) used to make these trips. In order to predict this travel behavior, forecasts of a variety of inputs that impact travel behavior such as population, travel costs, and transportation networks are prepared as inputs for each forecast year. However, the precise values of these inputs are uncertain (e.g., future auto operating costs). Furthermore, other factors, such as travel patterns and travelers' sensitivities to travel time, evolve over time. Thus, to fully understand the uncertainty in the HSR forecasts of revenue and ridership, the full range of probable values for these input variables should be analyzed.

The purpose of this risk analysis is to incorporate the uncertainty associated with model inputs and assumed travel behavior into the 2018 Business Plan HSR ridership and revenue forecasts. This risk analysis approach builds on the previous risk analysis procedures used for the 2014 and 2016 Business Plans. The approach allows the California High-Speed Rail Authority to express probabilities of achieving different forecast results for ridership and revenue.

To develop the full range of possible ridership and revenue forecasts, 150 full model runs were performed for each forecast year to estimate relationships between forecast revenue and ridership and selected input risk variables. These runs were used to create two models of the model outputs, or "meta-models," for each forecast year. The revenue meta-model and the ridership meta-model were used to generate thousands of revenue and ridership forecasts over the entire ranges of identified risk variables without requiring computationally expensive and time-consuming full model runs.

The initial step in the risk analysis was the identification of potential risk factors that could impact ridership and revenue forecasts (e.g., potential changes in auto operating costs or the impact of new technologies, such as autonomous vehicles). Second, the impact of each risk factor was assigned to a model variable or variables, and the variables were systematically narrowed to the set of inputs that would have the highest combination of uncertainty and impact on the forecasts. Third, the meta-model was coupled with distributions of the model inputs developed and used in a Monte Carlo simulation to develop 100,000 unique forecasts of revenue and ridership. Finally, probability distributions of total revenue and ridership were estimated from the results of the Monte Carlo simulation.

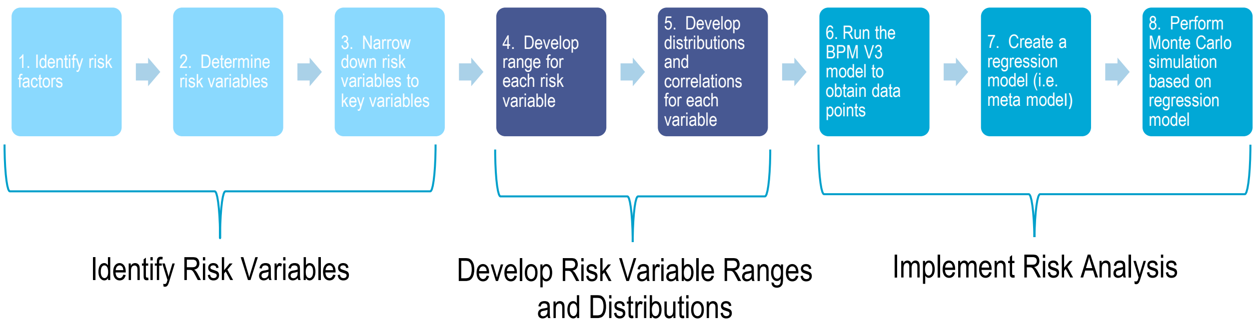
This methodology is similar to the methodology employed for the 2016 Business Plan Risk Analysis. It was refined and enhanced by the following:

- Adopting a combined Gaussian Process Regression (GPR) and Linear Regression meta-model to create inputs for the Monte Carlo risk model.
- Replacing the two-step experimental design process with a modified Latin hypercube experimental design.
- Increasing the number of risk factors considered in the risk analysis.

1.1 Overview of the Risk Analysis Approach

An eight-step risk analysis approach was employed to forecast revenue and ridership for the 2018 Business Plan, as shown in Figure 1.1.

Figure 1.1 Eight-Step Risk Analysis Approach



The steps to identify the model assumptions are described below.

Step 1. Develop a list of possible risk factors to be considered for the revenue and ridership risk analysis

- Risk factors are defined as any circumstance, event, or influence that could result in the HSR revenue and ridership deviating from its forecasted value.
- A panel of experts was used to develop a set of potential risk factors that could impact future HSR ridership and revenue.
- The final risk factors for each forecast year were chosen based on their likelihoods of affecting ridership and revenue for the forecast year.

Step 2. Identify risk variables for each risk factor

- Risk variables are actual variables and constants that can be adjusted in the BPM-V3. As an example, auto operating cost (i.e., cost, in dollars, per vehicle mile driven) is a variable that can be adjusted in the model. To address the possibility that fuel cost and fuel efficiency may be higher or lower than predicted, auto operating cost may be increased or reduced in the risk analysis to test how these two risk variables affect ridership and revenue.
- The risk variables have been chosen to represent one or more risk factors identified in Step 1.

Step 3. Narrow risk variables to key variables for inclusion within each forecast year of analysis

- Sensitivity runs of the BPM-V3 were performed for each risk variable that allowed for a quantitative comparison of the impacts of each risk variable on ridership and revenue.

- Based on the range and known sensitivity of the risk variables under consideration, final sets of risk variables were selected for inclusion for each forecast year.

Steps 4 and 5. Develop a range and distribution for each risk variable under consideration

- The uncertainty associated with each risk variable was quantified by assigning a range and distribution for each variable. For example, based on the research on each risk factor affecting auto operating cost, such as fuel cost and fuel efficiency, auto operating cost in year 2029 is predicted to range from $0.17 per mile to $0.35 per mile (stated in June 2017 dollars[1]), with a most likely value of $0.23 per mile.

- For each risk variable, the minimum, most likely, and maximum values for each forecast year were developed based on currently available research and analysis.

- The shape of the distribution of possible values for each variable determined the likelihood of the variable's value, within the set range, under random sampling. For example, it is very unlikely that auto operating cost will be the minimum value of $0.17 per mile or the maximum value of $0.35 per mile, but very likely it will be close to $0.23 per mile. The auto operating cost distribution is defined such that the most likely value will be chosen at a much higher rate than the extreme values, and thus the simulated model runs will be more representative of potential future outcomes.

Steps 6 and 7. Run the BPM-V3 using defined sets of risk variable levels to obtain data points for estimation of two sets of regression models (i.e., meta-models) for each forecast year that estimates the values of the dependent variables, either HSR revenue or ridership, based on values of the selected input risk variables

- The sets of BPM-V3 specified model runs were developed using a modified Latin hypercube sample design process to ensure that the data points represented the solution space effectively[2].

- A Gaussian Process Regression (GPR) was used to develop the meta-model. GPR does not impose a restriction on the functional form of the output (e.g., it does not need to be linear or any particular defined non-linear function). Instead, the functional form is developed on the reasonable assumption that, if two observations have inputs that are similar, then the output should also be similar.

Step 8. Perform a Monte Carlo simulation by running the GPR model 100,000 times with varying levels of the input variables based on the distributions assigned to the variables

- The simulation results in probability distributions of HSR revenue and ridership.

---

[1] All dollar figures presented in this document are base year as of June 2017. The 2018 Business Plan escalates all reported dollar amounts to December 2017 dollars for consistency with base year Capital and Lifecycle costs.

[2] Latin hypercube sampling is a statistical method for producing a close to random sample of values from a multidimensional distribution.

- The results of the simulation were analyzed to determine the relative contribution of each risk factor on revenue and ridership.

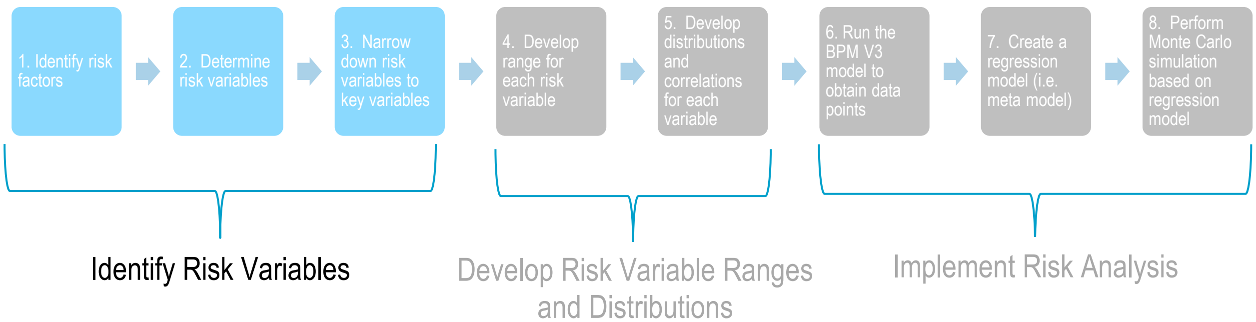
Each step in the risk analysis required thorough evaluation to ensure key risk factors were understood and addressed appropriately. The remainder of this technical supporting document is divided into three sections that provide insight into the steps taken to produce the simulation results followed by eight appendices that provide additional technical detail. The three sections and the associated appendices that provide additional detail are:

- **Section 2.0. Identification of Risk Variables** (Steps 1 to 3).
- **Section 3.0. Development of Risk Variable Ranges and Distributions** (Steps 4 to 5).
  - Appendix A – Risk Variable Component Specification for Monte Carlo Simulation (a summary of the final risk variables and ranges used for the risk analysis)
  - Appendix B – High-Speed Rail Constants
  - Appendix C – Trip Frequency Constants
  - Appendix D – Auto Operating Cost
  - Appendix E – Coefficient on Transit Access-Egress Time/Auto Distance Variable
  - Appendix F – Quantifying the Effects of Autonomous and Shared Use Vehicles on Year 2040 Risk Variables
  - Appendix G – Exceptionally Long Access and Egress: Experience from Japan
- **Section 4.0. Risk Analysis implementation (**Steps 6 to 8).
  - Appendix H – Technical Details for the Application of GPR

2.0 Identification of Risk Variables

This section details the steps taken to identify the risk variables included in the risk analysis, as shown in Figure 2.1.

Figure 2.1 Eight-Step Risk Analysis Approach: Identify Risk Variables (Steps 1 to 3)



To develop a set of potential risk factors (Step 1), Cambridge Systematics, Inc. (CS) started by holding a series of meetings among staff to review the potential risks originally identified by a panel of experts for the 2016 Business Plan Risk Analysis, and identify any changes to those potential risks or new risks that could impact ridership and revenue forecasts. Appendix A summarizes the risk factors considered.

The meetings sought to answer the following question: What real-world risks could impact ridership and revenue in years 2029, 2033, and 2040? These forecast years were chosen based on the Business Plan opening year dates for the Silicon Valley to Central Valley line and for Phase 1 in the 2018 Business Plan as well as a Phase 1 horizon year as shown in Table 2.1.

Table 2.1 Description of Each Phase of the HSR System

Operating Phase	Year	High-Speed Rail Segment	Frequency of Service	HSR Bus and Conventional Rail Connections
Silicon Valley to Central Valley Line (VtoV)	2029	San Francisco to Bakersfield	2 trains per hour during the peak period and 1 train per hour during the off-peak period	Includes bus connections between Bakersfield and Los Angeles, and rail and bus connections between Madera and Sacramento
Phase 1 (PH1)	2033 & 2040	San Francisco and Merced to Los Angeles and Anaheim	Up to 8 trains per hour (from all destinations) during the peak period and 5 trains per hour during the off-peak period	Includes rail and bus connections from Madera to Sacramento and rail connections in Southern California

Table 2.2 identifies the risk variables (i.e., assumptions built into the BPM V3 model) used to represent each risk factor (Step 2). The risk variables identified for each risk factor were determined by answering the following questions: What model inputs and variables best represent or are influenced by the risk factors, and how are the risks best accounted for by the variables in the model? Sensitivity runs of the BPM V3 model were run for each risk variable that allowed for a quantitative comparison of the impacts of each risk variable on ridership and revenue. Based on this sensitivity analysis, the risk variables that were determined to have the greatest effect on HSR ridership and revenue and the highest potential uncertainty for each forecast year were selected for inclusion (Step 3). Because the impact of each risk factor and the variables representing each factor could vary over the life of the project, the list of risk variables differed depending on the operating plan and forecast year under consideration. For example, the uncertainty and impact of HSR bus connections are a concern for earlier years when they are a critical access mode, while the likelihood of significant autonomous vehicle use affecting HSR ridership is not likely until 2040. Appendix A also summarizes the final risk variables and ranges used to represent the risk factors.

Table 2.2 Variables Included in Risk Analysis for Each Analysis Year

Number	Risk Variable	Reasons for Considering Model Variable and Risk Factors Represented
1 (All Years)	Business HSR Mode Choice Constant	The mode constants capture the unexplained variation in traveler mode choices after system variables and demographics are taken into account. Unexplained variation may include factors, such as comfort aboard trains, opinions regarding HSR, need for a car at the destination, level of familiarity with HSR, etc.
2 (All Years)	Commute HSR Mode Choice Constant	
3 (All Years)	Recreation/Other HSR Mode Choice Constant	
4 (All Years)	Business/Commute Trip Frequency Constant	Since the trip frequency model is a logit-based choice model, the constants capture the unexplained variation in the number of long-distance trips that travelers will take after accounting for household demographics and the accessibility of available destinations. In addition, risks associated with the state of the economy are accounted for within the trip frequency constant risk variable.
5 (All Years)	Recreation/Other Trip Frequency Constant	
6 (All Years)	Auto Operating Costs	This variable reflects the inherent risks in forecasting future fuel costs; fuel efficiencies; the adoption of alternative fuels/electric vehicles; maintenance costs; changes in gas taxes; potential impacts of cap and trade on fuel costs; and for 2040, market penetration of autonomous connected vehicles, autonomous vehicle fuel economy, higher shares of "shared use" vehicles, and shared use vehicle operating costs.
7 (All Years)	HSR Fares	A number of issues could affect actual fares charged to travelers, especially as the system is being opened: institution of discount/premium fares (advance purchase, peak/off-peak, first/second class seating); adjustments needed to respond to changing auto operating costs or air fares; yield management strategies; etc.
8 (All Years)	HSR Frequency of Service	With final service plans expected to be developed by a private operator, there is uncertainty around the amount of service that will be provided based on the markets and strategies that the operator may employ.
9 (Year 2029)	Availability and Frequency of Service of Conventional Rail and HSR Buses that connect with HSR	Access to and egress from the system include connections with both conventional rail services and HSR buses (as well as many other modes). Levels of conventional rail service are forecasted based on the State Rail Plan, but there is some uncertainty around the availability of the exact amount of conventional rail service. Similarly, the amount of connecting bus service could be different than currently forecasted. These connections are most critical in the early years of the program when the high-speed rail system does not yet connect the whole state.
10 (2033 and 2040)	Airfares	Airfares change and fluctuate over time. Some possible reasons that airlines may change airfares from currently forecasted levels include changes in fuel or personnel costs or airport landing fees; changes in equipment or efficiency, such as NextGen technology; competitive response to HSR to maintain air market shares; acceptance of HSR as a replacement for inefficient; short-haul air service; etc.

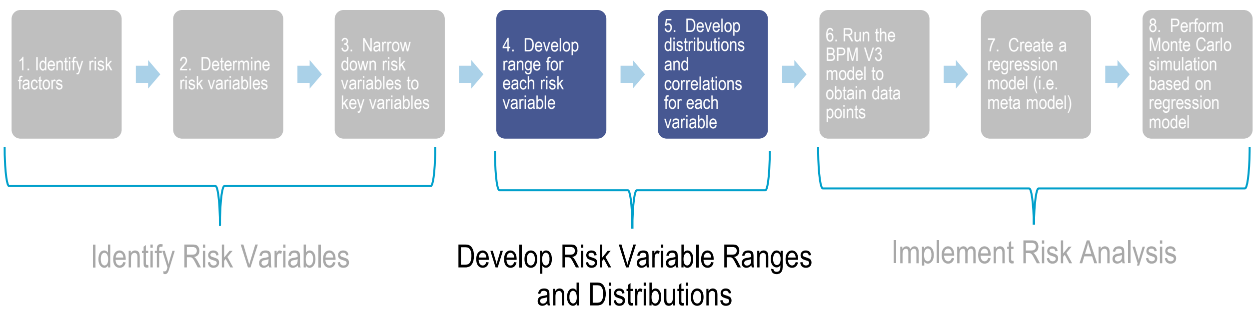
Number	Risk Variable	Reasons for Considering Model Variable and Risk Factors Represented
11 (All Years)	Coefficient on Transit Access-Egress Time/Auto Distance Variable	Between some regions in California, especially in the Silicon Valley to Central Valley line scenario, individuals who wish to travel primarily by transit to reach their destination must transfer from an HSR bus or conventional rail system before or after traveling on HSR. Experience in France has suggested some uncertainty regarding how the need to make these transfers affects overall HSR ridership. The model includes a variable that makes HSR less attractive for trips that require a long access or egress trip in relation to the time spent on HSR (or another public mode such as conventional rail or air), and the variation in this variable is used to estimate the uncertainty around the effect of these transfers on HSR ridership and revenue.
12 (All Years)	Number and Distribution of Households throughout the State	The forecasted number of statewide households can fluctuate for a variety of reasons, such as inherent uncertainty with population forecasts, national and statewide economic cycles, impacts of natural disasters, such as continuing draught, changes in U.S. immigration policy, etc. The uncertainty of population forecasts and the divergence between different forecasts increase the further out the forecasts are considered. The number and distributions of households throughout the state impact household characteristics used in the BPM-V3, such as household size, income group, number of workers, and auto ownership. These household characteristics impact travel behavior.
13 (Year 2040)	Auto In-Vehicle Travel Time Coefficient	The introduction of autonomous vehicles may change the way travelers view auto travel due to the substitution of other activities, such as sleeping, reading, Internet communications for the time spent driving.
14 (All Years)	HSR Reliability	Early implementation issues with equipment and operations could affect HSR reliability in the early stages of each phase. Overall HSR reliability may not match international experience on which the original 99 percent reliability assumption is based.
15 (All Years)	Exceptionally Long Access and Egress	Reliably estimating parameters for exceptionally long access and egress from currently available survey data is nearly impossible. There are very few observed trips with these attributes (e.g., there are no observations of access and egress by any mode over three hours). In addition, access and egress times to main modes are generally correlated: if your origin is very far from an airport, you are usually also very far from a train station, and vice versa. This will not necessarily be the case for HSR, since it is possible to be far from HSR but close to an airport or CVR stations. This risk variable is used as a way to estimate the uncertainty around the effect of exceptionally long access and egress on HSR ridership and revenue.[3]
16 (All Years)	Induced Travel	While the model forecasts induced travel resulting from improved accessibility, the relationships are based on travel made on existing modes. Induced travel forecasted by the model for HSR is low compared to what has been observed on international HSR systems.
17 (All Years)	Visitor Travel	The model only forecasts intra-state travel by California residents. However, in 2016, there were 60 million annual visitors to California. These visitors, especially those that travel by air to arrive in California, may find HSR a desirable option for traveling between various locations in California.

[3] Note that this risk variable is focused on exceptionally long access and egress by any mode in distance ranges where there were virtually no observed data. In contrast, Risk Variable 11 focused on transit access or egress in relation to the total trip distance; while observed data existed to estimate the coefficient, the applicability for ranges of access/egress to HSR is less certain.

3.0 Development of Risk Variable Ranges and Distributions

The uncertainty surrounding each risk variable must be quantified by assigning a range and distribution to each variable. As shown in Figure 3.1, determining the ranges of the risk variables corresponds to *Step 4*, and developing the distributions corresponds to *Step 5* of the risk analysis approach.

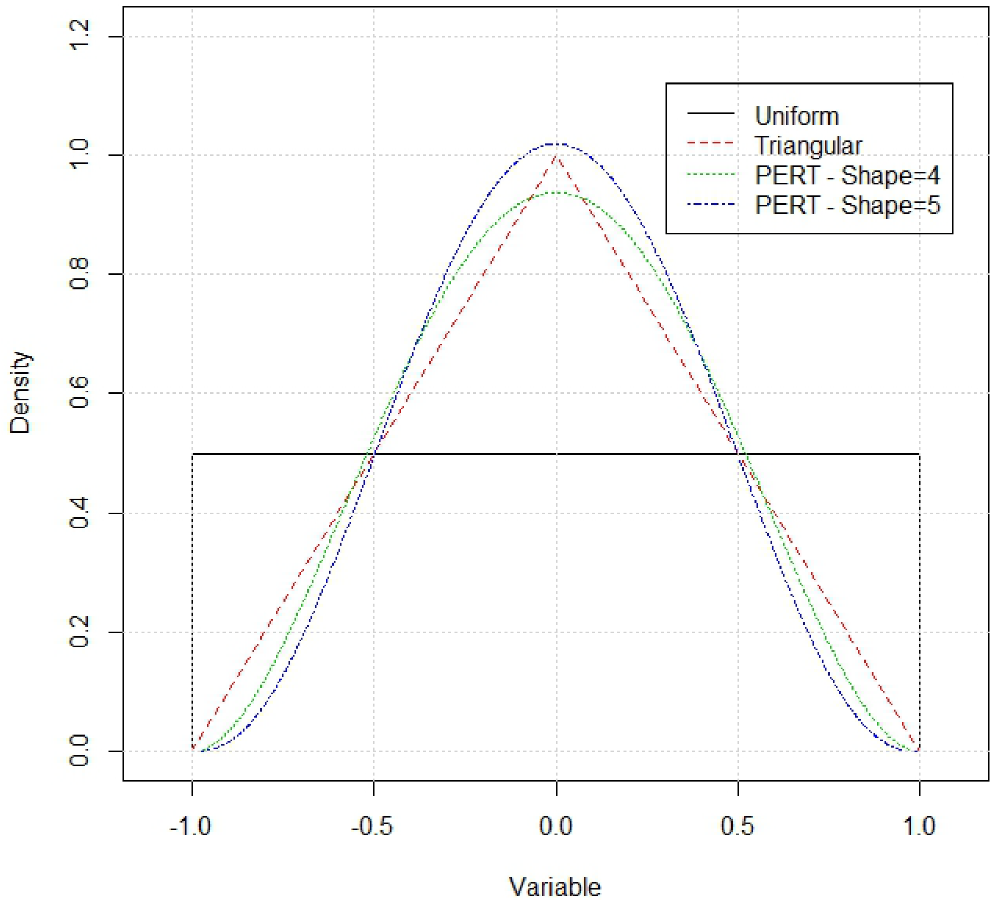
Figure 3.1 Eight-Step Risk Analysis Approach: Develop Risk Variable Ranges and Distributions (Steps 4 to 5)



To perform the risk analysis, a range of possible values for each risk variable has to be established in order to quantify the uncertainty related to that variable. The absolute minimum and absolute maximum values of the variable sets the range of the variable's forecasted value, while the most likely value represents the peak of the variable's distribution. For each risk variable, the absolute minimum, most likely, and absolute maximum values were based on research and analysis of currently available sources.

A distribution around the minimum, most likely, and maximum values of each risk variable was determined based on the characteristics of these three points. The shape of the distribution determines the likelihood of the variable's value, within the set range, under random sampling. The most likely value has the greatest likelihood of occurring within the distribution. The shape of the distribution can be triangular, PERT, uniform, or another form. PERT distributions were used for variables where there are significant tails based on the values assumed for the minimum and maximum (i.e., the minimum and maximum are extreme values). A Shape = 4 PERT distribution was assumed to be standard with a higher Shape used for Auto Operating Costs, because the maximum and minimum involve several independent downside or upside events taking place at the same time, which makes the extreme values less likely (and justifies longer, thinner tails). Triangular distributions were used where there is less information about the exact shape, but values around the most likely are more likely to occur than the values closer to the minimum and maximum (though not to the same extreme as for the PERT distributions). Uniform distributions were used where there is high uncertainty regarding the forecast values for the risk variable. Figure 3.2 illustrates the shapes of the different distributions used in the risk analysis.

Figure 3.2 Shapes of the Distributions Used in the Risk Analysis



The following sections identify the ranges of values and distribution for each risk variable, and summarize the research and methodology for developing the absolute minimum, most likely, and absolute maximum value. The risk variables are used for all forecast years unless noted (see Table 2.2). Appendix A summarizes the final risk variables and ranges used to represent the risk factors.

3.1 High-Speed Rail Constant

The HSR constant for each of the four trip purposes (i.e., business, commute, recreation, and other) is composed of two components: 1) unexplained variation, and 2) terminal and wait time. The unexplained variation component represents the desirability of HSR as a mode that is not captured directly by the system variables (e.g., travel time, cost, etc.) included in the model. Terminal time is the out-of-vehicle time spent traveling from the point of departure from the access mode to the train platform. Wait time is the out-of-vehicle time spent waiting on the platform for the train to arrive and the time spent waiting for the train to leave the platform once boarded[4]. For all forecast years, the range for the HSR constant was defined as:

- **Minimum**. HSR will be perceived by travelers as an equivalent mode to Conventional Rail (CVR), and terminal plus wait time will be 45 minutes;
- **Most Likely.** Calibrated HSR constant with terminal plus wait time of 25 minutes; and
- **Maximum**. Assumes the impact of unexplained variation is symmetrical around the calibrated constant (i.e., the difference between the calibrated and CVR constants, excluding terminal and wait time effects, can be added to the calibrated HSR constant to represent that HSR is even more desirable), but terminal plus wait time will be 15 minutes.

The bundling of the unexplained variation and terminal/wait time into a single constant for the BPM-V3 allows for the estimation of a single regression model parameter in the meta-model used for the risk analysis. For the Monte Carlo risk analysis, the unexplained variation and terminal/wait time components of the HSR constant are, effectively, unbundled and considered as separate risk variables with independent distributions.[5] This approach does not require an additional risk variable in the experimental design framework yet makes it possible to understand the uncertainty associated with HSR terminal/wait time on ridership and revenue independently from the uncertainty associated with the unexplained variation of the HSR constant on ridership and revenue since the two variables do not necessarily move together. Table 3.1 summarizes the offsets to the HSR constants used for the unexplained variation and terminal/wait time components. See Appendix B for additional information on the development of the range and distribution of the components of the HSR constant.

Table 3.1 Ranges of HSR Mode Specific Constant Offsets

		Implied Annual Long-Distance Round Trips per Capita After Applying Offsets		
Trip Purpose	**Constant Component**	**Minimum**	**Most Likely**	**Maximum**
Business	Unexplained Variation	-2.335	0.0	2.335
	Terminal/Wait Time	-0.3264	0.0	0.1632
Commute	Business/Commute	-1.222	0.0	1.222
	Recreation/Other	-0.3264	0.0	0.1632
Recreation/Other	Business/Commute	-1.354	0.0	1.354
	Recreation/Other	-0.1388	0.0	0.0694

[4] In the BPM-V3, wait time is mode specific and not a function of frequency of service. Frequency of service is an explicit variable in the BPM-V3 for the public modes, air, CVR, and HSR.

[5] The unexplained variation uses a PERT distribution, while the terminal/wait time uses a triangular distribution.

3.2 Trip Frequency Constant

The trip frequency constants include the unexplained variation in the propensity of households to make long-distance trips within California. The range and variance of the trip frequency constants (i.e. business, commute, recreation, and other) are included in the risk analysis in order to reflect both the unexplained variation in long distance trip making and the effect of the state of the economy on the proclivity of households to undertake long-distance travel. While "the economy" is an overarching risk that affects many different decisions regarding travel, one of the most direct and principal impacts on HSR ridership and revenue is whether a long-distance trip is even made. The state of the economy affects household income and employment levels; the levels of these variables are directly taken into account within the model to determine long-distance trip frequency and, by extension, average trip frequency rates[6] (e.g., people who are out of work or have reduced income due to a recession make fewer long-distance trips). Instead of including the distributions of households by various socioeconomic strata directly as risk variables in the risk analysis model to account for changes in the state of the economy, the effects of these risk variables on trip frequency levels are accounted for within the trip frequency constant risk variable.

The unexplained variation range is based on the range seen in forecasted annual long-distance trip rates produced by the model. The most likely value for each forecast year is the calibrated constant. The minimum value of the trip frequency constants is specified such that, for year 2040, the trip frequency constants produce average trip rates equal to the 2010 rates by trip purpose (long-term trends show people's propensity for making long-distance trips increasing over time). For the maximum value, the trip frequency constant is specified to mirror the deviations from the calibrated constants for the minimum values (i.e., symmetry of the constant offsets is assumed).

The range of trip frequency constant offsets for the economic cycles provides a proxy for the underlying economic-cycle risk variables being considered. This approach provides a method for specifying a continuous range of outcomes, rather than developing multiple input socioeconomic datasets. The economic-cycle range was developed by calculating the implied trip rates based on changes in the number of workers and income levels from the following scenarios:

- **Minimum**. Based on HSR-implied trip rate decrease resulting from a three-percent per year decrease in employment from the low-growth scenario for three years preceding the forecast year. The direct impact of the low economic cycle on trip frequency is determined by changing the distributions of households by number of workers and households by income group to reflect the three-percent per year decrease in employment.

- **Most likely**. Resulting trip rates obtained using calibrated trip frequency constants.

- **Maximum**. Based on HSR-implied trip rate increase resulting from a three-percent per year increase in employment from the high-growth scenario for five years preceding the forecast year. The direct

[6] Households stratified by household size, income group, auto ownership, and number of workers are input to the BPM-V3. The trip frequency choice model includes coefficients for different levels and combinations of these variables (e.g., low or high income). Thus, changing the distribution of the households to account for different socioeconomic conditions effectively changes the overall average per household trip rates.

impact of the high economic cycle on trip frequency is determined by changing the distributions of households by number of workers and households by income group to reflect the three-percent per year increase in employment.

The offsets for unexplained variation and the economic cycles were combined to represent the full range of possible outcomes for the development of the risk analysis regression equations. After the constant offsets were added together, it was possible to estimate the resulting implied ranges of annual long-distance round trips per capita as shown in Table 3.2. The implied trip rates vary by year since the trip frequency choice model considers the demographic make-up of the population as well as accessibility to possible destinations in the determination of whether or not a household member makes a long-distance trip. For comparison, the trip frequency constants for the BPM-V3 were calibrated to match “most likely” 2010 trip rates of 1.87 and 5.14 annual long-distance round trips per capita for business/commute and recreation/other, respectively.

Table 3.2 Ranges of Implied Annual Round Trips per Capita for Full Model Runs Based on Trip Frequency Constant Offsets

		Implied Annual Long-Distance Round Trips per Capita After Applying Offsets		
Model Year	**Purpose**	**Minimum**	**Most Likely**	**Maximum**
2029	Business/Commute	1.41	2.21	3.44
	Recreation/Other	4.83	5.86	7.12
	Total	6.24	8.07	10.56
2033	Business/Commute	1.46	2.28	3.54
	Recreation/Other	4.90	5.95	7.22
	Total	6.36	8.23	10.76
2040	Business/Commute	1.57	2.46	3.79
	Recreation/Other	5.15	6.27	7.59
	Total	6.72	8.73	11.38

For the Monte Carlo risk analysis, each component of the trip frequency constant is considered as a separate risk variable with independent distributions (i.e., 0 percent correlation). The unexplained variation uses a PERT distribution, while the economic cycle component uses a triangular distribution. A 50-percent correlation is assumed between the business/commute and recreation/other risk components for unexplained variation, since there is likely to be some relationship (though not perfect correlation) in changes to overall trip-making for different purposes. Perfect correlation is assumed between economic-cycle risk components for business/commute and recreation/other purposes. More information on the development of the range and distribution of the trip frequency constant components is detailed in Appendix C.

3.3 Auto Operating Cost

The auto operating cost forecasts for year 2029 and 2033 is assumed to be associated only with privately owned non-autonomous vehicles. Auto operating cost is calculated from the following components:

1. Retail fuel prices in California, which are projected using the U.S. Energy Information Administration (EIA) forecasts with an assumption that California prices are 12.8 percent higher than the national average (based on past trends).

2. An estimate of the market penetration rate of electric vehicles, along with accompanying costs for electricity, miles per gallon equivalent (MPGe) rating to determine energy costs for electric vehicles, and the cost of electricity. These estimates were developed from the 2017 Annual Energy Outlook produced by the EIA.

3. Additional fees and charges based on two scenarios:

    a. Cap and Trade implementation (i.e., 5 to 22 percent impact on retail fuel prices);[7] and

    b. A potential increase in Federal excise taxes.

4. The fuel economy of the entire "on the road" fleet, calculated from the 2017 Annual Energy Outlook (AEO).

5. Nonfuel costs, which were obtained from the Bureau of Transportation Statistics (BTS).

More information on the development of each of these components can be found in Appendix D. The minimum, most likely, and maximum were set based on the combined impacts of these components. Note that the minimum and maximum scenarios are intended to be extreme ends, and so individual components do not correlate as one might expect. For instance, the minimum scenario assumes the highest penetration rate of electric vehicles (EV) and the lowest price of gasoline and electricity. As found in the 2017 AEO, Californians are more likely to adopt electric vehicles as the cost of owning a conventional internal combustion engine (ICE) vehicle rises, but that likely correlation is set to the side in developing the minimum and maximum in order to provide the widest reasonable range of values.

Thus, the minimum combines the lowest fuel price projection, the greatest percentage of electric vehicles, high fuel efficiency for the entire vehicle fleet, the least impact from cap and trade, no increase in Federal taxes, and low nonfuel costs. This approach is reflected in the following formulas, which were used to calculate the minimum, most likely, and maximum auto operating cost:

[7] The exact impact of Cap and Trade on fuel prices is unknown and could change over time based on the industry response to reduce emissions. The California Air Resources Board estimated in 2010 that gasoline price changes in 2020 could range between 4 percent and 19 percent due to Cap and Trade rules (http://www.arb.ca.gov/regact/2010/capandtrade10/capv4appn.pdf). The minimum assumption assumes that Cap and Trade would not result in an increase in gas prices.

(Footnote continued on next page...)

*Minimum Auto Operating Cost = (1 - %EVs) * (Low CA Gas Price + Low C&T Impact + No Increase in Federal Gas Tax) / High ICE Fuel Efficiency + %EVs * (Low CA Electricity Price * 33.7[8]) / High EV Fuel Efficiency + Low Nonfuel Operating Costs*

*Most Likely Auto Operating Cost = (1 - %EVs) * (Most Likely CA Gas Price + Avg(Low C&T Impact, High C&T Impact) + No Increase in Federal Gas Tax) / Most Likely ICE Fuel Efficiency + %EVs * (Most Likely CA Electricity Price * 33.7) / Most Likely EV Fuel Efficiency + Most Likely Nonfuel Operating Costs*

*High Auto Operating Cost = (1 - %EVs) * (High CA Gas Price + High C&T Impact + Increase in Federal Gas Tax) / High ICE Fuel Efficiency + %EVs * (High CA Electricity Price * 33.7) / High EV Fuel Efficiency + High Nonfuel Operating Costs*

Table 3.3 gives the auto operating cost component values and the resulting minimum, most likely, and maximum auto operating cost for each forecast year. Since auto operating cost comprises individual components that each has minimum and maximum values (as described above), auto operating costs utilize a Shape = 5 PERT distribution. This distribution has somewhat longer tails since the very low or high end of the range requires each of the individual components to end up on the low or high end, which is a very unlikely occurrence.

For year 2040, in addition to privately owned non-autonomous vehicles, it is possible that autonomous vehicles and shared-use vehicles will have high enough market penetration to affect the overall auto operating cost for long-distance trips. Appendix F provides background on auto operating costs for autonomous and shared use vehicles and their impacts on overall auto operating costs as used for the 2040 analysis. Based on the adjustments for autonomous and shared-use vehicles, the year 2040 auto operating cost ranges from $0.12 per mile to $0.38 per mile, with a most likely of $0.23 per mile.

Table 3.3 Range of Auto Operating Cost for Each Forecast Year by Auto Operating Cost Component (*June 2017 Dollars*)

	Minimum	Most Likely	Maximum
2029 Auto Operating Cost ($/Mile)	$0.17	$0.23	$0.35
U.S. Gas Price ($/gal)	$1.90	$3.06	$5.33
California Gas Price ($/gal)	$2.14	$3.45	$6.01
California Electricity Price ($/kWH)	$0.17	$0.17	$0.18
% Electric Vehicles	10.51%	7.86%	6.70%
MPG	31.5	30.5	29.8
MPGe	78.64	76.14	74.53
Non-fuel cost ($/mi)	$0.10	$0.11	$0.12
Cap & Trade ($/gal)	$0.23	$0.39	$0.73

[8] The EPA estimates that each gallon of gasoline contains 33.7 kWH of energy. This number provides a common conversion factor between the two kinds of vehicles. (For more information see the May 2011 Regulatory Announcement EPA-420-F-11-017 published by the Office of Transportation and Air Quality.)

	Minimum	Most Likely	Maximum
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14
2033 Auto Operating Cost ($/Mile)	$0.17	$0.23	$0.34
U.S. Gas Price ($/gal)	$1.95	$3.19	$5.52
California Gas Price ($/gal)	$2.20	$3.59	$6.22
California Electricity Price ($/kWH)	$0.18	$0.18	$0.18
% Electric Vehicles	13.35%	9.75%	8.10%
MPG	34.3	32.8	31.9
MPGe	85.84	82.11	79.85
Non-fuel cost ($/mi)	$0.10	$0.11	$0.12
Cap & Trade ($/gal)	$0.26	$0.52	$0.77
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14
2040 Auto Operating Cost ($/Mile)[1]	$0.17	$0.23	$0.33
U.S. Gas Price ($/gal)	$2.05	$3.39	$5.75
California Gas Price ($/gal)	$2.32	$3.82	$6.49
California Electricity Price ($/kWH)	$0.19	$0.19	$0.19
% Electric Vehicles	17.16%	12.23%	9.96%
MPG	37.9	35.5	34.2
MPGe	92.76	88.79	85.55
Nonfuel cost ($/mi)	$0.10	$0.11	$0.12
Cap & Trade ($/gal)	$0.33	$0.58	$0.84
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14

[1] The 2040 auto operating costs presented in the table do not include adjustments for autonomous and shared-use vehicles. Once the adjustments for autonomous and shared-use vehicles are accounted for, the year 2040 auto operating cost ranges from $0.12 per mile to $0.38 per mile with a most likely of $0.23 per mile.

3.4 High-Speed Rail Fares

The base average HSR fare for the Northern California to Southern California market was originally set at 83 percent of airfares for that market. A fare model with boarding and mileage fare components was developed to determine station-to-station fares reflective of the 83 percent standard. Maximum HSR fares were later capped at $93 in June 2017 dollars. This fare structure is assumed to be the most likely HSR fare scenario for each forecast year in the risk analysis.

Conventional rail fares and airfares were used to bracket the HSR fares, with the conventional fares used to guide the development of the minimum fare values, and the airfares were used to guide the development of the maximum HSR fares. The percent difference between the HSR fares and an average of a recent Bay Area to LA Basin airfares is used to set the maximum HSR fares (i.e., 42 percent change or 1.42 factor difference from Base fares). The percent difference between base HSR fares and an average of the Bay Area to LA Basin conventional rail fares is used to set the minimum HSR fares (i.e., -26 percent change or 0.74 factor difference from Base fares). This market was chosen since it is the market originally used to set HSR fares, and because it is the largest travel market for the Phase 1 HSR system. Table 3.4 shows the range in HSR fares for the terminus-to-terminus interchanges for each operating plan. HSR fares use a triangular distribution, with the most likely fares set as the base fares.

Table 3.4 Range of Terminus-to-Terminus High Speed Rail Fares
(June 2017 Dollars; Rounded to Nearest Dollar)

Alternative	Origin Station	Destination Station	Minimum	Most Likely	Maximum
Silicon Valley to Central Valley	San Francisco	Bakersfield	$69	$93	$133
Phase 1	San Francisco	Anaheim	$69	$93	$133
Phase 1	Merced	Anaheim	$69	$93	$133

3.5 High-Speed Rail Frequency of Service

The number of roundtrip HSR trains in service during operations may vary from the forecasted service levels. The most likely scenario matches the current planned levels of service in the base model runs. The minimum is based on the absolute least amount of service that could be expected to be run once the system is constructed. The maximum service frequency is based on the maximum amount of service that could be expected to run on a Silicon Valley to Central Valley line system for year 2029 and a level of service that approaches maximum track capacity, subject to a flexible service plan, for year 2033 and year 2040. Table 3.5 shows the range in trains per day for each forecast year. HSR frequency of service uses a triangular distribution.

Table 3.5 Range in High Speed Rail Frequency of Service

Forecast Year	Minimum[a] (Roundtrips/Day)	Base/Most Likely (Roundtrips/Day)	Maximum (Roundtrips/Day)
2029 (Silicon Valley to Central Valley)	14	22	76
2033 & 2040 (Phase 1)	44	98	152

[a] For comparison, in 2017 the Capitol Corridor ran 15 roundtrips per day.

3.6 Availability and Frequency of Service of Conventional Rail and High-Speed Rail Buses

This risk variable is only considered for the year 2029 Silicon Valley to Central Valley line scenario because these connections have less of an impact once the Phase 1 system is completed. The availability and frequency of service of CVR and HSR buses are discrete variables that consider the presence, or lack of, specific improvements to connecting rail services and HSR bus connections. The variable is composed of three potential future scenarios (1, 2, and 3) with a probability assigned to each scenario. Only one of the three scenarios is chosen in each draw of the Monte Carlo simulation. The scenarios and respective probabilities are as follows:

- **Scenario 1** (5 percent):
  - The Caltrain Electrification project *is* complete (66 trains per day with terminus at San Francisco Transbay).
  - No service enhancements above 2017 levels on the San Joaquin Line and Capitol Corridor Line.
  - No HSR connecting buses are provided.
- **Scenario 2** (40 percent):
  - The Caltrain Electrification project *is* complete (66 trains per day with terminus at San Francisco Transbay).
  - San Joaquin and Capitol Corridor Route set at 2017 State Rail Plan (SRP) Conservative Plan with Service Frequency set to ¾ between Current Operation and Conservative Plan, but with Sacramento – Madera Line extending all the way down to Bakersfield to continue providing local service within the San Joaquin Valley.
  - About 75 percent of the originally planned HSR buses are in service to meet HSR trains.
- **Scenario 3** (55 percent): **Same as Base Case.**
  - The Caltrain Electrification project *is* complete (66 trains per day with terminus at San Francisco Transbay).
  - San Joaquin and Capitol Corridor Lines set at 2017 SRP Conservative Plan.
  - Full set of HSR connecting buses is provided.

3.7 Coefficient on Transit Access-Egress Time/Auto Distance Variable

Between some regions in California, especially during Silicon Valley to Central Valley operations, individuals who wish to travel primarily by transit to reach their destination must transfer from an HSR bus or the CVR system before or after traveling on HSR. International experience has shown that there is uncertainty around how the need to make these transfers affects overall HSR ridership. The uncertainty in the impact of transfers can have a significant impact on ridership and revenue, especially when the CVR or HSR bus leg of the journey is relatively long in relation to the HSR travel length. Thus, this uncertainty was included as a risk variable.

The transit transfer uncertainty is addressed by varying the range for the parameters associated with transit access/egress travel times relative to origin-destination (OD) distances variable. This variable appears in the access and egress modal utility functions as follows:

$$\boldsymbol{\beta}\times\max\left(0,\frac{[Acc\ or\ Egr\ Time]}{[OD\ Distance]}-Threshold\right)$$

In the base model, several threshold parameter options were tested in model estimation, and a value of 0.2 was ultimately identified. The values of beta (the variable coefficient) were estimated directly, and were found to be negative. Separate coefficients were estimated for auto access/egress modes versus non-auto access/egress modes (transit and walk/bike), with the magnitude of auto coefficients estimated to be much larger. This variable essentially provides a disincentive for selecting a main mode that requires a long access or egress time, relative to the entire trip length. The uncertainty associated with the variable is only applied for the HSR main mode (i.e., not air or CVR).

There are few comparable examples of these kinds of transfers in HSR systems around the world, but a French rail linkage was identified to serve as a guide. In the French experience, moving from a direct CVR connection between Paris and Grenoble to an HSR trip from Paris to Lyon and a connection to CVR from Lyon to Grenoble saved 90 minutes of total travel time, but did not result in increased ridership. The observed “90-minute penalty” in France served as a rough benchmark for determining a lower bound on the model parameters.

Appendix E details the process taken to develop the minimum parameter values for this variable. The minimum threshold value is set to 0.1, since a lower threshold would start to impact local transit access and other unrelated trips. The minimum coefficient value is set to -2.0 for business/commute purpose and -1.3 for recreation/other purpose. These are set to achieve penalty values of 51 and 66 minutes. These penalty value benchmarks come from the penalties the model suggests for the French scenario for drive access/egress modes. The lower bound on the transit penalty should not exceed the penalty suggested by the model for drive access/egress modes. A 51-minute and 66-minute penalties were used, instead of the 90-minute penalty observed in the French experience, because it offered more reasonable model behavior overall, and it was not desirable to change the long-distance models in unreasonable ways to match a single observed data point. The coefficient and threshold values vary in parallel (i.e., perfect correlation) for the full model runs and Monte Carlo simulation.

The maximum threshold and coefficient values are set to be identical to the calibrated base/most likely values since there is no evidence to suggest that the penalty to transfer from transit to HSR should be less than the penalty used for CVR and Air that was developed based on observed data. A PERT distribution was used for this variable.

3.8 Airfares

Airfares are considered as a risk variable for 2033 and 2040. The airfare uncertainty is based on the variability in airfares from 2009 to 2016 from routes that serve the major airports of the Northern California-Southern California market. Mean, minimum, and maximum annual weighted airfares by route were calculated for each year between 2009 and 2016 using the BTS Transtats Airline Origin and Destination Survey (DB1BMarket) fare data. Since, the base airfares (i.e., year 2009) represent the lowest point from the range analyzed, the base fares were set as the minimum value. The most likely value was set as the decimal factor difference from the base fare and the average of the calculated mean airfares across the analyzed routes (i.e., 15 percent higher airfares compared to the base fares). The maximum value was set as the decimal factor difference from the base fare and the average of the calculated maximum airfares across the analyzed routes (i.e., 31 percent higher airfares compared to the base fares). A triangular distribution was used for this variable.

3.9 Number and Distribution of Households throughout the State

The number of households (stratified by household size, income group, number of workers, and number of autos owned) is a key driver of the amount of long-distance travel in the BPM-V3 model. Statewide forecasts of population, households, and employment were assembled from various sources, as shown in Figure 3.3, Figure 3.4, and Figure 3.5. The four independent forecasting sources along with the year they were produced were the following:

- The California Economic Forecast (CEF) for the California Department of Transportation (Caltrans) Transportation Economics Branch (2016).
- Moody Analytics (2016).
- Metropolitan Planning Organization (MPO) data (assembled by CS from plans available through April 2017).
- California Department of Finance (DOF), Demographic Research Unit (Baseline 2016).

The figures also show the forecasts of population, households, and employment used for the 2016 Business Plan, along with averages from the sources listed above. Only population data were obtained from the DOF.

As shown in Figure 3.3 and Figure 3.5, population and employment forecasts are consistent across the sources, while there is a wide variation in the forecasts of households for the state (Figure 3.4). The differences in the household distributions are driven by different assumptions regarding future household sizes; the Moody Analytics forecasts are based on an assumption of decreasing average household sizes over time, while the CEF data are based on an assumption that average household sizes will increase over time.

Figure 3.3 Statewide Forecasts of Population by Source of Forecast

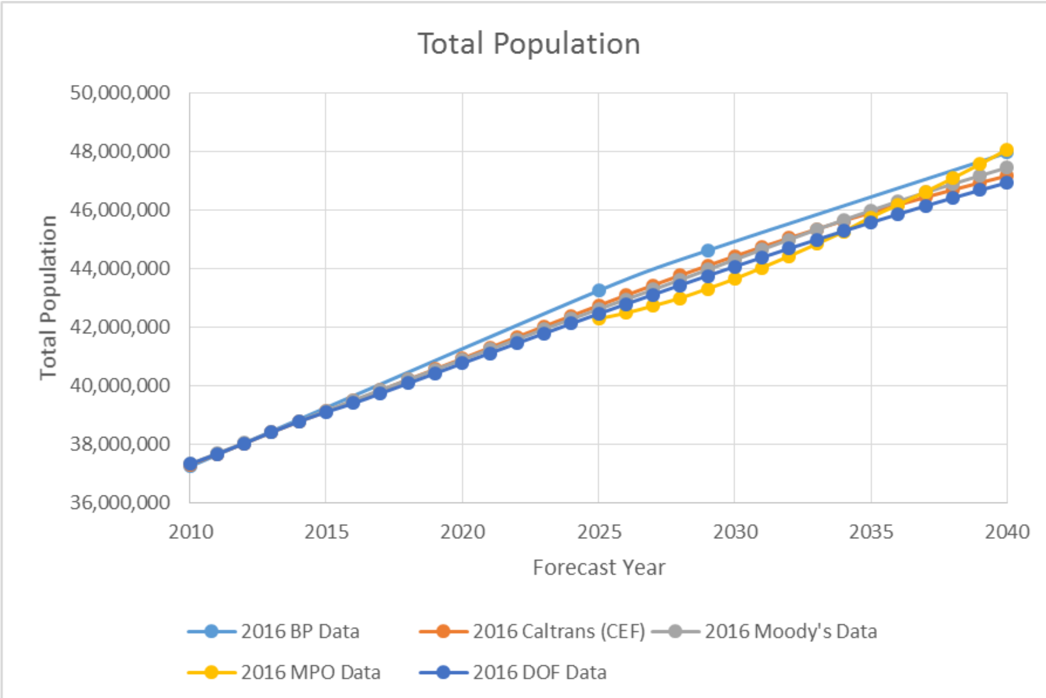


Figure 3.4 Statewide Forecasts of Households by Source of Forecast

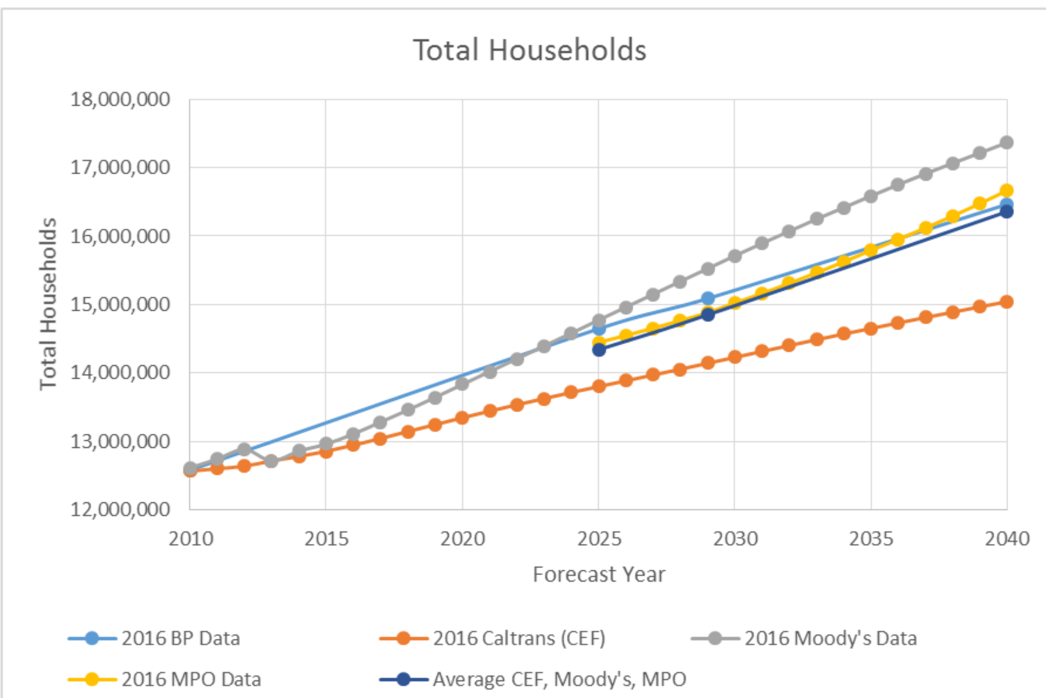
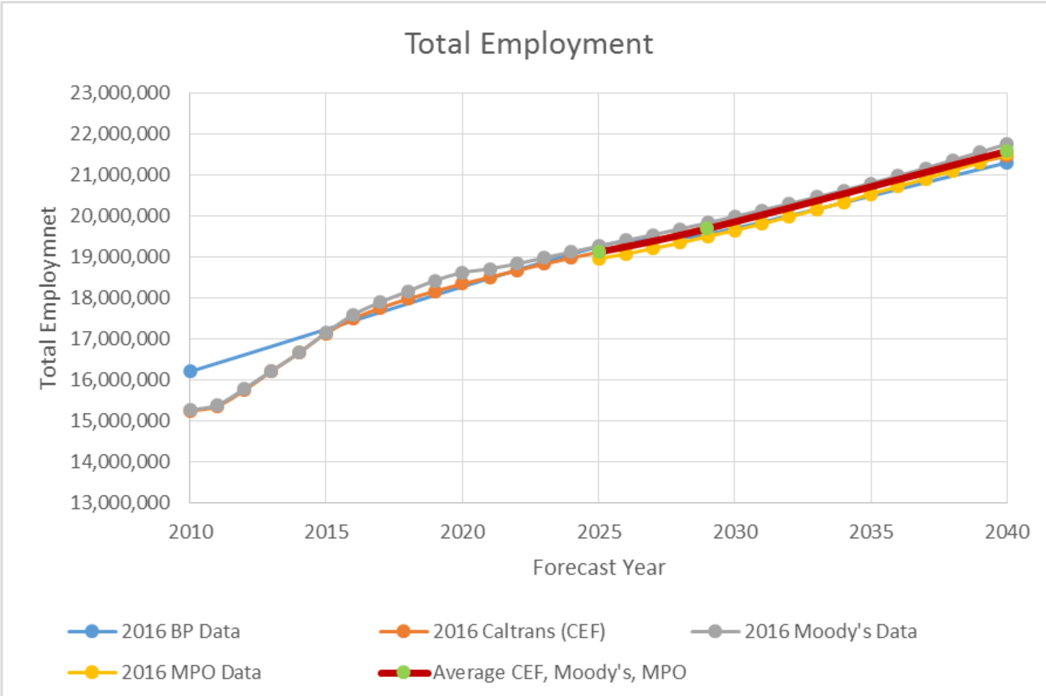


Figure 3.5 Statewide Forecasts of Employment by Source of Forecast



For the risk analysis, minimum, most likely (base case), and maximum forecasts of population, households and employment are outlined in Table 3.6. Since the minimum, most likely, and maximum population forecasts are relatively similar, but household forecasts vary, there are implied differences in average household sizes for the minimum, most likely, and maximum forecasts. Likewise, there are implied differences in the numbers of workers per household since, like population, the minimum, most likely and maximum employment forecasts are relatively similar. The implied differences in average household sizes and numbers of workers per household have been reflected by varying the distributions of households by household size and households by number of workers in the input socioeconomic datasets used by the BPM-V3.

Table 3.6 2040 Statewide Population, Household, and Employment Forecasts

Forecast	Source of Forecast	Population	Households	Employment
Maximum	• Moody Analytics forecasts	47,164,732	17,366,581	21,750,362
Most Likely (Base Case)	• Averages of Moody's, CEF, and MPO data for Households and Employment • Average of Moody's, CEF, MPO, and DOF data for Population	47,399,849	16,359,211	21,574,851
Minimum	• CEF forecasts	47,443,634	15,044,342	21,461,166

The Moody Analytics, CEF, DOF, and MPO forecasts were provided on a county-by-county basis, and disaggregated to transportation analysis zones based on detailed MPO forecasts. County level estimates for any given risk analysis run were developed using a draw from a uniform distribution ranging between –1 and +1, by creating a weighted combination of the base case and either the minimum (for negative draws) or maximum (for positive draws) distributions. For example, if a run calls for a population based on a random draw of -0.4, then the input data for the BPM-V3 model run performed in Step 6 of the risk analysis would be the weighted combination of 40 percent of the population, households, and employment in the minimum distribution; and 60 percent of the population, households, and employment in base case distribution (i.e., the remaining weight). Alternatively, if a run calls for a population based on a random draw of +0.9, then the input would be the weighted combination of 90 percent of the population, households, and employment in the maximum distribution; and 10 percent of the population, households, and employment in base case distribution.

3.10 Auto In-Vehicle Time Coefficient

The auto in-vehicle time coefficient is considered as a risk variable for 2040. By 2040, it is likely that autonomous vehicles (AV) will compose a nontrivial share of all automobile travel. Because travelers will be able to engage in other activities in AVs (e.g., checking email, reading, or even sleeping), AVs offer the possibility that being in one's car may be less onerous than it is today. This will be considered in the risk analysis by adjusting the in-vehicle time (IVT) coefficient associated with the auto mode. For the purposes of selecting a range for the IVT coefficient, the risk analysis considers the factor by which the IVT coefficient would be multiplied. A factor of 1.0 would indicate no change, while a factor of 0.9 would correspond to a reduction in traveler sensitivity to auto IVT.

The Federal Transit Administration (FTA) offers informal guidance on IVT bonuses to use in mode choice models for fixed guideway transit modes (relative to bus or auto) in the context of regional travel models. Such modes include commuter rail, light rail, subway, etc. The FTA's guidance suggests that, under optimal conditions, the IVT factor could be as low as 0.75. The considerations in determining what factor to select include seat availability, travel time reliability, ride quality, and vehicle amenities. There are a couple of key considerations in making comparison with the FTA guidance. First, the guidance is specific to regional travel, so it is not clear whether it is transferable to long-distance travel. Second, in terms of the actual variables mentioned by FTA impacting the selection of a factor, it is not clear that any apply in a comparison of AVs to traditional autos. Nonetheless, the value of 0.75 provides a benchmark of what has worked in the context of transit. A 2017 study of travel time sensitivities across a variety of modes found that a passenger's sensitivity to travel time was about 30 percent less than that of the driver (implying a factor of about 0.70).[9] As noted below, a number of studies have considered how AVs will impact traveler perception of travel time:

[9] Litman, T. 2017. Transportation Cost and Benefit Analysis II – Travel Time Costs, Victoria Transport Policy Institute (VTPI).

(Footnote continued on next page...)

- A 2015 study for the Seattle region considered IVT factors of 1.00 and 0.65 and concluded that the value of 0.65 was best based on the region's travel demand model for the commuter rail mode in the mode choice model (which was developed based on observed data).[10]
- A 2016 scenario analysis in Germany and the U.S. assumed a value of 0.75, but also provided bounds of 0.5 and 1.0.[11] One important distinction made in this study was that it assumed no impact on the first 10 minutes of travel time for a trip (i.e., 1.0 factor for the first 10 minutes).
- IVT factors ranging from 0.5 to 1.0 (in increments of 0.125) were assumed across 24 scenarios in a 2016 study of long-distance travel made by Michigan travelers.[12]
- A 2017 scenario-based approach in the Austin region assumed a mean IVT factor for AVs equal to 0.5 with a range from 0.25 to 0.75.[13]
- A 2016 survey of 45 modeling experts from around the world regarding the appropriate IVT factor to use for AVs in travel models relative to the factor for traditional autos resulted in a suggested average factor from 0.8 to 0.9, with a standard deviation of 0.2, depending on trip purpose.[14]

The research above indicates a clear upper bound: the sensitivity to travel time would be unaffected by the introduction of AVs. This would correspond to a value of 1.0.

The lower bound and base values are more difficult to identify. On the one hand, IVT factors in the range of about 0.75 are commonly accepted for premium transit modes. Experts seem to believe a value close to 0.8 or 0.9 would be appropriate in the context of AVs relative to traditional autos. However, for long-distance travel, it might be reasonable for the bonus given to AVs to be even larger, as it might be easier to engage more fully in non-driving activities with a larger amount of time to allocate.

We recommend a triangular distribution with base of 0.75 and bounds of 0.5 and 1.0. The triangular distribution reflects that the likelihood of the auto IVT coefficient at the upper or lower bound values is unlikely. There will be some reduction in IVT coefficient to account for the impact of autonomous vehicles; the base value will be 0.75 of the estimated for the BPM-V3 coefficient.

[10] Childress, S., B. Nichols, B. Charlton, and S. Coe. 2015. Using an Activity-Based Model to Explore Possible Impacts of Automated Vehicles, *Transportation Research Record*, 2493, 99-106.

[11] Kroger, L., T. Kuhnimhof, and S. Trommer. 2016. Modelling the Impact of Automated Driving – Private Autonomous Vehicle Scenarios for Germany and the U.S., Proceedings of the 2016 European Transport Conference, Barcelona, Spain.

[12] LaMondia, J., D. Fagnant, H. Qu, J. Barrett, and K. Kockelman. 2016. Long-Distance Travel Mode Shifts Due to Autonomous Vehicles: A Statewide Mode-Shift Simulation Experiment and Travel Survey Analysis, *Transportation Research Record*, 2566, 1-11.

[13] Zhou, Y., and K. Kockelman. 2017. Anticipating the Regional Impacts of Connected and Automated Vehicle Travel in Austin, Texas, Proceedings of the 96th Annual Meeting of the Transportation Research Board (TRB), Washington, D.C.

[14] Kohli, S., and L. Willumsen. 2016. Traffic Forecasting and Autonomous Vehicles. Proceedings of the 2016 European Transport Conference, Barcelona, Spain.

To account for the fact that there is effectively no change in the types of activities a passenger in an auto can pursue, we made an adjustment to the range for group travelers. Effectively, the range above suggests a discount on the IVT coefficient of between 0.0 and 0.5. To reflect the passenger component for shared-ride travel, the range will be divided by the average group size. Then, the effect on the IVT coefficient range for shared-ride travel will be to replace 0.5 and 0.75 in the triangular distribution above with values equal to [1- (0. 5/group size)] and [1 – (0.75/group size)]. The assumption of a triangular distribution will remain the same. Table 3.7 shows the range of factor applied to the IVT coefficient for alone and group travelers.

Table 3.7 Factor Applied to In-Vehicle Time Coefficient

Risk Variable	Minimum	Most Likely	Maximum
Alone Travelers	0.50	0.75	1.0
Group Travelers	0.80	0.90	1.0

Since all autos are treated by the existing model as a single mode, there is no way to model the impact for AVs differently than non-AVs. To account for this, if we assume AVs provide a 10-percent reduction in IVT coefficient and our scenario suggests that 50 percent of autos are AVs, we would model this as a 5-percent reduction in IVT coefficient for all autos. The range in market penetration of AVs for year 2040 is discussed in Appendix F.

3.11 High-Speed Rail Reliability

Experience from international systems suggests that high-speed rail will arrive within 15 minutes of the scheduled arrival time 99 percent of the time. There is a risk that this assumption is too high, particularly for a new system in the United States and one with final alignments still to be confirmed.

The shared right-of-way with Caltrain in the Bay Area creates a risk that performance will degrade over this corridor, but Caltrain's current performance makes it likely that the 99-percent reliability target will be achieved. A 2011 report regarding Caltrain on-time performance showed a very high reliability of 98.2 percent over the December 2010 to May 2011 timeframe, based on the BPM-V3 definition of reliability (percent of trains arriving within 15 minutes of scheduled time). An unofficial 2016 review of Caltrain API[15] data by Silicon Valley Data Science suggests that Caltrain reliability may have improved since then. Their analysis of 10,918 local train and 9,667 limited train departure delays from stations suggests virtually no delays greater than 6 minutes. These results suggest that operations issues in the Caltrain corridor are unlikely to increase possible delays to high-speed rail.

The high-speed rail reliability risk variable will be set at 90 percent as the minimum value, 99 percent as the most likely value, and 99.7 percent as the maximum value with a PERT – standard distribution. This range and distribution assumption concentrates the reliability around the 99 percent most likely value, but

[15] Silicon Valley Data Science, "Analyzing Caltrain Delays: What We Can Learn," March 10, 2016, https://svds.com/the-trains-project-analyzing-caltrain-delays/.

allows for the unlikely possibility that the high-speed rail system will have reliability more traditionally experienced in CVR systems.

3.12 Exceptionally Long Access/Egress to High-Speed Rail

Project Finance Advisory Limited (PFAL) provided an independent review of the BPM-V3 ridership and revenue forecasts developed for the 2016 Business Plan and identified a risk that the BPM-V3 model may be over-estimating the usage of high-speed rail on trips that have an exceptionally long access or egress leg. This was mostly a concern in the model for the Silicon Valley to Central Valley line as defined in the 2016 Business Plan, where some origins and destination within California are quite distant from any HSR station.

Reliably estimating parameters for exceptionally long access and egress from our available survey data is very difficult, because so few observed trips have these attributes, and they are generally correlated: if your origin is very far from an airport, you also are usually very far from a train station, and vice versa. This will not necessarily be the case for high-speed rail; many places are far from high-speed rail, but close to Air or CVR stations, especially during the Silicon Valley to Central Valley phase. For example, conventional rail connects San Diego to the San Diego International Airport, but the closest HSR station in Bakersfield will be over 200 miles away during a Silicon Valley to Central Valley Line. Thus, the exceptionally long access/egress to high-speed rail risk variable addresses the uncertainty in the attractiveness of high-speed rail when accessing or egressing high-speed rail involves very long travel times. From our estimation data, we observe almost no access or egress trips to CVR or Air that exceed three hours in travel time, so there is no way to rigorously estimate model coefficients that affect such long access or egress travel differently than shorter access or egress. As a result, the BPM-V3 model may forecast trips with these characteristics.

These exceptionally long access and egress trips are particularly concerning when made using taxi, although existing model parameters already make taxi trips very uncompetitive at long distances. Of 684 access or egress taxi trip observations from our model estimation data, only 10 were over 50 miles, and none was over 70 miles. Although there are very few high-speed rail trips where taxi is used as an access/egress modes for a very long distance, evidence suggests that these trips should not exist at all.

Similarly, when the access or egress trip is made via pick up or drop off, the implication is generally that a friend or relative will drive a deadhead leg of the trip roughly equivalent to the overall access or egress distance. However, this becomes more difficult when the access or egress leg is exceptionally long.

For further guidance on these trips we researched international high-speed rail systems, which is discussed in Appendix G.

The exceptionally long access/egress to high-speed rail risk variable is composed of a set of penalties that are added to the access/egress mode choice utilities to limit long-distance trips that rely on exceptionally long access/egress:

- A Taxi Cost penalty for long access/egress is added; whereby, taxi costs double after 50 miles.[16]
- Disutility of Pick Up/Drop Off IVT increases by 0 to 150 percent after two hours.
- Disutility of Pick Up/Drop Off Cost increases by 0 to 150 percent after 90 miles.[17]
- Disutility of Access and Egress IVT by Auto (i.e., Drive and Park, Rental, Pick Up/Drop Off) increases by 0 to 150 percent after three hours, and again by a further 0 to 150 percent after four hours.[18]
- Disutility of Access and Egress IVT by Transit (sum of transit time and auto-to-transit time) increases by 0 to 150 percent after four hours.

All components vary together, through an index that has a uniform distribution with a minimum value of 0 percent and a maximum value of 150 percent. The minimum value represents no penalty on long access and egress, other than as noted above for Taxi, and is equivalent to the BPM-V3 base run forecast. We adopt the uniform distribution because this risk factor is based on behaviors that are not observed in our model estimation data nor contemplated in the stated preference survey, so we are unable to make reasonable estimates based on data. Second, as this risk factor represents an exclusively "down-side" risk, adopting a uniform distribution represents a conservative approach.

3.13 Visitor Travel

The BPM-V3 ridership and revenue forecasts only include intra-state travel made by residents of California. However, over 60 million people visit California each year; some of whom may choose to travel via high-speed rail. Using professional judgement, PFAL reported, with 90 percent confidence, that visitors to California may increase ridership by 5 to 10 percent for each operating phase. In 2029, the Silicon Valley to Central Valley high-speed rail will have just opened and will not include a direct connection to the Southern California region. Since it is likely that visitors to California will be more likely to visit sites in the Bay Area and Southern California regions, the increase in ridership for the Silicon Valley to Central Valley phase resulting from visitors was reduced to a range of 3.5 to 7 percent from the PFAL prediction of 5 to 10 percent. **Error! Reference source not found.** shows the PFAL estimate of high-speed rail visitor trips for Phase 1 and the reduced estimate of high-speed rail trips made by visitors for the Silicon Valley to Central Valley phase.

[16] Of 684 access or egress taxi trip observations from our model estimation data, over 99 percent were under 50 miles.

[17] For valid observed Air and CVR trips in the mode choice model estimation data made with Pick Up/Drop Off as the access mode, over 98 percent had access times under two hours and access distance under 90 miles.

[18] For valid observed Air and CVR trips in the mode choice model estimation data, over 99.5 percent had access times under three hours.

Table 3.8 Range in High-Speed Rail Visitor Trips

	Minimum (based on 90% confidence)	Median	Maximum (based on 90% confidence)
Year 2029 Silicon Valley to Central Valley			
Percentage Increase in Base Run Ridership	3.5%	5.0%	7.0%
High-Speed Rail Trips by Visitors (millions)	0.50	0.72	1.00
Year 2033 Phase 1			
Percentage Increase in Base Run Ridership	5.0%	7.5%	10.0%
High-Speed Rail Trips by Visitors (millions)	1.78	2.67	3.56
Year 2040 Phase 1			
Percentage Increase in Base Run Ridership	5.0%	7.5%	10.0%
High-Speed Rail Trips by Visitors (millions)	1.97	2.95	3.93

For the risk analysis, we apply a uniform distribution to the ranges discussed above. Since the PFAL range was developed within a 90-percent confidence range, the risk analysis range was widened to represent 100 percent probability of occurrence by adding +/- [(Maximum Visitor Ridership –Minimum Visitor Ridership)*.05] to each year. The risk analysis minimum, median, and maximum visitor ridership for each year are shown in **Error! Reference source not found.**.

Table 3.9 Visitor Travel High-Speed Rail Ridership *(Millions)*

Risk Variable	Minimum	Mid-point (PFAL Median)	Maximum
Year 2029 VtoV	0.48	0.72	1.03
Year 2033 Ph1	1.69	2.67	3.65
Year 2040 Ph1	1.87	2.95	4.03

We assume a 50-percent positive correlation between visitor travel high-speed rail ridership and total California resident high-speed rail ridership. Some risk analysis factors that contribute to lower high-speed rail resident ridership will also affect visitor ridership. However, there are some risk analysis variables that will only affect resident travel, or will affect resident travel differently from visitor travel. It is assumed that the geographic distribution of visitor high-speed rail ridership (i.e., the distribution of station-to-station boardings) is identical to resident high-speed rail ridership, and thus revenue can be calculated by multiplying ridership by average fare of resident travel for each simulation run. This risk factor results in an increase in high-speed rail ridership and revenue regardless of where it falls on the minimum to maximum range of values.

3.14 Induced Travel

The BPM-V3 forecasts approximately 4 percent of the high-speed rail ridership are induced trips. Induced trips are new long-distance trips made by households due to the introduction of high-speed rail. For example, with the introduction of high-speed rail, households can now travel from Fresno to San Jose in ~60 minutes via high-speed rail that before took ~150 minutes by car. This significant decrease in travel time may result in households taking more trips to San Jose on top of all other long-distance trips they took before. These induced trips may be replacing short-distance trips (e.g., a shopping trip made within Fresno is replaced by a shopping trip made in San Jose); or may be due to a household increasing the number of overall trips they make (e.g., a new work trip to San Jose to attend a work meeting in person rather than taking the meeting via phone).

PFAL, as part of their independent review of the 2016 Business Plan high-speed rail ridership, believed with 90 percent confidence that induced travel represents 0 to 20 percent additional ridership (with a most likely value of 10 percent) on HSR systems. It is our judgement that this assessment is on the high-side; and to be conservative, we have set the additional induced travel range for the risk analysis at 0 to 15 percent, with a most likely value of 7.5 percent. A triangular distribution is assumed.

The same characteristics that drive uncertainty in the BPM-V3 results drive induced travel uncertainty. Thus, the additional induced travel high-speed rail ridership for each simulation run is calculated by multiplying the induced travel percentage by the resident high-speed rail ridership for each simulation run. This results in lower additional induced travel when overall high-speed rail ridership is lower. In addition, we assume a 50 percent negative correlation with the trip frequency constant. This assumption is based on the fact that there is a finite number of total trips that households will make. The more trips accounted for in the BPM-V3 model, the less induced trips that can potentially be made. With the exception of Monte Carlo simulations where 0 percent additional induced travel is selected, this risk factor results in an increase in high-speed rail ridership and revenue regardless of where it falls on the minimum to maximum range of values.

4.0 Implementation of Risk Analysis

Mathematical models such as the BPM-V3 provide a simplified understanding of causal processes affected by thousands of variables to provide shortcuts to predicting their outcomes, turning a complex reality into a streamlined process. However, as models more closely approach reality with highly disaggregate results (e.g., the number of travelers and their mode choices between all transportation analysis zones in California), they become more complex and computationally intensive. While the full BPM-V3 model provides highly disaggregate results, it takes hours to run.[19]

There are inherent uncertainties associated with estimated parameters used in the BPM-V3 and the model input data; consequently, uncertainty surrounds model output as well. In order to assess the likelihoods of achieving different levels of ridership and revenue, multiple applications of the BPM-V3 with varied values for the parameters and model inputs are used to provide ranges of future ridership and revenue levels as well as the probability of achieving those levels.

The risk analysis for the 2018 Business Plan focuses on revenue and ridership totals and requires thousands of data points considering the numbers of risk factors and risk variables described in Section 3. It is not feasible to run thousands of full BPM-V3 model simulations, but by building models of the model forecasts of ridership or revenue based on the results of 100 to 200 applications of the BPM-V3 with varied inputs for the identified risk variables for each forecast year, it is possible to predict the revenue or ridership totals that would be forecast by a full model run in a fraction of a second using a Monte Carlo simulation. These meta-models provide effective and efficient tools for performing the risk analysis.

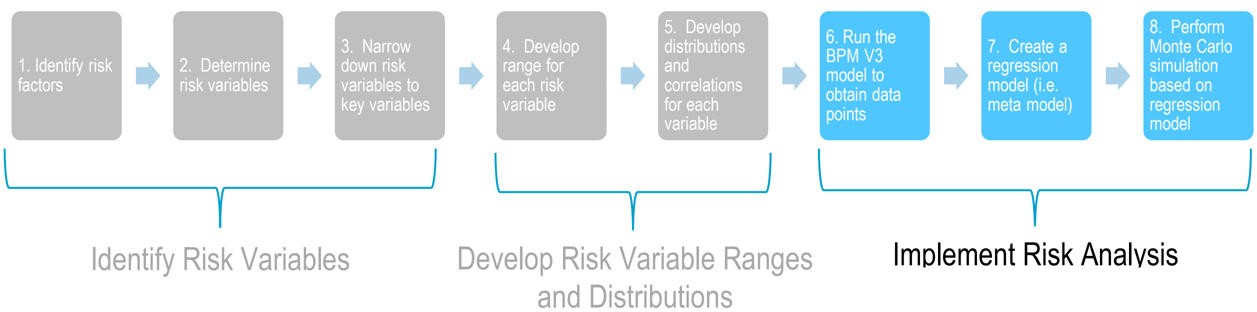
As shown in Figure 4.1, the risk analysis is implemented in three steps. Each meta-model is developed from a set of full BPM-V3 runs (*Step 6*). The independent variables used to develop the meta-model are the risk analysis variables, and the dependent variables are either HSR revenue or ridership. Each full BPM-V3 model run provides one data point for use in estimating the regression equations (*Step 7*). Monte Carlo simulations with 100,000 draws each for each forecast year are then run using the ridership and revenue meta-models. The Monte Carlo simulations use different combinations of values for the risk variables, with the values drawn from the assigned risk variable distributions (*Step 8*). The ridership and revenue outputs from these runs are then used to develop the ranges of ridership and revenue along with their probabilities of occurrence.

As noted in Section 1, the 2018 Business Plan risk analysis has built on the procedures used for the 2016 Business Plan risk analysis by enhancing the meta-models developed in *Step 7*. While linear regression is still used to develop initial meta-models, a Gaussian Process Regression (GPR) approach is also used to refine the linear regression models for the final meta-models. The remainder of Section 4 is structured as follows:

[19] It takes approximately 12 hours to run the BPM-V3 model using a "one-thread set-up." It takes one hour to run the BPM-V3 model using a "12-thread set-up," which is the maximum possible threads that can be run on one standard computer.

- Section 4.1 describes the derivation of the meta-models including the linear regression and GPR enhancements
- Section 4.2 describes the approach for developing the data used to estimate the meta-models
- Sections 4.3 and 4.4 describe the resulting revenue and ridership meta-models for each forecast year
- Sections 4.5 and 4.6 summarize the revenue and ridership risk analysis results for each forecast year
- Section 4.7 discusses the relative importance of the various risk variables in the revenue and ridership forecasts

Figure 4.1 Eight-Step Risk Analysis Approach: Implement Risk Analysis (Steps 6 to 8)



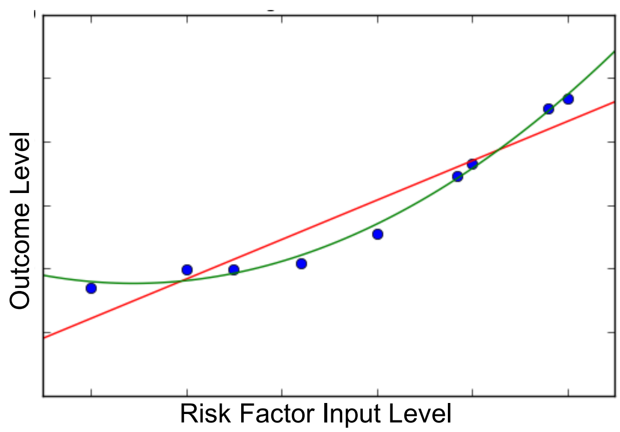
4.1 Meta-Model Derivation

Just as travel demand models (TDM), such as the BPM-V3, are simplified mathematical abstractions of the very complex "real" travel behavior processes they represent, meta-models are, in turn, simplified representations of TDMs. Meta-model runs can be produced much more quickly than full TDM models since they produce only a single headline number (e.g., total ridership) using a limited number of key input variables. In contrast, TDMs produce detailed, trip-level information (e.g., interchange mode shares or station-level boardings and alightings) from detailed input datasets with millions of data points. While the underlying TDM model may require hours or days to generate outcomes from one set of inputs, a meta-model typically can make a forecast of the headline number in a fraction of a second. This facilitates the generation of outputs for thousands or millions of different combinations of input variables through the use of the meta-model in a Monte Carlo simulation. This approach allows for the estimation of the range and probability distribution of outcomes and the analysis of the relationships between the various inputs and headline forecasts that would be produced by the TDM.

The selection of the mathematical form of the meta-model is an important part of meta-model development. Because the speed of meta-model computation is critical, it is common to employ a linear regression (LR) model form, as this kind of model offers exceptionally fast computational performance for generating point forecasts for each new set of inputs. However, the adoption of a LR meta-model does come with some limitations. Most importantly, the use of LR requires the modeler to develop a defined functional form of the relationship between the inputs and output. The parameters of the defined function are set to provide the best fitting model for the observed data but, if the defined functional form does not fit the data well, the meta-model may not provide accurate results.

Figure 4.2 provides an illustrative example of such a limitation with LR models in one dimension. Nine observations (blue circles) of an outcome (e.g., ridership or revenue output by the BPM-V3) are plotted on the Y axis, with input variable levels plotted on the X axis. A simple LR (slope + intercept) is constructed, shown by the red line. This model fits the data only moderately well. Alternatively, a second order polynomial LR, shown in green, fits better, although still not perfectly. Increasingly complex LR model functions can be created to iteratively improve the fit of the LR by including various alternative linear transformations and interactions of explanatory variables. However, in doing so, there is a risk of "overfitting" the LR model by too closely matching the observed data points and degrading the quality of the LR forecast for new points.

Figure 4.2 A Linear Regression Example in One Dimension



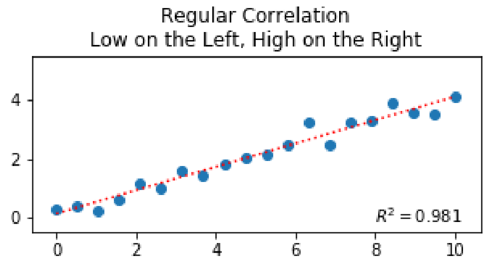
Source: Cambridge Systematics

The figure also illustrates another important shortcoming of the LR model. The two observations (blue dots) in the center of the figure are both notably lower than the initial LR fitted line (red), and both also lower than the improved LR model (green). The LR model is predicated on the assumption that residuals (the deviation between the observed points and the fitted regression line) are identically and independently distributed (IID). If that were correct, each observation should be equally likely to be above or below the fitted regression line, regardless of whether other nearby points are above or below the line. However, when the underlying model is smooth and deterministic,[20] this IID assumption is violated: other nearby observations will tend to be high or low together, not independently. Such a relationship between observations is called auto-correlation (see Figure 4.3). The presence of auto-correlation provides an opportunity to create a meta-model that better captures the output data points than a meta-model based on simple LR.

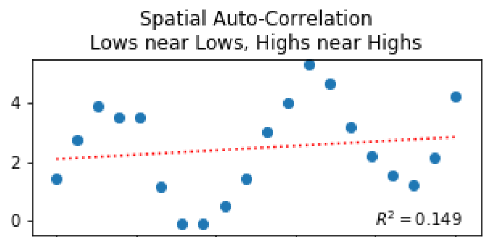
[20] A model is smooth if an infinitesimally small change in the inputs will produce an infinitesimally small change in outputs, and there are not sudden jumps in the output values when input values cross some threshold. A model is deterministic if it always produces exactly the same output value for the same set of input values, and there is no random variation in the outputs. The BPM-V3 model is a deterministic model.

Figure 4.3 Auto-Correlation

Auto-correlation occurs when two observations tend to be similar if they are close to each other, while observations that are far apart have lesser similarity. With typical correlation, the magnitude of the explanatory data is important, and all data points contribute equal information:



For auto-correlated data, high values tend to cluster around other high values, and lows with lows. More distant observations provide less information than nearby ones, and global trends may exist, but are not required to exist:



Source: Cambridge Systematics.

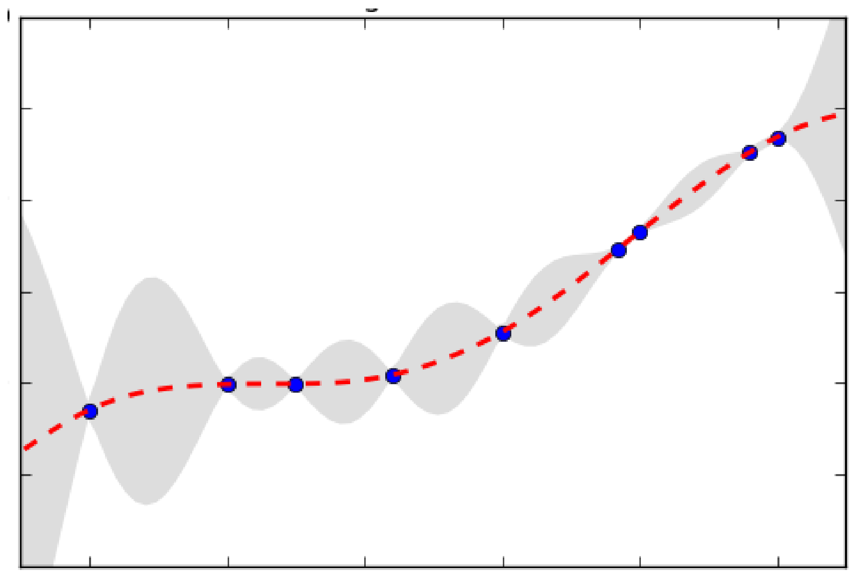
Gaussian Process Regression (GPR) is a nonparametric "machine learning" tool for regression analysis, which explicitly harnesses the auto-correlation property of the underlying main model. Like LR methods, GPR estimates the output value (i.e. the dependent variable) for any set of inputs; it is developed using the set of observed inputs and their observed outputs[21]. However, as a nonparametric model, GPR does not impose a restriction on the functional form of the output (e.g., it does not need to be linear or any particular defined nonlinear function). Instead, it is simply assumed that output is auto-correlated: if two observations have inputs that are similar, then the output should also be similar. To achieve this, a GPR computes the expected value of the output for any set of inputs as the weighted average of other observations where both the input and output are known, setting the weights higher for nearer observations and lower for more distant observations. In this context, the "distance" between two observations is not based on an actual physical distance, but rather an abstract measure of Euclidean distance across an *N*-dimensional space, where *N* is the number of distinct input variables being varied.

[21] An example for the risk analysis would be the forecast revenue based on different input auto operating costs.

Although GPR has not been widely used for transportation planning meta-model applications, it is widely used for computer simulation meta-models in other fields, and represents a reliable and well-documented approach for improving meta-model results in relationship to simpler, underlying models such as LR. Because of its desirable properties, discussed below, and the relative ease and speed of application, GPR is the state-of-the-practice approach for modern meta-models of computer experiments.[22]

The results of a GPR are illustrated in Figure 4.4, which constructs a GPR predictor on the same observations from the LR models contemplated in Figure 4.2. The expected value of the GPR at any point is shown by the dashed red line, and the distribution around the expectation (as measured by two standard deviations) is shown by the gray area. A few salient features of the GPR can be observed in Figure 4.4.

Figure 4.4 An Illustrative Gaussian Process Regression



Source: Cambridge Systematics.

Most notably, when GPR is applied as a meta-model for a deterministic travel demand model, there are no residuals in predictions for sampled observations used in model estimation. That is, the difference between the predicted value from the meta-model and the observed value is always zero. Instead, the expectation line always passes exactly through every observation point. This has implications for evaluating the model goodness of fit (discussed below). Moreover, because the simulation model (e.g., the BPM-V3 for the 2018 Business Plan risk analysis) is deterministic, there is also no variance in the predictions at those points. In Figure 4.4, this is reflected in the collapse of the height of the gray ribbon to zero at each observation.

An important consequence of this is that GPR models cannot be evaluated based on traditional "goodness of fit" measures (e.g., R^2) derived from the estimation data. Instead of measuring goodness of

[22] Santner, T. J., B. J. Williams, and W. I. Notz. 2013 *The design and analysis of computer experiments*, Springer Science & Business Media.

fit directly based on estimation data, it is necessary to measure fit on a validation data set that is not used for model estimation. Because it is usually expensive to collect additional validation data, it is preferred to conduct K-fold cross validation (CV). For this, the set of observations is randomly partitioned into K subsets (typically 5 to 10; for this risk analysis, we have used K = 10). The entire model is re-estimated using K-1 subsets of the data (leaving one out), and a model score is calculated by using the result to predict the outcomes on the remaining held-out subset of observations. The entire process is repeated iteratively holding out each of the K subsets one at a time, and then averaging the resulting scores. The CV score is interpreted in roughly the same manner as R^2 for LR models, such that a score of 1.0 indicates a perfect prediction, and a score of 0.0 is achieved by predicting the global mean of the dependent variable. When the GPR is applied to de-trended data (i.e., on top of a LR model), the resulting CV scores are calculated based on the residuals from the LR model, so they represent the relative improvement in fit over the LR result, in which case they provide insight into the value of the GPR process, but are not directly comparable to the R^2 values.

Another feature of the GPR is that it is not neatly summarized by a limited set of parameters, but instead the entire estimation dataset is explicitly incorporated into the model, and must be available to generate predictions for new points. This contrasts with LR models, where the estimation data is used to develop parameter values, but then only those parameters (and not the entire estimation dataset) are needed to generate new values. Fortunately, this requirement is not limiting for meta-modeling applications such as this, as the number of observations used is manageably small.

4.2 BPM-V3 Model Runs

An experimental design for model runs lays out the number of model runs needed to support the risk analysis and the combination of risk variable values that compose each model run. For a complex model such as BPM-V3, it is important to design experiments to provide data to the risk analysis in an efficient manner, as the computational cost of each individual experiment is high.

The BPM-V3 is a deterministic simulation model whose meta-models are best supported by a "space filling" design of experiments, such as Latin hypercube draws.[23] A Latin hypercube sample for one dimension is constructed by subdividing the distribution of each input factor into *N* equally probable ranges, and drawing one random sample within each range. For example, if an input factor is assumed to be uniformly distributed between 0 and 100, that distribution can be divided into four regions (0-25, 25-50, 50-75, and 75-100), and one random draw can be made in each region. This ensures better coverage of the entire input range than making four random draws from the full 0-100 range, which could easily result in a cluster of observations in one part of the range and a large void elsewhere. In other words, for the high-speed rail risk analysis, this approach ensures that the random draws for the risk variables used in the full BPM-V3 runs used to generate the data for the estimation of the meta-model are not clustered around the minimum, most likely, or maximum values for any of the variables. Generating a multidimensional Latin Hypercube sample for use with multiple input variables follows this same basic technique. However, the various draws in each dimension are randomly reordered before being joined

[23] Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. Design and analysis of computer experiments, Statistical Science, 409-423.

with draws from other dimensions to avoid unintended correlation (such as joining values from multiple dimensions that tend to decrease ridership).

Various techniques exist to conduct this random reordering, and there is no agreement in the literature about the "best" Latin hypercube design. The method we have implemented to achieve a relatively good design for this risk analysis is to generate numerous possible orderings of the draws for each dimension. Then, moving sequentially through the dimensions, we iteratively select the ordering that minimizes the maximum correlation with the draws for each of the previously selected dimensions. This approach, which minimizes correlation for previous dimensions, without accounting for the possible impact on the correlation of dimensions yet to be considered, is reasonably effective if the pool of candidate orderings is sufficiently large.

The use of the Latin hypercube design is a change from the fractional factorial design used for the 2016 Business Plan risk analysis. The Latin hypercube design of experiments is advantageous over a factorial or grid-based design, as every experimental observation can provide useful information, even when some input factors are potentially unimportant or spurious. Figure 4.5 provides an illustration of this: because every run is unique in each input dimension, all of the observations provide unique and useful information even if one dimension is spurious.

One additional advantage of a Latin hypercube design of experiments is that the required number of experimental runs is not explicitly dependent on the dimensionality of the input variables. With a factorial or grid-based design, the number of experiments required expands exponentially with the number of input dimensions. This can be partly mitigated by using a fractional factorial design (that does not require an experiment at every grid point), but even fractional factorial designs can require very large numbers of experiments to support high dimensional exploratory analysis.[24] A large fractional factorial design can also be subject to irregular requirements for sample size: depending on the resolution and existing dimensionality of the design, increasing the dimensionality of the risk factor space by one dimension may result in no need to expand the number of experiments, or it may require tripling the number.

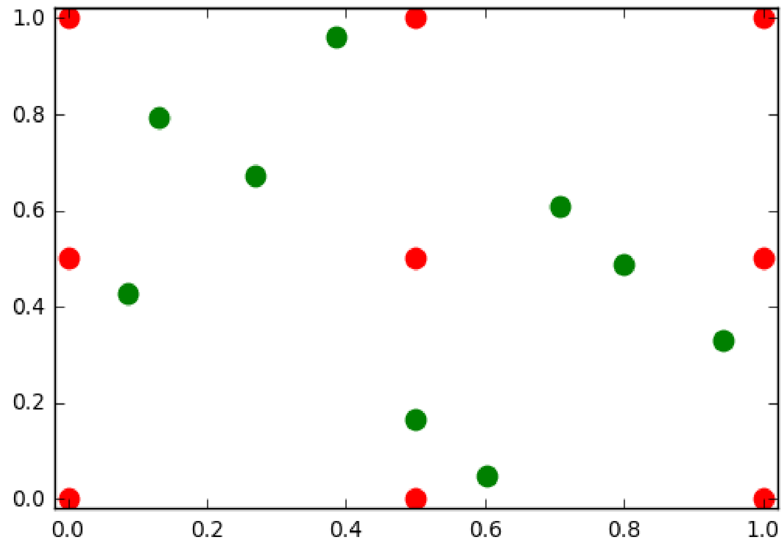
The Latin hypercube design does not demand any particular number of experiments. Adding a dimension without changing the number of experimental runs marginally degrades the efficiency of the design. Conversely, increasing the number of experimental runs while holding the dimensionality of the problem content can marginally improve the results. Practical experience across multiple domains has led to a "rule of thumb" that good results for prediction can be obtained from 10 experimental data points per input variable dimension.[25] Using this "rule of thumb" 150 BPM-V3 model runs were run for each model year and operating plan, as none included more than 15 unique risk variables.

[24] As a practical example, the fractional factorial design used for the 2016 Business Plan risk analysis effectively limited the number of risk variable that could be evaluated to 10.

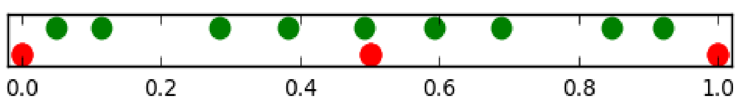
[25] Loeppky, J., J. Sacks, and W.J. Welch. 2009. Choosing the sample size of a computer experiment: A practical guide. Technometrics, 366-376.

Figure 4.5 Contrasting Experimental Designs

A representation of a factorial design (in red) which covers the edges of the solution space, and a Latin hypercube (green) with the same number of runs, with better coverage of the middle of the space.



If the Y dimension is not important, the factorial design collapses to only 3 data points, while the Latin hypercube still has 9.



Source: Cambridge Systematics.

4.3 Final Revenue Regression Models

The forecast revenues from the 150 BPM-V3 runs were used as data points for developing LR equations of the log of revenue as a function of the risk variables, used as the initial step in defining the meta-model for each forecast year. The final set of linear regression models for each model year and operating plan took the following functional form:

$$\ln(\text{Revenue}) = \text{Constant} + \beta_1 \times \text{Var}_1 + \beta_2 \times \text{Var}_2 \ldots$$

This model is a main effects model with no interaction terms and de-trended the observed data well. The estimated models are shown in Table 4.1. All models have R^2 values above 0.9, indicating that the linear regression model fits the BPM-V3 data points very well, and all of the signs and magnitudes of model coefficients are sensible. For example, a positive value on auto operating cost indicates that, as auto operating cost increases (i.e., it becomes more expensive to drive), HSR revenue also increases.

Table 4.1 Revenue Linear Regression Model Coefficients

Constant and Regression Model Variables	2029 – VtoV		2033 – Phase 1		2040 – Phase 1	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Constant	18.9132	65.16	20.3300	74.06	20.4995	73.53
HSR Mode Choice Constant – Business	0.2239	46.33	0.1955	43.88	0.2040	44.97
HSR Mode Choice Constant – Commute	0.0648	7.08	0.0588	7.08	0.0569	6.72
HSR Mode Choice Constant – Recreation/Other	0.4431	52.52	0.4196	54.77	0.4189	53.72
Trip Frequency Constant – Business/Commute	0.4177	11.93	0.4054	12.61	0.4109	12.50
Trip Frequency Constant – Recreation/Other	0.5182	6.58	0.6474	8.97	0.6430	8.77
Auto Operating Cost	0.0122	6.71	0.0117	6.82	0.0125	8.72
HSR Fare	0.1126	3.67	0.1546	5.52	0.1450	5.09
HSR Headway	-0.2926	-16.47	-0.2300	-16.39	-0.2301	-16.09
HSR Reliability	1.4659	5.20	0.6992	2.73	0.6561	2.51
HSR Access/Egress Connecting Service – Scenario 1	-0.0510	-2.40	NA	NA	NA	NA
HSR Access/Egress Connecting Service – Scenario 3	0.0034	0.38	NA	NA	NA	NA
Airfare	NA	NA	0.0293	0.47	0.0230	0.36
Coefficient on Transit Access-Egress Time/Auto Distance Variable	0.1107	2.79	0.0081	0.23	0.0079	0.22
Demographic Forecast CEF	-0.0362	-2.14	-0.0377	-2.44	-0.0892	-5.67
Demographic Forecast Moody's	0.0629	3.72	0.0652	4.19	0.0529	3.33
Automobile IVT Coefficient	NA	NA	NA	NA	-0.4675	-6.45
Extremely Long Access/Egress	-0.0288	-2.89	0.0115	1.26	0.0122	1.31
Model Statistics						
R^2	0.978		0.977		0.977	
Adjusted R^2	0.975		0.975		0.975	

The linear regression trend model provides an initial prediction for the (log of) revenue generated by the BPM-V3 model for each run. The difference between the linear regression prediction and the actual revenue observed for each run represents the residual, which is used as the dependent variable of a GPR model. As noted in previous sections, the GPR model reduces the differences between the forecast revenue using the linear regression model for the meta-model and the revenue that would actually be forecast using the full BPM-V3. The GPR model is fitted using the 150 residual values and the same explanatory risk factor data.

The cross-validation results for the revenue GPR regression models are provided in Table 4.2. It is important to note that these scores in the first row represent a measure of improvement in model fit above and beyond the fit achieved from the linear regression model alone, instead of the absolute model fit. The 2029 cross validation score is lower than for the 2033 and 2040 forecasts in large part due to the presence of three discrete connecting service scenarios in this forecast (see Section 3.6), and in particular the singly unlikely Scenario 1, for which predictions are hampered by the scarcity of data (only eight BPM-V3 runs are used to represent this unlikely scenario). Nevertheless, the GPR still provides a notable improvement in model fit above and beyond the linear regression model for this forecast. The second row shows the root mean squared error (RMSE)[26] of the cross-validation prediction of annual system revenue (in millions of dollars) across all observations in each 150 run dataset. In order to provide an order of magnitude comparison of the average error measured by the RMSE, the table also shows the long-distance revenue forecast for the 2018 Business Plan and the RMSE as a percent of the long-distance revenue. These results indicate that for the Phase 1 forecasts, the meta-model is generally predicting the output of BPM-V3 within 0.5%.

Table 4.2 Revenue Gaussian Process Regression Model Cross Validation Results

	2029 – VtoV	2033 – Phase 1	2040 – Phase 1
GPR Cross Validation Score	0.747	0.987	0.983
RMSE of Cross Validation Predictions (millions of 2017$)	$14.4	$7.1	$9.0
Long Distance HSR Revenue – 2018 Business Plan Base Runs (millions of 2017$)	$823	$2,085	$2,329
RMSE as a percent of Base Run Long Distance HSR Revenue	1.7%	0.3%	0.4%

To generate a final prediction of the revenue for a particular set of risk factor inputs, the revenue arising from additional induced and visitor travel risk factors is added to the revenue resulting from the GPR meta-model. The revenue arising from the induced and visitor travel risk factors is over and above the revenue resulting from HSR travel modeled using the BPM-V3 model and captured as part of the GPR.

4.4 Final Ridership Regression Models

The ridership forecasts from the 150 BPM-V3 runs were used as data points for developing the meta-model linear regression equations of the log of ridership as a function of the risk variables for each

[26] RMSE is a measure of average error literally calculated as the square root of the mean of the squared differences for a set of observed and modeled values. In this case, the observed values were the revenues forecast using the BPM-V3 for each of the 150 full model runs used to estimate the LR and GPR. The modeled values were the revenues estimated by applying the GPR using the input values for the risk variables defining the 150 full model runs.

forecast year. The final set of ridership linear regression models for each model year and operating plan took the following functional form:

$$\ln(\text{Ridership}) = \text{Constant} + \beta_1 \times \text{Var}_1 + \beta_2 \times \text{Var}_2 \ldots.$$

This model is a main effects model with no interaction terms. The estimated models are shown in Table 4.3. All models have R^2 values above 0.9, indicating that the regression model fits the BPM-V3 data points very well, and all of the signs and magnitudes of model coefficients are sensible. For example, a negative value on HSR fare indicates that, as HSR fare increases, HSR ridership decreases. Note that for revenue this is not always the case since for certain values of HSR fare; the increase in HSR ridership offsets the loss of revenue from a decrease in HSR fare.

Table 4.3 Ridership Regression Model Coefficients

Constant and Regression Model Variables	2029 – VtoV		2033 – Phase 1		2040 – Phase 1	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Constant	15.9131	59.07	17.4099	59.98	17.5535	59.63
HSR Mode Choice Constant – Business	0.2063	45.99	0.1913	40.60	0.1993	41.62
HSR Mode Choice Constant – Commute	0.0771	9.08	0.0696	7.92	0.0683	7.64
HSR Mode Choice Constant – Recreation/Other	0.4602	58.77	0.4567	56.37	0.4544	55.20
Trip Frequency Constant – Business/Commute	0.4109	12.65	0.3767	11.08	0.3826	11.02
Trip Frequency Constant – Recreation/Other	0.4975	6.81	0.6629	8.69	0.6581	8.51
Auto Operating Cost	0.0107	6.36	0.0089	4.92	0.0097	6.40
HSR Fare	-0.6945	-24.37	-0.6956	-23.51	-0.7057	-23.46
HSR Headway	-0.2998	-18.18	-0.2257	-15.21	-0.2257	-14.95
HSR Reliability	1.5447	5.91	0.8038	2.97	0.7777	2.82
HSR Access/Egress Connecting Service – Scenario 1	-0.0477	-2.42	NA	NA	NA	NA
HSR Access/Egress Connecting Service – Scenario 3	0.0008	0.08	NA	NA	NA	NA
Airfare	NA	NA	-0.0286	-0.44	-0.0337	-0.51
Coefficient on Transit Access-Egress Time/Auto Distance Variable	0.1494	4.05	0.0640	1.68	0.0663	1.71
Demographic Forecast CEF	-0.0437	-2.80	-0.0458	-2.81	-0.0971	-5.85
Demographic Forecast Moody's	0.0568	3.62	0.0548	3.32	0.0424	2.53
Automobile IVT Coefficient	NA	NA	NA	NA	-0.4191	-5.48
Extremely Long Access/Egress	-0.0233	-2.51	0.0136	1.41	0.0143	1.46
Model Statistics						

Constant and Regression Model Variables	2029 – VtoV		2033 – Phase 1		2040 – Phase 1	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
R^2	0.982		0.978		0.978	
Adjusted R^2	0.980		0.976		0.976	

The cross-validation results for the ridership GPR regression models are provided in Table 4.4. It is important to note that these scores represent a measure of improvement in model fit above and beyond the fit achieved from the linear regression model alone, instead of the absolute model fit. The 2029 cross validation score is lower than for the 2033 and 2040 forecasts in large part due to the presence of three discrete connecting service scenarios in this forecast, and in particular the singly unlikely Scenario 1, for which predictions are hampered by the scarcity of data (only 8 BPM-V3 runs are used to represent this unlikely scenario). Nevertheless, the GPR still provides a notable improvement in model fit above and beyond the linear regression model for this forecast. The second row shows the root mean squared error (RMSE) of the cross-validation prediction of annual system ridership (in millions) across all observations in each 150 run dataset. As with Table 4.2 for the revenue GPR, Table 4.4 also shows the long-distance ridership forecast as well as the RMSE as a percent of the long distance ridership for the 2018 Business Plan to provide an order of magnitude comparison of the average error measured by the RMSE.

Table 4.4 Ridership Gaussian Process Regression Model Cross Validation Results

	2029 – VtoV	2033 – Phase 1	2040 – Phase 1
GPR Cross Validation Score	0.834	0.986	0.983
RMSE of Cross Validation Predictions (millions of HSR riders)	0.24	0.16	0.19
Long Distance HSR Ridership – 2018 Business Plan Base Runs (millions of riders)	14.4	35.6	39.4
RMSE as a percent of Base Run Long Distance HSR Revenue	1.7%	0.4%	0.5%

To generate a final prediction of the ridership for a particular set of risk factor inputs, the ridership arising from additional induced and visitor travel risk factors is added to the ridership resulting from the GPR meta-model. The ridership arising from the induced and visitor travel risk factors is over and above the ridership resulting from HSR travel modeled using the BPM-V3 model and captured as part of the GPR.

4.5 Revenue Results of the Monte Carlo Simulation

A Monte Carlo simulation using the regression meta-model was run 100,000 times using different combinations of values of the risk variables, with the values being drawn from the assigned risk variable distributions. Note that some risk factors include multiple components that are sampled in the

Monte Carlo analysis. For example, values are sampled from both the uncertainty component distribution and the terminal/wait time component distribution for the HSR Mode Choice Constant risk variable. Appendix A details the components of each risk variable, the range of values and distributions for each component, and correlation between distributions of risk variables. Setting a positive correlation between two risk variable components results in the Monte Carlo simulation having a higher probability of sampling from the same point on the distribution (e.g., a 100-percent positive correlation would result in two risk variables always being chosen from the same percentile point on the distribution).

The revenue output from these 100,000 Monte Carlo runs was used to develop the revenue range and probability of occurrence, as shown in Table 4.5. Short-distance trips less than 50 miles within the Southern California Association of Governments (SCAG) and the Metropolitan Transportation Commission (MTC) contribute $13 million in revenue in year 2033 and $14 million in 2040. This short-distance revenue was added to the year 2033 and year 2040 long-distance revenue for all probability levels to obtain total HSR revenue.[27]

Table 4.5 Year 2029 to 2040 HSR Revenue Range and Probability of Occurrence[a]

	Revenue (Millions of June 2017 Dollars)		
Probability	**2029 VtoV**	**2033 PH1**	**2040 PH1**
Minimum	$230	$645	$674
1%	$358	$983	$1,037
10%	$517	$1,409	$1,480
25%	$666	$1,777	$1,872
Median	$887	$2,301	$2,436
75%	$1,167	$2,937	$3,112
90%	$1,451	$3,556	$3,790
99%	$1,961	$4,650	$5,018
Maximum	$2,757	$6,311	$7,082
Base Run	$823	$2,098	$2,344

[a] The results are raw model output and do not account for ramp-up.

The "base run" is the revenue for the year and scenario forecast using the BPM-V3 model with the base input variable values. The forecast base run revenue is less than the median revenue estimated using the risk analysis for all forecast years. The risk analysis demonstrates that, overall, there is a greater than even chance that revenue will be higher than the base run for all forecast scenarios.

[27] Given the small proportion of short-distance trip HSR revenue, variability associated with forecasting short-distance HSR trips was deemed low risk in contributing to overall uncertainty of HSR revenue, and thus was not included as part of the risk analysis.

Figure 4.6, Figure 4.7, and Figure 4.8 plot the cumulative distribution of HSR revenue for years 2029, 2033, and 2040, respectively.

Figure 4.6 Year 2029 Cumulative Distribution of High-Speed Rail Revenue *(June 2017 Dollars)*

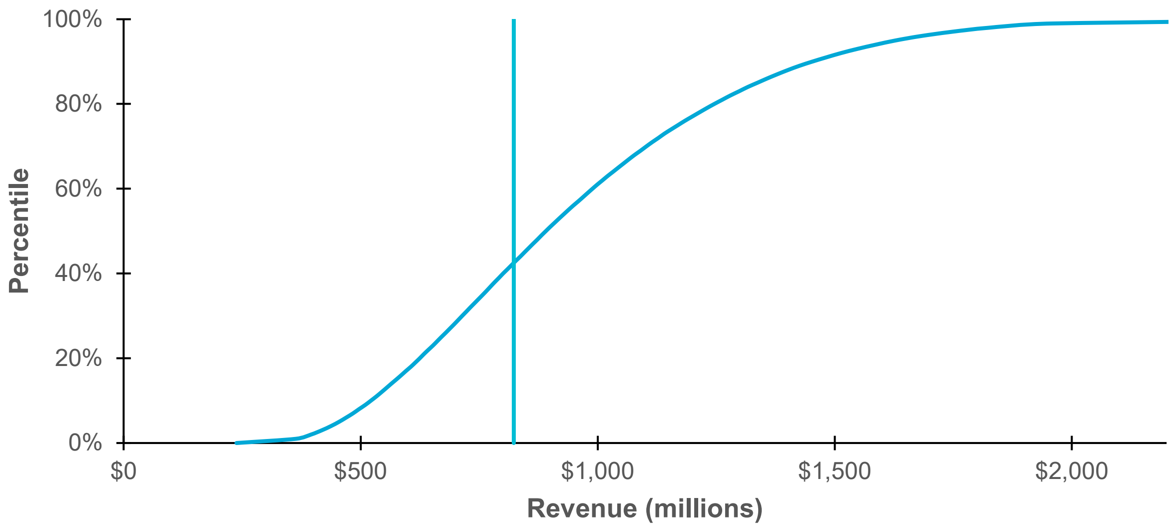


Figure 4.7 Year 2033 Cumulative Distribution of High-Speed Rail Revenue *(June 2017 Dollars)*

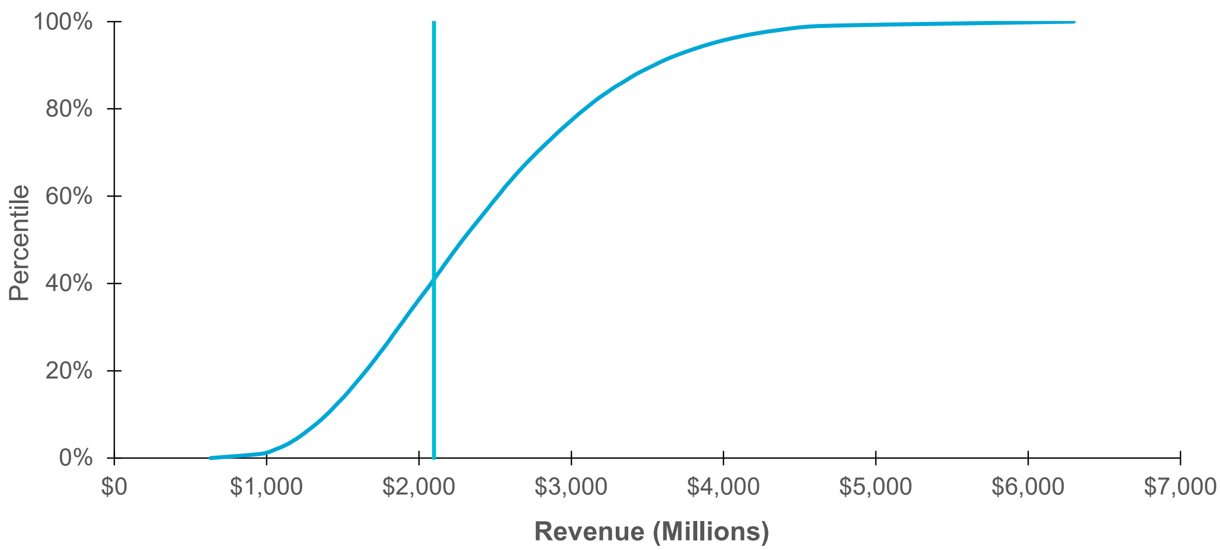
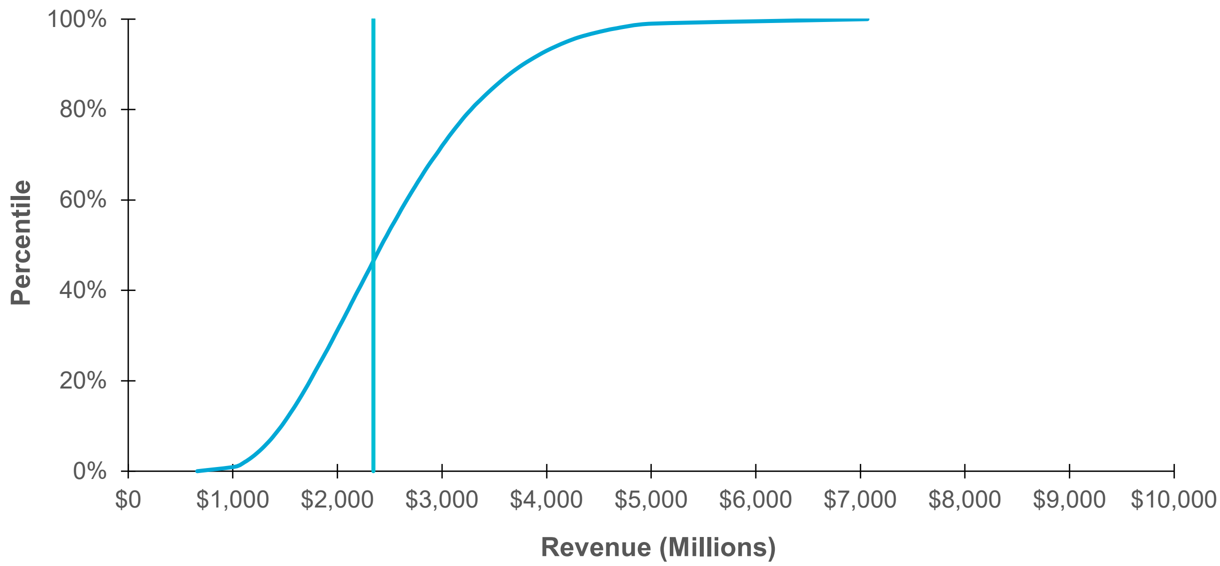


Figure 4.8 Year 2040 Cumulative Distribution of High-Speed Rail Revenue *(June 2017 Dollars)*



4.6 Ridership Results of the Monte Carlo Simulation

A Monte Carlo simulation using the ridership regression meta-model was applied to the same 100,000 runs developed for the revenue analysis. The ridership output from these runs was used to develop the ridership range and probability of occurrence, as shown in Table 4.6. Short-distance trips less than 50 miles within SCAG and MTC contribute 0.58 million in ridership in years 2033 and 0.63 million in 2040. This short-distance ridership was added to the year 2033 and year 2040 long-distance ridership for all probability levels to obtain total HSR ridership.[28]

The "base run" is the ridership for the year and scenario forecast using the BPM-V3 model with the base input variable values. The forecast base run ridership is less than the median ridership estimated using the risk analysis for 2029 and 2033 and higher than the median for 2040. This variation reflects the differential impact that some risk factors have for longer or shorter (i.e., higher or lower revenue) trips.

[28] Given the small proportion of short-distance trip HSR ridership, variability associated with forecasting short-distance HSR trips was deemed low risk in contributing to overall uncertainty of HSR ridership, and thus was not included as part of the risk analysis.

Table 4.6 Years 2029 to 2040 High-Speed Rail Ridership Range and Probability of Occurrence[a]

	Ridership (Millions)		
Probability	2029 VtoV	2033 Ph1	2040 Ph1
Minimum	3.3	8.9	9.7
1%	5.6	14.9	15.6
10%	8.3	21.6	22.6
25%	10.8	27.4	28.9
Median	14.5	36.1	38.0
75%	19.1	46.7	49.2
90%	23.9	57.5	60.6
99%	32.9	76.9	81.8
Maximum	47.3	111.7	117.6
Base Run	14.4	36.2	40.0

[a] The results are raw model output and do not account for ramp-up.

Figure 4.9, Figure 4.10, and Figure 4.11 plot the cumulative distribution of HSR ridership for years 2029, 2033, and 2040, respectively.

Figure 4.9 Year 2029 Cumulative Distribution of High-Speed Rail Ridership

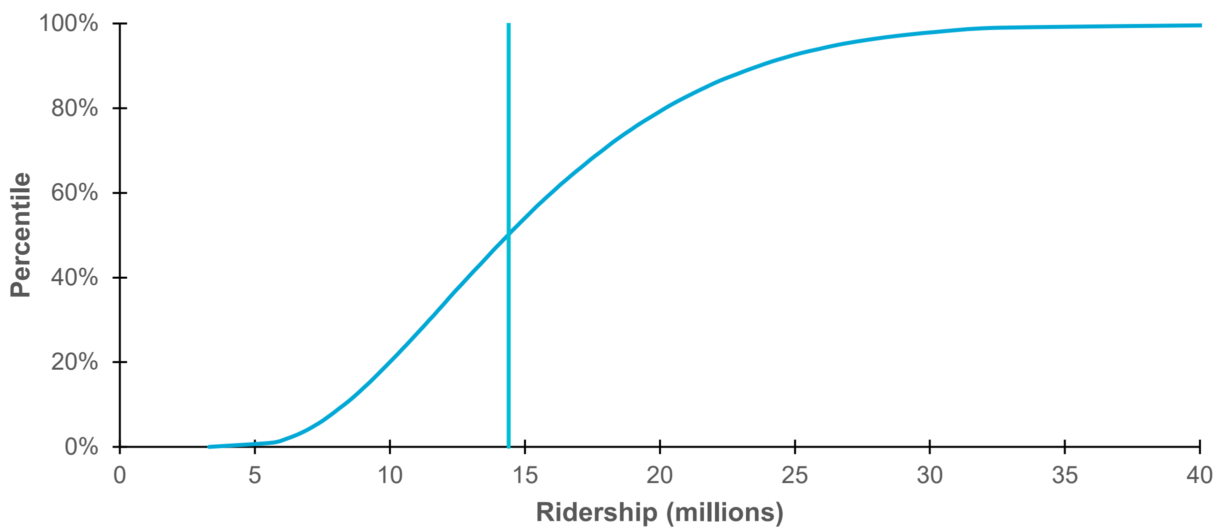


Figure 4.10 Year 2033 Cumulative Distribution of High-Speed Rail Ridership

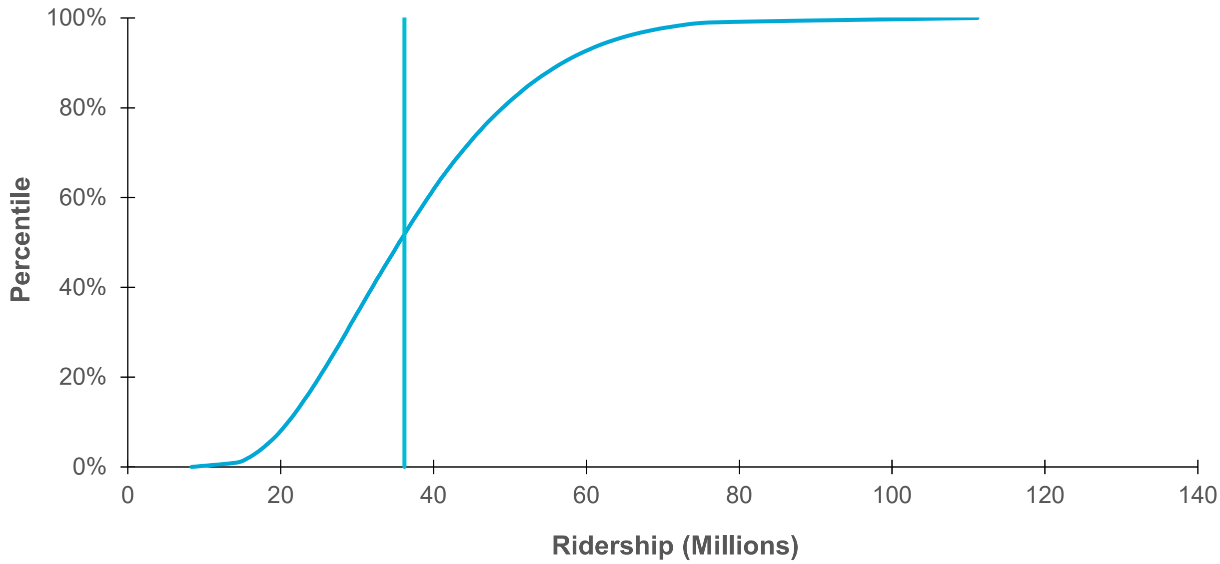
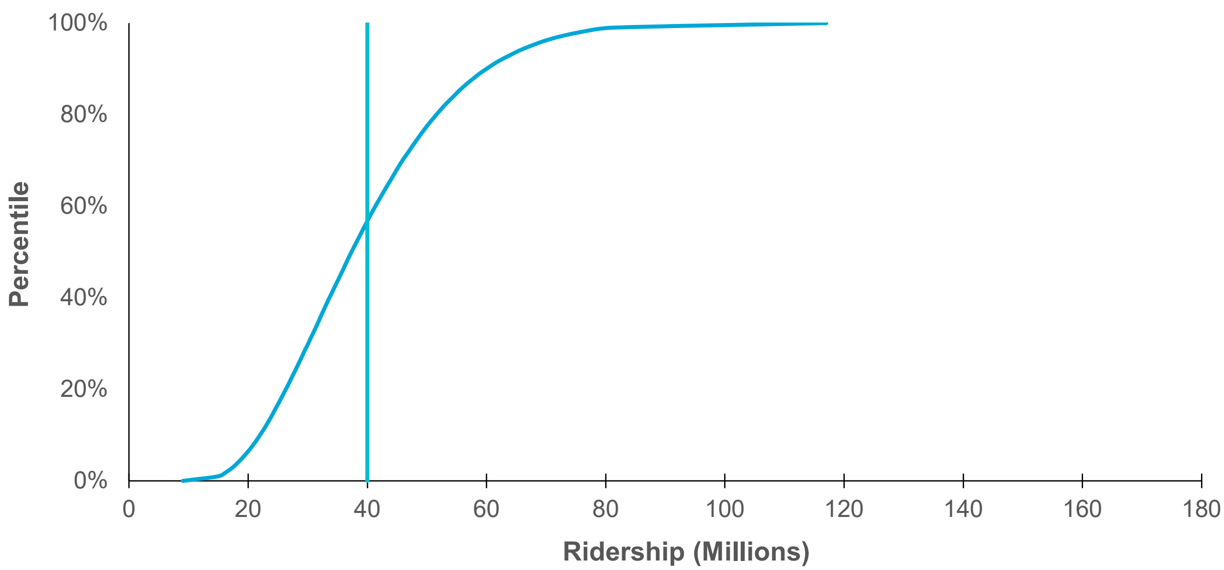


Figure 4.11 Year 2040 Cumulative Distribution of High-Speed Rail Ridership



4.7 Relative Importance of Risks

One feature of the risk analysis approach taken here is that the probability distribution of forecasts of high-speed rail ridership and revenue result from the underlying uncertainty in several variables that have direct impacts on high-speed rail ridership and revenue. Each of those variables contributes to the uncertainty in different ways, which can be quantified by examining the overall distribution of the forecasts. There are several methods outlined in the literature for expressing the relative importance of risk factors with a GPR model, although there is no well-established and universally accepted "best" method for this analysis.

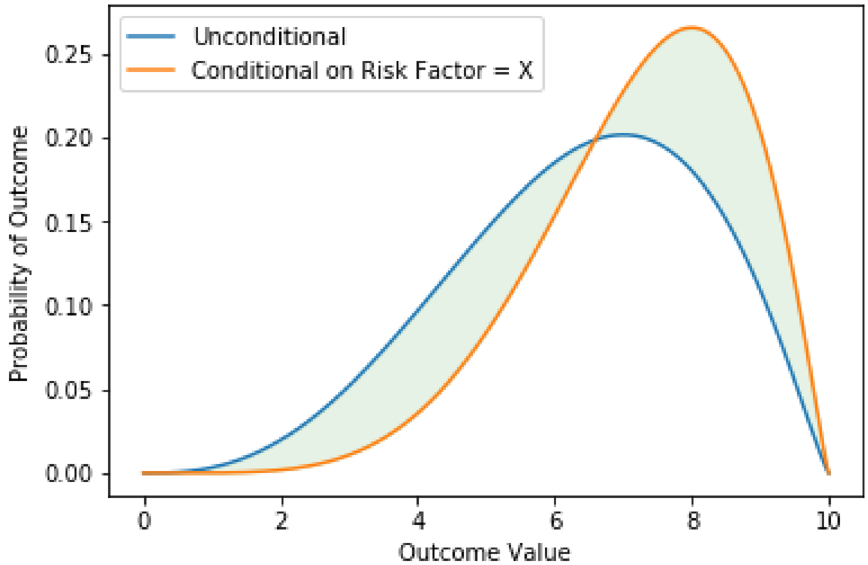
To provide a measure of the relative importance of the risk factors for the high-speed rail risk analysis, we adopt an alternative method called the "Delta Moment-Independent Measure".[29,30] Instead of measuring the contribution to variance, the Delta measure calculates the volume of change in the probability density function (pdf) that is a result of the change in any particular risk factor. It is called "moment-independent" because it can measure not just changes in the mean or variance (i.e., the first and second moments), but also changes to the shape of the distribution (e.g., changes in asymmetric tails, referred to as skewness).

The adoption of the Delta measure for risk factor importance analysis also is supported by the detail of the Monte Carlo risk analysis. The Delta measure can be computed efficiently from the same 100,000 Monte Carlo draws used to develop the primary risk analysis result, thus minimizing the additional computational effort required to develop this measure. It is also compatible with correlated input factors, which represent a significant portion of the relative importance of various factors. The Delta measure is computed by calculating the absolute difference between the unconditional distribution of the outcome measure (i.e., the overall total distribution) and the distribution conditional on any particular value of an individual risk factor. This absolute difference is illustrated by the shaded region in Figure 4.12. In the figure, the blue line represents the unconditional probability distribution of the outcome measure (e.g., the overall probability of various levels of revenue) and the orange line represents a conditional probability distribution (e.g., the probability of various levels of revenue when auto operating costs are set to 27 cents per mile). The area of the shaded area is multiplied by the probability of observing the particular value of the risk factor (e.g., the probability that auto operating costs will be 27 cents per mile), and this weighted value is integrated across all possible values of the risk factor. Thus, if the net impact of a risk factor is large but only for unlikely values of that risk factor, the overall Delta importance measure can be small.

[29] Plischke, E., E. Borgonovo, and C. L. Smith. 2013. "Global sensitivity measures from given data," European Journal of Operational Research 226.3: pages 536 to 550.

[30] Borgonovo, E. 2007. "A new uncertainty importance measure," Reliability Engineering & System Safety 92.6: pages 771 to 784.

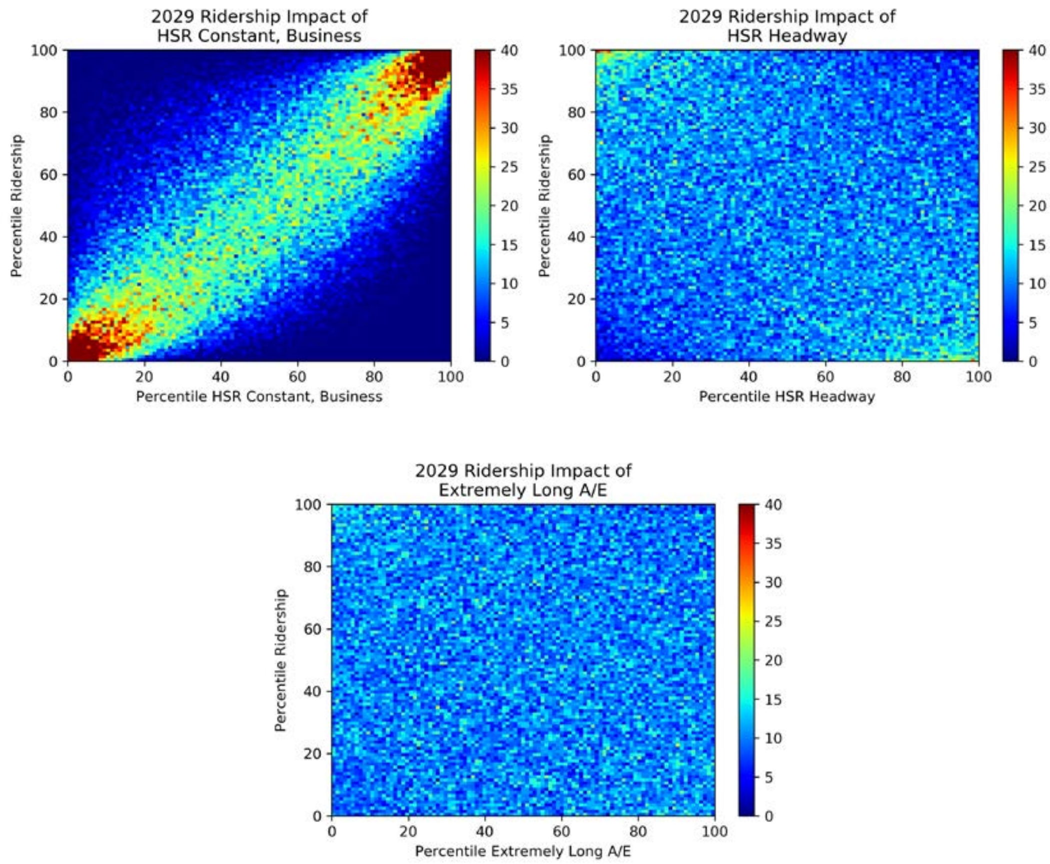
Figure 4.12 Illustration of Delta Moment-Independent Measure



Source: Cambridge Systematics.

The development of the Delta measure can also be understood graphically by plotting a heat map from the Monte Carlo results, where both the risk factor and the outcome measure are represented using percentile scales. The cells of the heat map can be defined by a sufficiently dense grid (e.g., 100 by 100) overlaid on the data, with each grid cell value determined by the number of Monte Carlo draws falling within the cell. The use of the percentile scaling ensures that every column and row of the heat map will contain substantially the same number of total observations (subject to possible trivial variations from rounding). When the risk factor is unimportant, any value of the risk factor will result in substantially the same distribution of outcomes, yielding a homogenous column of heat map cells. If on the other hand, the risk factor is important, then some or all of the columns of the heat map will show a nonhomogeneous pattern. An example of a relatively important, moderately important, and comparatively less important risk factor (respectively, the HSR constant for business travel, HSR headway, and the extremely long access/egress penalty) is shown in Figure 4.13. The magnitude of the Delta measure can be numerically approximated from the heat map by taking the total deviation of all heat map cells from the average per-cell value.

Figure 4.13 Three Illustrative Heat Maps from the 2029 SV to CV Scenario



This graphical approach also highlights one feature of the Delta measure: even for a spurious risk factor (i.e., something with no impact on the outcome whatsoever), the presence of simulation noise will result in a small, but nonzero, estimated Delta value. Graphically, this appears as a heat map that appears as random noise, but not as a perfectly homogenous surface.

Delta measures for various individual risk factors can be compared against each other to gain insight into the relative importance of each risk factor, but in the presence of correlated risk factors, the Delta measures do not neatly partition the total importance. This is because risk factors can impact outcomes both directly and indirectly, as any particular risk factor realization can be associated with a change in the marginal likelihood of other risk factor values. For example, the alternative specific constants for business, commute, and recreation/other are all positively correlated. Although the overall fraction of ridership by commuters is relatively small, high commute constants tend to happen alongside high business and recreation/other constants, which both directly trigger much larger shifts in ridership than the direct effect of the commute constant on commute ridership. Using this approach, we are not able to represent the relative impact of individual risks as a percentage of the total impact.

The Delta measure of relative importance of each risk variable component is shown in Table 4.7. While the values in the table are not percentages, they can be compared against each other to consider relative importance, with larger values indicating more important factors, and smaller values indicating less important factors. The high-speed rail constants' variation is the most important factor in determining the revenue distribution. This results both from the high level of uncertainty on these risk values (with a very wide distribution of possible values), as well as the large sensitivity of high-speed rail revenue and ridership to these constants. There remains a significant amount of uncertainty associated with how travelers will view high-speed rail, because there is no way to observe and collect data related to it until high-speed rail opens. After the constants, the next most important risk factor is the level of visitor travel, which is tied both to the overall uncertainty about how travelers will view high-speed rail, as well as a high degree of uncertainty over the overall number of visitors to California in the future. Other risk factors, such as the HSR service frequency (headways), have a small importance in the risk analysis, in part, because we are able to forecast controllable factors with a higher level of confidence.

Table 4.7 Delta Measure of Relative Importance for Revenue Risk Factors

Risk Variables	2029 – VtoV	2033 – Phase 1	2040 – Phase 1
HSR Constant – Business	0.3901	0.3675	0.3693
HSR Constant – Commute	0.2298	0.2259	0.2238
HSR Constant – Recreation/Other	0.4295	0.4277	0.4137
Terminal and Wait Time	0.0416	0.0383	0.0394
Trip Frequency Constant – Business/ Commute	0.0385	0.0405	0.0410
Trip Frequency Constant – Recreation/ Other	0.0309	0.0344	0.0341
Trip Frequency – Economic Cycle	0.0338	0.0325	0.0340
HSR Reliability	0.0214	0.0204	0.0182
HSR Headway	0.0517	0.0562	0.0558
HSR Fares	0.0255	0.0329	0.0354
Airfares	NA	0.0191	0.0183
Exceptionally Long Access/Egress	0.0179	0.0169	0.0170
HSR Access/Egress by Transit Variable	0.0190	0.0173	0.0184
Demographic Forecast CEF	0.0258	0.0338	0.0430
Demographic Forecast Moody's	0.0255	0.0330	0.0418
Automated Vehicle Market Penetration	NA	NA	0.0173
Automated Vehicle Shift in Disutility of Travel Time	NA	NA	0.0164
Automated Vehicle Fuel Economy	NA	NA	0.0177
Shared Vehicle Market Penetration	NA	NA	0.0194
Shared Vehicle Operating Cost	NA	NA	0.0250
Standard Automobile Operating Cost	0.0242	0.0215	0.0288
HSR Access/Egress Connecting Service – Scenario 1	0.0227	NA	NA
HSR Access/Egress Connecting Service – Scenario 3	0.0226	NA	NA
Induced Travel Ratio	0.0201	0.0188	0.0206
Visitor Ridership	0.1799	0.1836	0.1833

Appendix A. Risk Variable Component Specification for Monte Carlo Simulation

Table A.1 lists the risk factors considered for the risk analysis. Table A.2 details the components of each risk variable used to specify the risk factors, the range of values and distributions for each component, and correlation between distributions of risk variables. Some risk factors include multiple components that are sampled in the Monte Carlo analysis. For example, values are sampled from both the error component distribution and the terminal/wait time component distribution for the High-Speed Rail (HSR) Mode Choice Constant risk variable. The sampled values are combined, as appropriate, prior to inputting the value into the regression model used for the Monte Carlo simulation. Setting a positive correlation between two risk variable components results in the Monte Carlo simulation having a higher probability of sampling from the same point on the distribution (e.g., a 100-percent positive correlation would result in two risk variables always being chosen from the same percentile point on the distribution).

Table A.1 Risk Factors Considered for Risk Analysis

Risk Factor	Risk Factor in 2016 Risk Analysis?	Comments
HSR Main Mode Constants	Yes	• Main contributor to 2016 HSR revenue variance
Trip Frequency Constants	Yes	• Second largest contributor to 2016 HSR revenue variance
Total Population, Households, Employment (both statewide total and distribution)	Yes	• Almost 1:1 relationship of changes in households (accessible to HSR) to HSR ridership • Uncertainty introduced by Federal changes regarding immigration policy • Uncertainty regarding average household size changes
Auto Operating Cost	Yes	• Potentially more uncertainty than in 2016 regarding gas costs, CAFE standards, Cap & Trade, plug-in electric vehicles
Airfares	Yes	• Air travel is a direct competitor for HSR travel and there is uncertainty regarding how airlines will respond to HSR
Impacts of Automated & Shared Vehicles	Yes	• Re-evaluate impacts of automated & shared vehicles in 2029 & 2033 as well as 2040 • Incorporated uncertainty represented in 2016 by auto in-vehicle travel times in this factor
Coefficient on Transit Access-Egress Time/Auto Distance Variable	Yes	• Impact on travel with long access times outside ranges presented or observed in model estimation data
HSR Service Frequency	Yes	• Large service related contributor to 2016 HSR revenue variance
HSR Fares	Yes	• Contribution to 2016 HSR revenue variance was mixed • Possible improved capture of non-linearities using GPR • Important to understand risk levels of factors the Authority or its operator can control
HSR Connecting Service Via HSR Bus	Yes	• Response to HSR bus connecting service is uncertain but important for Silicon Valley to San Joaquin Valley scenario
HSR Reliability	No	• Other HSR systems show very high reliability, however, some uncertainty may be introduced by operations on Caltrain tracks in Bay Area
Induced demand	No	• Model forecasts some induced travel, but low in comparison to some international estimates
Visitor travel	No	• Visitor travel is not modeled using BPM-V3 • International experience suggests this will be a contribution to ridership and revenue

Table A.2 Risk Variable Distributions Used in Monte Carlo Analyses

Risk Variable	Components	Years	Minimum	Most Likely	Maximum	Distribution	Notes
HSR Mode Choice Constant – Business	Error Component	All	-2.335	0.0	2.335	PERT – Standard (Shape = 4)	Unit = Offset from calibrated coefficient. 50% Correlation with Commute & Recreation/Other HSR Error Components.
	Terminal/Wait Time	All	-0.3264	0.0	0.1632	Triangular	Unit = Offset from calibrated coefficient. 100% Correlation with Commute & Recreation/Other Term./Wait Times.
HSR Mode Choice Constant – Commute	Error Component	All	-1.222	0.0	1.222	PERT – Standard	Unit = Offset from calibrated coefficient. 50% Correlation with Business & Recreation/Other HSR Error Components.
	Terminal/Wait Time	All	-0.3264	0.0	0.1632	Triangular	Unit = Offset from calibrated coefficient. 100% Correlation with Business & Recreation/Other Term./Wait Times.
HSR Mode Choice Constant – Recreation/Other	Error Component	All	-1.354	0.0	1.354	PERT – Standard	Unit = Offset from calibrated coefficient. 50% Correlation with Business & Commute HSR Error Components.
	Terminal/Wait Time	All	-0.1388	0.0	0.0694	Triangular	Unit = Offset from calibrated coefficient. 100% Correlation with Business & Commute Term./Wait Times.
Trip Frequency Constant – Business/Commute	Error Component	All	-0.275	0.0	0.275	PERT – Standard	Unit = Offset from calibrated coefficient. 50% Correlation with Recreation/Other Error Components.

Risk Variable	Components	Years	Minimum	Most Likely	Maximum	Distribution	Notes
	Economic Component	2029 2033 2040	-0.176 -0.176 -0.177	0.0 0.0 0.0	0.173 0.169 0.161	Triangular	Unit = Offset from calibrated coefficient. 100% Correlation with Recreation/Other Economic Component.
Trip Frequency Constant – Recreation/Other	Error Component	All	-0.132	0.0	0.132	PERT – Standard	Unit = Offset from calibrated coefficient. 50% Correlation with Business/Commute Error Components.
	Economic Component[1]	2029 2033 2040	-0.063 -0.064 -0.067	0.0 0.0 0.0	0.066 0.065 0.062	Triangular	100% Correlation with Business/Commute Economic Component.
Auto Operating Costs	Combined Components	2029 2033 2040	0.17 0.17 0.17	0.23 0.23 0.23	0.35 0.34 0.33	PERT – Shape=5 2040 n/a	Unit = 2017 dollar per mile. Full Model & Regression Model use 2005 dollar, rather than 2017 dollar. Conversion at following rate 202.6 / 262.286 based on CPI. 2040 values used in Full Model Runs, but not in Monte Carlo.
Auto Operating Costs Impacts of Autonomous and Shared-Use Vehicles	Owned Nonautonomous vehicle auto operating cost	2040	0.17	0.23	0.35	PERT – Shape=5	Unit = 2017 dollar per mile. Used in Monte Carlo, but not used in Full Model Runs
	Owned Autonomous Vehicle Market Penetration	2040	0.10	0.35	0.75	Triangular	Unit = Decimal percent of owned AVs used for long-distance trips. Used in Monte Carlo, but not used in Full Model Runs.
	AV Fuel Economy improvements	2040	0.10	n/a	0.50	Uniform	Unit = Decimal percent fuel economy improvements from base. Used in Monte Carlo but not used in Full Model Runs.

Risk Variable	Components	Years	Minimum	Most Likely	Maximum	Distribution	Notes
	Shared-use vehicle market share	2040	0.02	0.05	0.20	Triangular	Unit = Decimal percent of shared-used vehicles used for long-distance trips. Used in Monte Carlo but not used in Full Model Runs.
	Shared-use vehicle auto operating cost	2040	0.18	n/a	0.85	Uniform	Unit = 2017 dollar per mile. Used in Monte Carlo, but not used in Full Model Runs.
HSR Fares	N/A	All	0.74	1.0	1.42	Triangular	Unit = Factor from Base.
HSR Headway	N/A	2029 2033 & 2040	0.29 0.65	1.0 1.0	1.58 2.25	PERT – Standard	Unit = Factor from Base/Most Likely Value.
HSR Connecting Service	N/A	2029	Scenario 1 – 5%	Scenario 2 – 40%	Scenario 3 – 55%	Multinomial	Unit = 1 if Scenario is chosen, 0 otherwise.
Coefficient on Transit Access-Egress Time/Auto Distance Variable	Business/Commute Coefficient	2029	-2.0	-1.215	-1.215	PERT – Standard	Unit = Coefficient. Used in Full Model Run but not used in regression. 100% Correlation with Recreation/Other coefficient & Threshold parameter.
	Recreation/Other Coefficient	2029	-1.3	-0.88	-0.88	PERT – Standard	Unit = Coefficient. Used in Full Model Run, but not used in regression. 100% Correlation with Business/Commute coefficient & Threshold parameter.
	Threshold Parameter	2029	0.1	0.2	0.2	PERT – Standard	Unit = Threshold value. Used in Full Model Run but not used in regression. 100% Correlation with Business/Commute coefficient & Recreation/Other coefficient.

Risk Variable	Components	Years	Minimum	Most Likely	Maximum	Distribution	Notes
	Index Variable	2029	-0.1	0.0	0.0	PERT – Standard	Unit = Index variable. Not used in Full Model Runs, but used in regression. Middle value set to 0.05 for Full Model Runs.
Airfares		2033 and 2040	1.0	1.15	1.31	Triangular	Unit = Factor from Base.
Number and Distribution of Households throughout the State	N/A	All	CEF	Blended	Moody's	Triangular	Unit = Interpolate from Base.
HSR Reliability	N/A	All	0.90	0.99	0.997	PERT – Standard	Unit = Decimal percent.
Auto In-Vehicle Time Coefficient	Owned Autonomous Vehicle Market Penetration	2040	0.10	0.35	0.75	Triangular	Unit = Decimal percent of owned AVs used for long-distance trips. Used in Monte Carlo but not used in Full Model Runs.
	IVT Coefficient – Alone Travel	2040	0.50	0.75	1.0	Triangular	Unit = Factor of Auto IVT applied to AV market only for alone travel (applied for both auto main mode and auto access/egress).
	IVT Coefficient – Group Travel	2040	0.80	0.90	1.0	Triangular	Unit = Factor of Auto IVT applied to AV market only for group travel (applied for both auto main mode and auto access/egress). Perfect correlation between alone and group travel IVT coefficients.
Exceptionally Long Access and Egress *(Percent increase in disutility)*		All	150%	n/a	0%	Uniform	

Risk Variable	Components	Years	Minimum	Most Likely	Maximum	Distribution	Notes
Visitor Travel	Visitor Travel HSR Trips	2029	0.48	n/a	1.03	Uniform	Unit = HSR Trips (millions). 50% positive correlation between visitor travel high-speed rail ridership and total California resident high-speed rail ridership.
		2033	1.69	n/a	3.65		
		2040	1.87	n/a	4.03		
Induced HSR Ridership	Percent of additional ridership	All	0	7.5	15	Triangular	Unit = Percent of additional ridership. 50% negative correlation with the trip frequency constant.

Appendix B. High-Speed Rail Constants

The high-speed rail (HSR) constant for each of the four trip purposes (i.e., business, commute, recreation, and other) is composed of two components: 1) unexplained variation, and 2) terminal and wait time. The unexplained variation component represents the desirability to choose HSR that is not captured directly by the system variables included in the model. Terminal time is the out-of-vehicle time spent traveling from the point of departure from the access mode to the train platform. Wait time is the out-of-vehicle time spent waiting on the platform for the train to arrive and the time spent waiting for the train to leave the platform once boarded. The risks associated with each of the components are different and should be specified separately for the Monte Carlo experiments, as discussed in the next sections.

For full model risk analysis runs, terminal and wait times are included with the unexplained variation within the HSR constant.[31] For Monte Carlo risk analysis, each component of the HSR constant is considered as a separate risk variable with completely independent distributions. The former allows for estimation of a single regression model parameter, and does not require an additional risk variable in the experimental design framework. The latter allows for an understanding of the terminal/wait time's effect on ridership and revenue uncertainty independent from the HSR constant's effect on ridership and revenue uncertainty.

B.1 Unexplained Variation

An important part of any mode choice model is a modal constant that explains factors that are not quantifiable by the stated-preference (SP) 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 in the California/U.S., this is impossible, and thus there is uncertainty in the asserted constant.

The asserted HSR constant is the average of the estimated constant value from two distinct approaches. The first approach considered offsets from air and CVR constants derived from 2013 estimated SP constants. The second approach averaged the calibrated air and CVR constants used in model application. Details of the derivation of the HSR constant are documented in the *California High-Speed Rail Ridership and Revenue Model Business Plan Model-Version 3 Model Documentation*. Both approaches were reasonable approaches to arrive at an HSR constant, but this analysis takes the average of these values. Since each approach is reasonable on its own terms, the values derived from each approach must fall within the uncertainty range considered in the risk analysis.

In order to better understand the uncertainty associated with the HSR constant, additional analysis of the 2013 RP/SP survey data was undertaken by performing additional mode choice model estimation using additional variables that were not included in the Business Plan Model – Version 3 (BPM-V3). This additional analysis was separate from the procedure described above to assert the HSR constant. The

[31] The decision to bundle the terminal and wait times with the unexplained variation in the constants was made when the BPM-V3 was developed since the terminal and wait times were considered to be mode specific. The model implementation code could be modified to unbundle the terminal and wait times from the constant without impacting the underlying BPM-V3. This approach, however, was unnecessary for the risk analysis; the unbundling could effectively be accomplished through the method used to implement the risk analysis.

variables included demographic characteristics, trip characteristics, and attitudinal questions, as shown in Table B.1.

Table B.1 Additional Variables Considered in Analysis of High-Speed Rail Constant

Demographic Characteristics	Trip Characteristics	Attitudinal Questions
• Gender • Age • Worker Status • Highest education level achieved • Schedule flexibility	• Car not available for trip • Car needed at destination • Duration of stay	• Respondent's stated likelihood of ever using HSR service in the future. • Respondent's perceived economic value of HSR to the State of California. • Respondent's perceived environmental value of HSR to the State of California. • Respondent's support/opposition level to HSR. • Respondent's familiarity with conventional Amtrak, Acela services in the Northeast, and HSR in foreign countries.

Using the best model with these new variables, the HSR constant was recalibrated using the constant offset method, assuming the same calibrated CVR and air constants.[32] The resulting HSR constants under this new model were nearly identical to those of the original model, suggesting that even after controlling for all these additional factors, the constants we would assert for the HSR mode would have been about the same in relation to the calibrated air and CVR constants. While the estimated coefficients for several of the variables in Table B.1 were found to be highly statistically significant with expected signs and appropriate magnitudes on their own, these coefficients do not say much about the size or magnitude of the HSR constant, or its relation with CVR or air constants. If they had explained a portion of the unexplained variation included in the HSR constant, the resulting HSR constant asserted using the offset method should have been different from the HSR constant asserted using the offset method for the BPM-V3.

Given that this additional model estimation did not provide additional insight into the uncertainty of the HSR constant, we had no basis to narrow the range in uncertainty from the range assumed in previous risk analyses (as long as the BPM-V3 is used for forecasting). In previous risk analyses, the CVR constant was assumed to represent an absolute worst case lower bound for the uncertainty range for the HSR constant. This was reasonable, since none of the unobserved characteristics for HSR should be less attractive than CVR to travelers.

The shape of the HSR constant is assumed to be symmetric around the base case value. The asserted baseline constants come from averaging two reasonable approaches, as outlined above. The asserted constants for both approaches are equidistant from the asserted value for the BPM-V3; and by extension, should have the same likelihood of occurrence in the constant distribution. Thus, due to the symmetry

[32] These constants would change under a different model specification, but this allowed for direct comparison of the resulting HSR constant to those of the original model.

assumption, the absolute minimum bound of the CVR constant also determines the absolute maximum bound.

Each trip purpose (i.e., business, commute, recreation/other) is treated individually as separate risk factors since the difference between CVR and HSR base constants is different for each purpose, and thus, the CVR lower bound is different for each purpose. In addition, some parts of the uncertainty captured in the constants are likely to be totally correlated amongst trip purposes (i.e., 100 percent correlation), while others would be unrelated between purposes (i.e., 0 percent correlation). A 50-percent correlation between the HSR constant trip purposes was assumed to capture the judgment that a portion of the constants would be correlated, but not necessarily every aspect of them.

For the Monte Carlo simulation, a PERT distribution is specified rather than a triangular distribution, because the CVR constant represents an absolute minimum possible value for the HSR constant, essentially a tail event. Since the triangular distribution does not have tails, it would overstate the likelihood of observing a very unlikely tail event.

B.2 Terminal and Wait Time

B.2.1 Terminal Time

Terminal time is the out-of-vehicle time spent traveling from the point of departure from the access mode to the train platform. It currently is assumed that terminal times for CVR and air are 3 and 22 minutes, respectively; and for HSR, 10 minutes is assumed. A lower bound based on the CVR value is considered, but given that HSR stations will be larger than many CVR stations, a lower bound for the risk analysis simulations of 5 minutes is more appropriate.

The upper bound on terminal time is based on the air terminal time. An upper bound of 22 minutes is used for HSR terminal time, which is identical to the terminal time assumed at airports. This conservative upper bound assumes that the time it takes to traverse an HSR station is similar to airports, and that HSR travelers will need to undergo security similar to current Transportation Security Administration (TSA) security at airports.

B.2.2 Wait Time

Wait time is the out-of-vehicle time spent waiting on the platform for the train to arrive and the time spent waiting for the train to leave the platform once boarded. Wait times are often related to service headways, except when headways grow so long that travelers coordinate their arrivals to coincide with train departure. Research supports this assumed behavior. When transit headways are less than 11 minutes, travelers arrive at the stop randomly because they are not attempting to coordinate their arrival with the bus departure time. However, when bus headways exceed 38 minutes, travelers carefully arrive when the bus is scheduled to depart.[33] Because HSR trains will arrive infrequently, it is reasonable to assume that travelers will coordinate their arrivals with the HSR schedule.

This coordination guides the value for the upper bound of the wait time component. For trains with 60-minute headways, the mean wait time of travelers who do not coordinate their arrival at the station with the train will be 30 minutes. However, given the evidence that travelers do not arrive randomly, it is reasonable to assume that the average wait time will be less than 30 minutes in such a case.

If it is assumed that with 30-minute headways, 25 percent of travelers have random arrivals with 15-minute average waits, 50 percent of travelers have coordinated arrivals with 10-minute average waits, and 25 percent of travelers have coordinated arrivals with 5-minute average waits. The overall average wait time is 10 minutes. Thus, 10 minutes is used as a lower bound on the distribution for risk analysis.

The base wait and terminal times for HSR are set to 15 and 10 minutes, respectively. These were the terminal and wait times that were stated in both the 2005 and 2012/2013 RP/SP survey. The wait time and terminal time risk variables for each trip purpose are 100 percent correlated with each other, since factors that contribute to shorter or longer terminal and wait times would not differ by trip purpose. The risk variable has a triangular distribution since the ranges do not reflect extreme or highly unlikely events.

[33] Fan, W., and R. Machemehl. 2009. Do Transit Users Just Wait for Buses or Wait with Strategies? Some Numerical Results That Transit Planners Should See, Transportation Research Record: Journal of the Transportation Research Board, Issue 2111, pages 169-176.

Appendix C. Trip Frequency Constants

The trip frequency constants include the unexplained variation in the propensity of households to make long-distance trips within California. Within the risk analysis model, variation in the trip frequency constants is used to reflect the effect of the state of the economy on the proclivity of households to take high-speed rail (HSR). Instead of including distributions of household and employment levels directly as risk variables in the risk analysis model to account for changes in the state of the economy, risks associated with the state of the economy are taken into account within the trip frequency constant risk variable. The risks associated with each of the components are different and should be specified separately for the Monte Carlo experiments, as discussed in the next sections.

C.1 Unexplained Variation

The trip frequency model was calibrated to 2010 conditions and applied using forecast year socioeconomic data and networks. The changes in the demographic composition and the networks in the modeled forecast years result in an increase in annual long-distance trip rates compared to the year 2010 trip rates. This increase in annual long-distance trip rates is consistent with findings from the 1995 American Traveler Survey and the 2001 National Household Travel Survey (NHTS), which found a 21-percent increase in annual round trips per household over the six-year period from 10.15 annual trips per household to 12.32 annual trips per household.[34] This occurred even though the economic conditions in 2001 were not as good as in 1995 due to the "dot-com" bust. In addition, since some surveys were collected after 9/11, the 2001 NHTS trip rates may have been affected.

Annual long-distance trip rates appear to be relatively independent of disruptions caused by economic conditions, changes in technology, and changes in traveler perceptions and behavior. Information and communication technologies have been found to be a complement, and even be an incentive for, business trips.[35] During recessions and hard economic times, research has found that households choose to make more leisure trips closer to home for shorter periods of time, rather than taking longer trips that involve more days away from home.[36] As the baby boomers continue to move into retirement age, leisure travel also may increase due to fewer family obligations, higher incomes compared to their younger peers, and fewer necessary expenditures.[37] Research suggests that, if anything, long-distance travel may increase with changing technologies and demographics.

[34] NCHRP Report 735, Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models, Transportation Research Board, 2012, page 51.

[35] Aguilera, A. Business Travel and Mobile Workers, Transportation Research Part A: Policy and Practice, Volume 42, Issue 8, October 2008, pages 1109 to 1116.

Mokhtarian, P. If Telecommunication is such a good substitute for travel, why does congestion continue to get worse? Transportation Letters, Volume 1, Issue 1, January 2009, pages 1 to 17.

[36] Lamonda, J., and C. Bhat. 2011. A study of visitors' leisure travel behavior in the northwest territories of Canada. Transportation Letters, Volume 3, Issue 1, January 2011, pages 1 to 19.

[37] Lamonda, J., C. Bhat, and D. Hensher. 2008. An annual time use model for domestic vacation travel. Journal of Choice Modeling, Volume 1, Issue 1, pages 70 to 97.

Since changes in economic conditions, technologies, and traveler perceptions and behaviors are not hypothesized as a significant risk to annual long-distance trip rates, the trip frequency constant risk factor range is based on the range seen in forecasted annual long-distance trip rates produced by the model. The most likely value for each forecast year is the calibrated constant. The minimum value of the trip frequency constants is specified, such that for year 2040, the trip frequency constants produce average trip rates equal to the 2010 rates by trip purpose. The maximum value of the trip frequency constant is specified to mirror the deviations from the calibrated constants for the minimum values (i.e., symmetry of the constant offsets is assumed).

For each trip purpose (i.e., business/commute, recreation/other), some parts of the uncertainty captured in the constants are considered likely to be correlated amongst trip purposes (i.e., 100 percent correlation), while others would be unrelated between purposes (i.e., 0 percent correlation). A 50-percent correlation between the trip frequency constant trip purposes was assumed to capture that a portion of the constants' uncertainty would be correlated, but not necessarily every aspect of it.

For the Monte Carlo simulation, a PERT is specified rather than a triangular distribution, because the minimum and maximum values represent unlikely events. Since the triangular distribution does not have tails, it would overstate the likelihood of observing a very unlikely tail event.

Table C.1 shows the approximate results in terms of annual long-distance round trips per capita resulting from the specification of the constant ranges to account for unexplained variation. Note that symmetry of the constant offsets does not produce symmetry of the implied trip rates. This is due to the trip frequency choice model being specified as a logit model with choices of no long-distance trip, one long-distance trip traveling alone, or one long-distance trip traveling in a group on a given day. Since the base shares for each of these choices are very low (e.g., about 0.2 percent), the model is more sensitive to the constants on the high end than the low end.

Table C.1 Unexplained Variation of Trip Frequency Constants – Implied Annual Long-Distance Round Trips per Capita

Purpose	Implied Annual Long-Distance Round Trips per Capita for 2029			Implied Annual Long-Distance Round Trips per Capita for 2033			Implied Annual Long-Distance Round Trips per Capita for 2040		
	Minimum	Most Likely	Maximum	Minimum	Most Likely	Maximum	Minimum	Most Likely	Maximum
Business/Commute	1.68	2.21	2.90	1.74	2.28	2.99	1.87	2.46	3.23
Recreation/Other	5.14	5.86	6.67	5.22	5.95	6.78	5.50	6.27	7.14
Total	6.82	8.07	9.57	6.96	8.23	9.77	**7.37**	**8.73**	**10.37**

C.2 Economic Cycle

Economic cycles potentially impact several different variables in the model, including the number of workers, household income levels, and overall trip making. However, incorporating each of these risk factors separately is infeasible. These impacts are interrelated and can be accounted for jointly. Sensitivity tests have shown that the economic-cycle variations can be reasonably accounted for by changes in trip frequencies. Thus, the effect of economic cycles on HSR ridership and revenue is accounted for as a separate risk component in the trip frequency constants.

In order to determine the appropriate range in the trip frequency constant, changes in employment and income need to be translated into changes in the trip frequency constants. The primary driver for long-distance trip-making in the BPM-V3 model is the number of households within the State. Households are stratified into 99 different groups based on 4 household size groups, 3 auto ownership groups, 3 number of workers groups, and 3 income groups. The 4 x 3 x 3 x 3 groups result in 108 strata; 9 of which are illogical (i.e., 2 or more worker, 1 person households for the 9 groups defined by auto ownership and income). Total trips are based on the modeled trip frequency and the numbers of households in the State.

Employment is the metric used to define the economic cycles for the State. Employment has a secondary impact on trip frequency and a more direct impact on destination choice. However, the employment levels also can be used to more directly impact the total numbers of trips through relationships with households by numbers of workers and households by income group. For a given forecast of households, the numbers of 0, 1, and 2+ worker households should vary so that total workers in the State track the total employment. Likewise, in a recession, it should be expected that the number of low-income households should increase at the expense of middle- and high-income households and, likewise, that the number of middle-income households might increase at the expense of high-income households.

Suggested low and high employment levels representing the economic cycles were based on historic observations through 2014. The Great recession produced a -2.8-percent Compound Annual Growth Rate (CAGR) for employment in California between 2007 and 2010. Thus, for the low economic growth scenario, annual declines of 3.0 percent per year for the three years preceding the forecast year were assumed, with those declines being applied to the new “Low Scenario” statewide control total. The period from 1994 to 2000 was the high-water period for job growth in California with a 3.0-percent CAGR for five years. Thus, for the high economic growth scenarios, annual increases of 3.0 percent per year for the five years preceding the forecast year were assumed, with the increase applied to the new “High Scenario” statewide control total.

It was assumed that the low employment forecast would result in a commensurate decrease in the number of household workers. This was accomplished through increasing in the number of 0 and 1 worker households, and decreasing 2+ worker households. The above changes could result from some households moving from 2+ worker households to 1 worker households, and 1 worker households moving to 0 worker households to reflect the increasing unemployment. It was assumed that the increase in 0 worker households would result in an increase in low-income households and a commensurate decrease in high-income households. The changes could result from some households moving from high-income households to middle-income households, and middle-income households moving to low-income households to reflect the increasing unemployment or underemployment.

Likewise, an assumption was made that the high employment forecasts would result in a commensurate increase in the number of household workers. This was accomplished through decreases in the number of 0 worker households and 1 worker households, and increases in 2+ worker households to maintain the statewide control total of households. Low-income households were assumed to decrease and middle-income households were assumed to increase. Table C.2 shows examples of the resulting joint distributions of households by number of workers and income group and the resulting factors for the base, low, and high employment scenarios.

For the 2029 and 2033 forecast years for the 2018 Business Plan, the economic cycle offsets were estimated by interpolating the implied economic cycle offsets for 2025 and 2040 (after adjusting for updated socioeconomic forecasts). The BPM-V3 model runs needed to generate the information necessary for the analysis were originally made for 2025 and 2040 in order to bracket potential forecast years. Since the same assumptions have been used for the minimum and maximum growth scenarios for each forecast year, this interpolation approach was more direct and produced results consistent with those that would have been obtained had the minimum and maximum scenario forecasts been performed for 2029 and 2033. Table C.3 shows the ranges of offsets for and the implied annual round trip per capita trip rates for 2029, 2033, and 2040.

Table C.2 Workers per Household by Income Group for Most Likely, Minimum, and Maximum Changes in Employment for 2040

	Income Group			
Workers/ Household	Low	Middle	High	Total
Base Scenario				
0	17%	9%	6%	**32%**
1	9%	12%	12%	**33%**
2+	3%	9%	23%	**35%**
Total	29%	30%	41%	**100%**
Minimum Economic Growth Scenario				
0	19%	11%	6%	**36%**
1	10%	14%	12%	**36%**
2+	2%	8%	18%	**28%**
Total	31%	33%	36%	**100%**
Maximum Economic Growth Scenario				
0	14%	8%	6%	**28%**
1	7%	11%	13%	**31%**
2+	2%	10%	29%	**41%**
Total	23%	29%	48%	**100%**

Table C.3 Minimum, Most Likely, and Maximum Economic-Cycle Trip Frequency Constant Offsets and Implied Trip Rates

		Constant Offsets			Implied Annual Per Capita Round Trip Rates		
Model Year	**Purpose**	**Minimum**	**Most Likely**	**Maximum**	**Minimum**	**Most Likely**	**Maximum**
2029	Business/Commute	-0.17607	0	0.17275	1.85	2.21	2.63
	Recreation/Other	-0.06278	0	0.06613	5.50	5.86	6.26
	Total	–	–	–	***7.35***	***8.07***	***8.89***
2033	Business/Commute	-0.17633	0	0.16865	***1.91***	***2.28***	***2.70***
	Recreation/Other	-0.06420	0	0.06479	***5.58***	***5.95***	***6.35***
	Total	–	–	–	***7.49***	***8.23***	***9.05***
2040	Business/Commute	-0.17680	0	0.16148	2.06	2.46	2.89
	Recreation/Other	-0.06670	0	0.06243	5.87	6.27	6.67
	Total	–	–	–	***7.93***	***8.73***	***9.56***

C.3 Trip Frequency Constant Ranges

For full model risk analysis runs, economic-cycle effects are included with the unexplained variation in the range specified for the trip frequency constant. The range of constant offsets for the uncertainty analysis is directly related to the calibrated constants. The range of constant offsets for impacts of economic cycles provides proxies for the actual economic-cycle risk variable being considered. This approach provides a useful method for specifying a continuous range of outcomes rather than developing multiple input socioeconomic datasets. The offsets must be combined to represent the full range of possible outcomes for the development of the risk analysis regression equations. The constant offsets for the Unexplained Variation and Economic Cycle are added, and the implied range of trip rates was estimated, as shown in Table C.4.

For the Monte Carlo risk analysis, each component of the trip frequency constant is considered as a separate risk variable with completely independent distributions (i.e., 0 percent correlation). The unexplained variation uses a PERT distribution, while the economic cycle component uses a triangular distribution. A 50-percent correlation is assumed between the business/commute and recreation/other risk components for unexplained variation, since there is likely to be some relationship (though not perfect correlation) in changes to overall trip-making for different purposes. Perfect correlation is assumed between economic-cycle risk components for business/commute and recreation/other purposes.

Table C.4 Range of Trip Frequency Constant Offsets and Implied Trip Rates for Full Model Runs

Model Year	Purpose	Composite Trip Frequency Model Constant Offsets			Implied Trip Rates Based on Composite Constant Offsets		
		Minimum	Most Likely	Maximum	Minimum	Most Likely	Maximum
2029	Business/Commute	-0.45059	0	0.44728	1.41	2.21	3.44
	Recreation/Other	-0.19516	0	0.19852	4.83	5.86	7.12
	Total	–	–	–	6.24	8.07	10.56
2033	Business/Commute	-0.45085	0	0.44318	1.46	2.28	3.54
	Recreation/Other	-0.19659	0	0.19717	4.90	5.95	7.22
	Total	–	–	–	6.36	8.23	10.76
2040	Business/Commute	-0.45132	0	0.43600	1.57	2.46	3.79
	Recreation/Other	-0.19908	0	0.19481	5.15	6.27	7.59
	Total	–	–	–	6.72	8.73	11.38

Appendix D. Auto Operating Cost

The approach for forecasting auto operating costs for the 2018 Business Plan is consistent with the methodology used for the 2016 Business Plan, with updates to recognize the following:

- The most current motor gasoline and electricity price projections based on U.S. Energy Information Administration's (EIA) 2017 Annual Energy Outlook (AEO).
- The most current projections of the market penetration of electric vehicles.
- Revised non-gasoline operating costs.
- The most current fuel efficiency projections of the on-the road vehicle fleet.
- Effects of Cap and Trade rules in motor fuel prices and potential effects of an increase in the Federal excise tax rate.

The auto operating costs documented in this appendix are for privately owned vehicles. Appendix F provides background on auto operating costs for autonomous and shared use vehicles and their impacts on overall auto operating costs as used for the 2040 Phase 1 – Blended risk analysis.

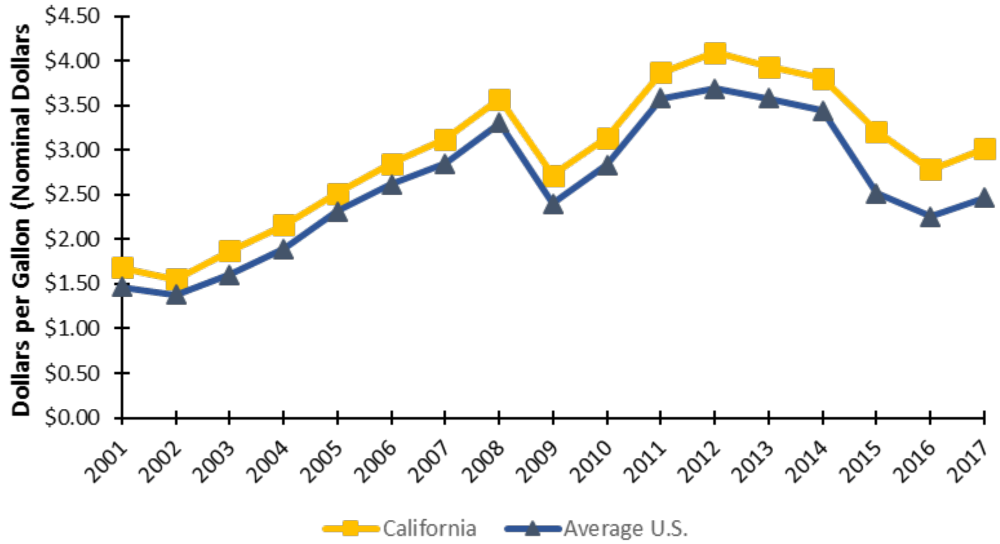
The following sequential steps were undertaken to calculate the auto operating cost:

1. Project retail fuel and electricity prices in California.
2. Project the market penetration rate for electric vehicles in California.
3. Adjust additional fees and charges based on two scenarios:
    a. Cap and Trade.
    b. Potential increase in Federal excise tax.
4. Project fuel economy of the electric and nonelectric "on the road" fleet.
5. Estimate nonfuel costs.
6. Combine fuel operating cost with nonfuel operating cost.

D.1 Fuel Prices

Historically, California retail gasoline prices have been higher than the U.S. average. As shown in Figure D.1, from year 2000 to 2017, the overall average for California prices was consistently 12.8 percent higher than the U.S. average.

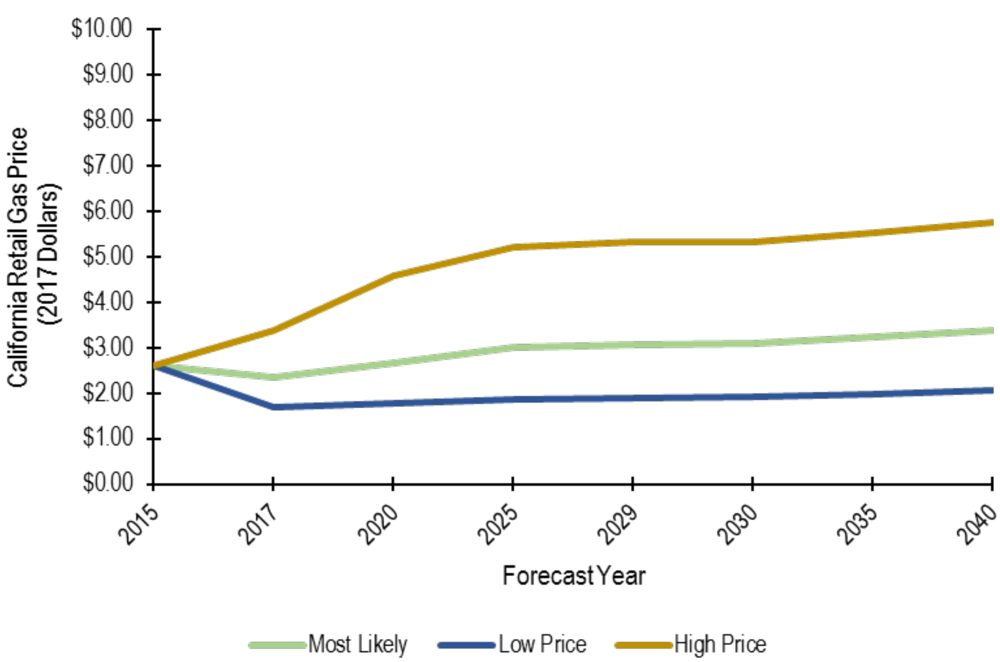
Figure D.1 Annual Retail Gasoline Prices



Source: U.S. Energy Information Administration: Annual All Grades All Formulations Retail Gasoline Prices http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_m.htm.

The EIA forecasts motor gasoline prices through 2050 for three different scenarios in its 2017 Annual Energy Outlook (AEO): reference, low, and high. The projections have been increased by 12.8 percent to develop projections of retail gas prices in California and are shown in Figure D.2.

Figure D.2 Low, Reference, and High California Retail Gas Price



Source: EIA, AEO2017 National Energy Modeling System.

D.2 Cap and Trade Effects on Fuel Prices

On January 1, 2015, the Cap and Trade rules came into effect for the fuel sector in California. The California Legislative Analyst's Office (LAO) estimated in 2017 that Cap and Trade could add $0.15 to $0.63 per gallon to retail gasoline prices in 2021 (Table D.1). The ranges were generated from the auction reserve price (ARP) and allowance price containment reserve price (APCRP) (i.e., the minimum and maximum prices per ton of carbon dioxide allowed at the Cap and Trade auctions). The LAO projected these figures out to 2031, and these figures were adopted as estimates of the minimum and maximum marginal impact of Cap and Trade on gasoline prices.

Table D.1 Gasoline Price Increase Due to Cap and Trade, LAO Estimate

	2021	2026	2031
Minimum (ARP)	$0.15	$0.19	$0.24
Maximum (APCRP)	$0.63	$0.67	$0.73

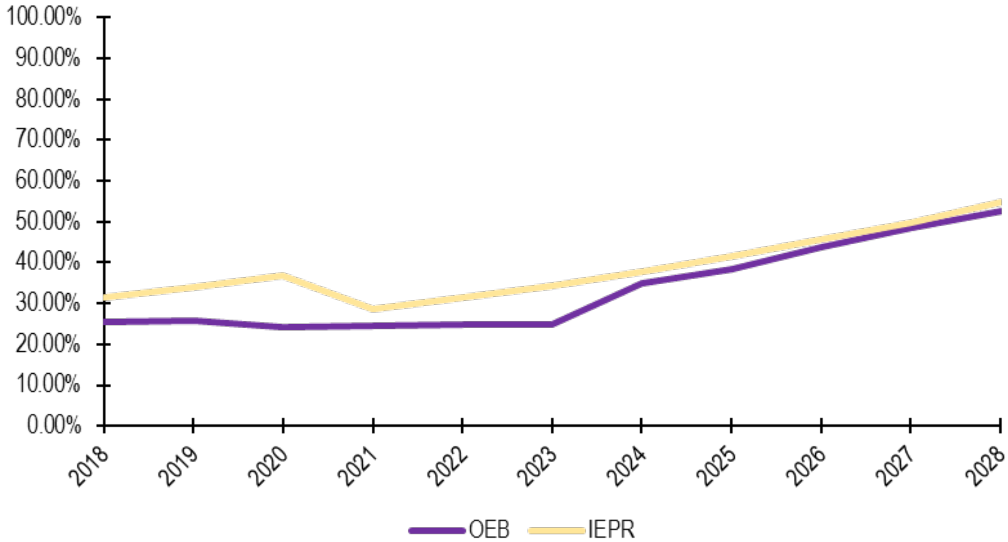
Source: March 29, 2017 Letter from California Legislative Analyst's Office to Assembly Member Vince Fong. http://www.lao.ca.gov/letters/2017/fong-fuels-cap-and-trade.pdf.

Independent projections of the long-term price of carbon on the Cap and Trade auction market provided estimates of the most likely price per ton of carbon as a function of the maximum price through 2028. These projections were created by consultants to the Ontario Energy Board (OEB) and the Integrated Energy Policy Report (IEPR) produced by the California Energy Commission. The results are found in Figure D.3.

The three forecast years represent, respectively, a year from the midst of the projected years, a year from the upper edge of the projected years, and a year well outside the OEB/IEPR forecast. Figure D.3 shows that, although the most likely price of carbon is only about one-third of the maximum price in 2025 that percentage rises by 2028.

Thus, to determine the most likely value for the impact of Cap and Trade, the maximum per gallon impact of Cap and Trade is interpolated from the 2017 LAO letter, and then multiplied by the average of the two percentages in Figure D.3 used for 2029 and 2033. By 2040, it is assumed that the price of carbon will stabilize at the midpoint between the minimum and maximum values. These marginal impacts are provided in Table D.2.

Figure D.3 Most Likely Price per Ton of Carbon as a Percentage of the Maximum Price per Ton of Carbon



Source: California Energy Commission. "Preliminary GHG Price Projections – Energy Assessment Division." December 19, 2016.
ICF Consulting Canada. "Long-Term Carbon Price Forecast Report." Submitted to Ontario Energy Board. Last updated July 19, 2017.

Table D.2 Minimum, Maximum, and Most Likely Marginal Cost per Gallon of Gasoline Due to Cap and Trade *(June 2017 Dollars)*

	2029	2033	2040
Minimum	$0.23	$0.26	$0.33
Most Likely	$0.39	$0.52	$0.58
Maximum	$0.73	$0.77	$0.84

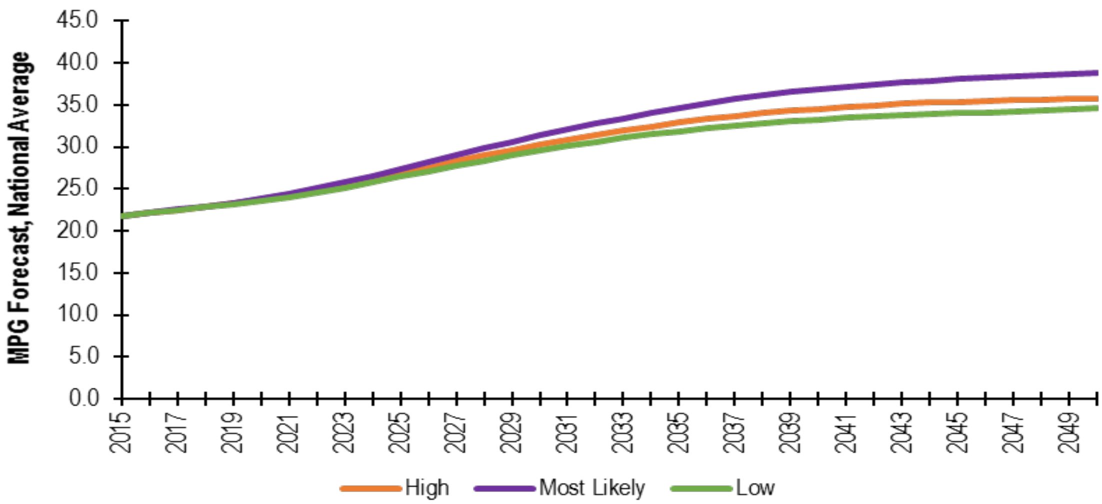
D.3 Federal Fuel Tax Increase Scenario

For the maximum auto operating cost scenario only, it is assumed that the Federal Government introduces a bill that links the Federal fuel tax to the Consumer Price Index. Today, the Federal Fuel Tax is $0.184 per gallon. If the Federal Fuel Tax is increased based on adjustment for Consumer Price Index (CPI) changes, which are assumed at 2.4 percent per year increase retroactive to year 1993 (i.e., last gas tax increase), then the Federal Fuel Tax would be $0.322 per gallon today. This results in the maximum scenario adding an additional $0.14 tax to the Fuel Cost projection (i.e., $0.32 – $0.18 = $0.14).

D.4 Projections of Fuel Economy of Light-Duty Vehicles

U.S. National Average, shown in Figure D.4, is used for the assumptions of Fuel Economy projections in California.[38] For calculating the minimum auto operating cost, the high miles per gallon (MPG) forecast was coupled with the low gasoline price forecast; and for the maximum auto operating cost, the low mpg forecast was coupled with the high gasoline price forecast. These fuel economy forecasts were used to calculate the fuel economy of only the non-electric portion of the vehicle fleet.

Figure D.4 National Average Fuel Economy Forecasts

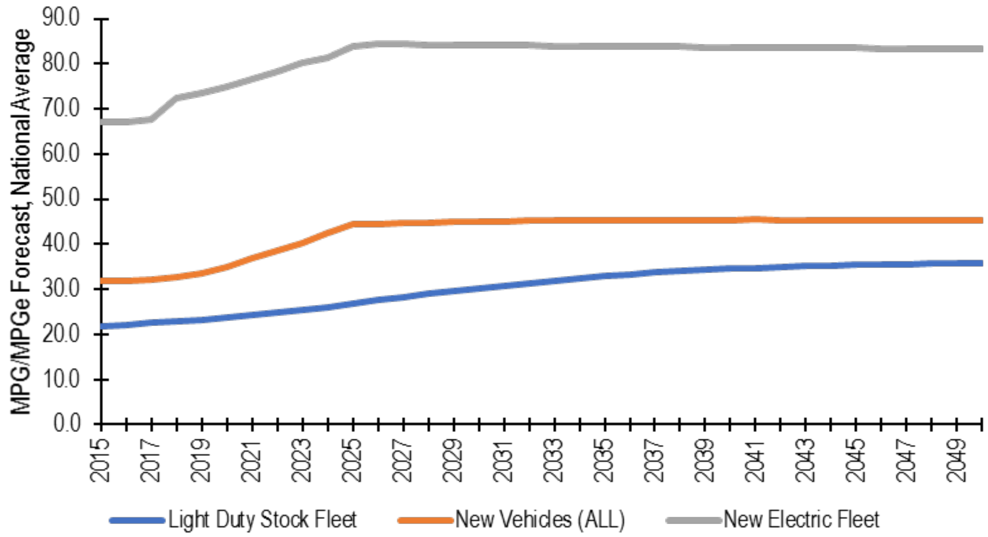


D.5 Projections of Fuel Economy of Electric Vehicles and Market Penetration

The 2017 Annual Energy Outlook provides an estimate for the fuel efficiency of new vehicles sold in each year of the forecast, including electric and other alternative fuel vehicles. It does not, however, include an estimate for the fuel efficiency of the on-the-road fleet of alternative fuel vehicles. In order to capture the higher fuel efficiency of the electric fleet, the equivalent miles per gallon efficiency of the electric fleet are set at 2.5 times the projected fuel efficiency of the Light-Duty Stock Fleet, based on EIA data. This is reasonable given the trends in Figure D.5, where the average fuel efficiency of all new electric vehicles is 2.6 times that of the stock fleet.

[38] U.S. Energy Information Association. 2017. Annual Energy Outlook 2017 with projections to 2050. DOE/EIA-0383 (2015), April 2017.

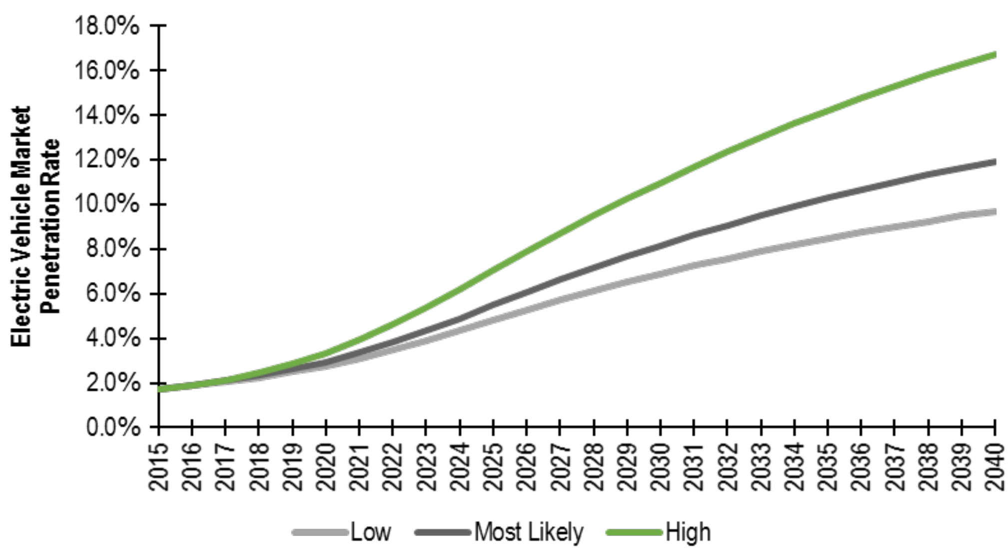
Figure D.5 National Fuel Economy Forecasts, Including Light-Duty Stock Fleet, All New Vehicles Sales, and New Electric Vehicle Sales



Source: EIA, AEO2017 Reference Case.

The market penetration of electric vehicles is estimated as the number of total electric fuel vehicles (including both cars and light trucks) divided by the total size of the stock vehicle fleet. Penetration rates were calculated under the 2017 AEO’s reference case, low oil case, and high oil case. The results are provided in Figure D.6.

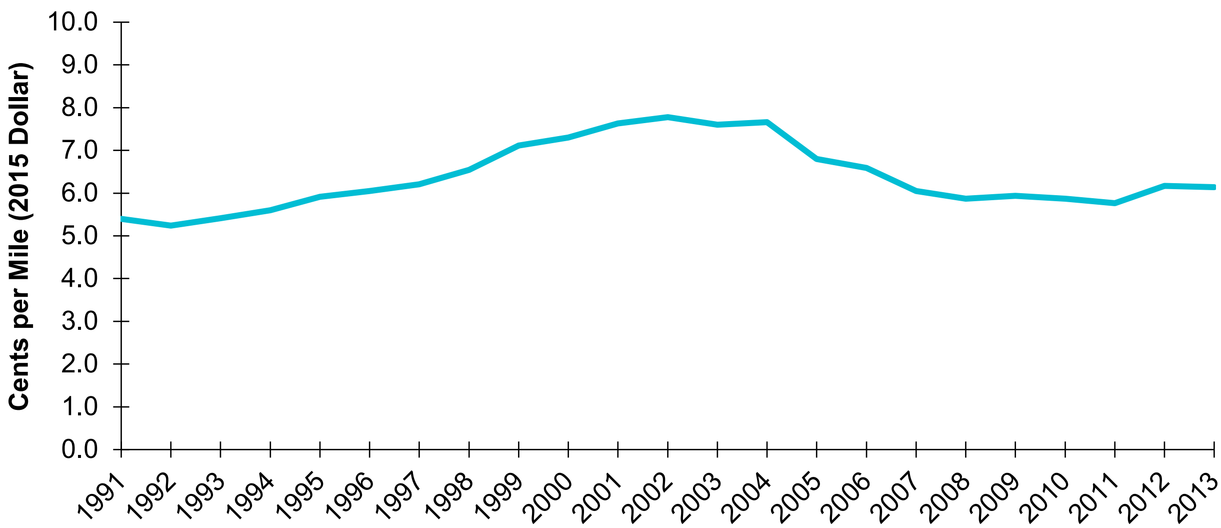
Figure D.6 Percent Market Penetration of Electric Vehicles by Year



D.6 Nonfuel Operating Vehicle Cost

The Bureau of Transportation Statistics (BTS) publishes historical average nonfuel auto operating costs. The total cost of owning and operating an automobile includes fuel, maintenance, tires, insurance, license, registration and taxes, depreciation, and finance costs. Figure D.7 illustrates the nonfuel auto operating cost per mile between 1991 and 2016. The low nonfuel auto operating cost scenario is calculated as the minimum nonfuel cost between 1991 and 2016 (i.e., 5 cents per mile). The high nonfuel auto operating cost scenario is calculated as the maximum nonfuel cost between 1991 and 2014 (i.e., 8 cents per mile). The most likely value is the current nonfuel auto operating cost (i.e., 7 cents per mile).

Figure D.7 Historical Nonfuel Operating Vehicle Cost



Source: CPI, BLS, All Urban Consumer, National Average: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_03_17.html.

D.7 Range of Auto Operating Cost

The following formulas were used to calculate the minimum, most likely, and maximum auto operating cost:

*Minimum Auto Operating Cost = (1 - %EVs) * (Low CA Gas Price + Low C&T Impact + No Increase in Federal Gas Tax) / High ICE Fuel Efficiency + %EVs * (Low CA Electricity Price * 33.7) / High EV Fuel Efficiency + Low Nonfuel Operating Costs*

*Most Likely Auto Operating Cost = (1 - %EVs) * (Most Likely CA Gas Price + Avg(Low C&T Impact, High C&T Impact) + No Increase in Federal Gas Tax) / Most Likely ICE Fuel Efficiency + %EVs * (Most Likely CA Electricity Price * 33.7) / Most Likely EV Fuel Efficiency + Most Likely Nonfuel Operating Costs*

*High Auto Operating Cost = (1 - %EVs) * (High CA Gas Price + High C&T Impact + Increase in Federal Gas Tax) / High ICE Fuel Efficiency + %EVs * (High CA Electricity Price * 33.7) / High EV Fuel Efficiency + High Nonfuel Operating Costs*

gives the auto operating cost component values and the resulting minimum, most likely, and maximum auto operating cost for each forecast year before adjusting for the impact of autonomous and shared vehicles for 2040 forecasts.

Table D.3 Range of Auto Operating Cost for each Forecast Year by Auto Operating Cost Component *(June 2017 Dollars)*

	Minimum	Most Likely	Maximum
2029 Auto Operating Cost ($/mile)	**$0.17**	**$0.23**	**$0.35**
U.S. Gas Price ($/gal)	$1.90	$3.06	$5.33
California Gas Price ($/gal)	$2.14	$3.45	$6.01
California Electricity Price ($/kWH)	$0.17	$0.17	$0.18
% Electric Vehicles	10.51%	7.86%	6.70%
MPG	31.5	30.5	29.8
MPGe	78.64	76.14	74.53
Nonfuel cost ($/mi)	$0.10	$0.11	$0.12
Cap and Trade ($/gal)	$0.23	$0.39	$0.73
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14
2033 Auto Operating Cost ($/mile)	**$0.17**	**$0.23**	**$0.34**
U.S. Gas Price ($/gal)	$1.95	$3.19	$5.52
California Gas Price ($/gal)	$2.20	$3.59	$6.22
California Electricity Price ($/kWH)	$0.18	$0.18	$0.18
% Electric Vehicles	13.35%	9.75%	8.10%
MPG	34.3	32.8	31.9
MPGe	85.84	82.11	79.85
Nonfuel cost ($/mi)	$0.10	$0.11	$0.12
Cap and Trade ($/gal)	$0.26	$0.52	$0.77
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14
2040 Auto Operating Cost ($/mile)[a]	**$0.17**	**$0.23**	**$0.33**
U.S. Gas Price ($/gal)	$2.05	$3.39	$5.75
California Gas Price ($/gal)	$2.32	$3.82	$6.49
California Electricity Price ($/kWH)	$0.19	$0.19	$0.19
% Electric Vehicles	17.16%	12.23%	9.96%
MPG	37.9	35.5	34.2
MPGe	92.76	88.79	85.55
Nonfuel cost ($/mi)	$0.10	$0.11	$0.12
Cap and Trade ($/gal)	$0.33	$0.58	$0.84
Federal Gas Tax Increase ($/gal)	$0.00	$0.00	$0.14

Appendix E. Coefficient on Transit Access-Egress Time/Auto Distance Variable

Between some regions in California, especially in the Silicon Valley to Central Valley scenario, individuals who wish to travel primarily by transit to reach their destination must transfer from a high-speed rail (HSR) bus or conventional rail (CVR) system before or after traveling on HSR. There is uncertainty around how the need to make these transfers affects the overall desirability of traveling by HSR. The uncertainty in the desirability of travel by HSR, when the CVR or HSR bus leg of the journey is relatively long in relation to the HSR travel length, has an impact on ridership and revenue. Thus, this uncertainty was included as a potential risk variable.

E.1 Options for Addressing Risk in Uncertainty Analysis

Two primary options were considered for addressing the transit transfer concern in the context of the risk analysis. The first option considers a range for the constant associated with the transit access/egress to the HSR main mode. The main advantage of this approach is its simplicity. The range used for the constant would come directly from conversion of a penalty value (in minutes) to utility. The main disadvantage is that the same range would need to be applied to all transfers between access/egress transit modes and HSR. This means that the penalty would apply equally to transfers between local transit (e.g., someone taking a city bus from their home to the station) and HSR, and transfers between CVR or HSR bus and HSR with longer access trips. Transfer between local transit and CVR exists today, and thus are accounted for within the model estimation of this variable, while transfers between CVR or HSR bus to HSR have not been observed in the estimation dataset. Moreover, it means the penalty would not vary on the basis of how long the trip was or how much of the trip was transit versus HSR.

The second option considers a range for the parameters associated with transit access/egress travel times relative to origin-destination (OD) distances. This variable appears in the access and egress modal utility functions as follows:

$$\beta \times \text{MAX}\left(0, \frac{[Acc\ or\ Egr\ Time]}{[OD\ Distance]} - Threshold\right)$$

In the base model, several threshold parameter options were tested in model estimation, and a value of 0.2 was ultimately identified. The values of beta (the variable coefficient) were estimated directly and were found to be negative. Separate coefficients were estimated for auto access/egress modes versus nonauto access/egress modes (transit and walk/bike), with the magnitude of auto coefficients estimated to be much larger. This variable provides a disincentive for selecting a main mode that requires a long access or egress time, relative to the entire trip length.

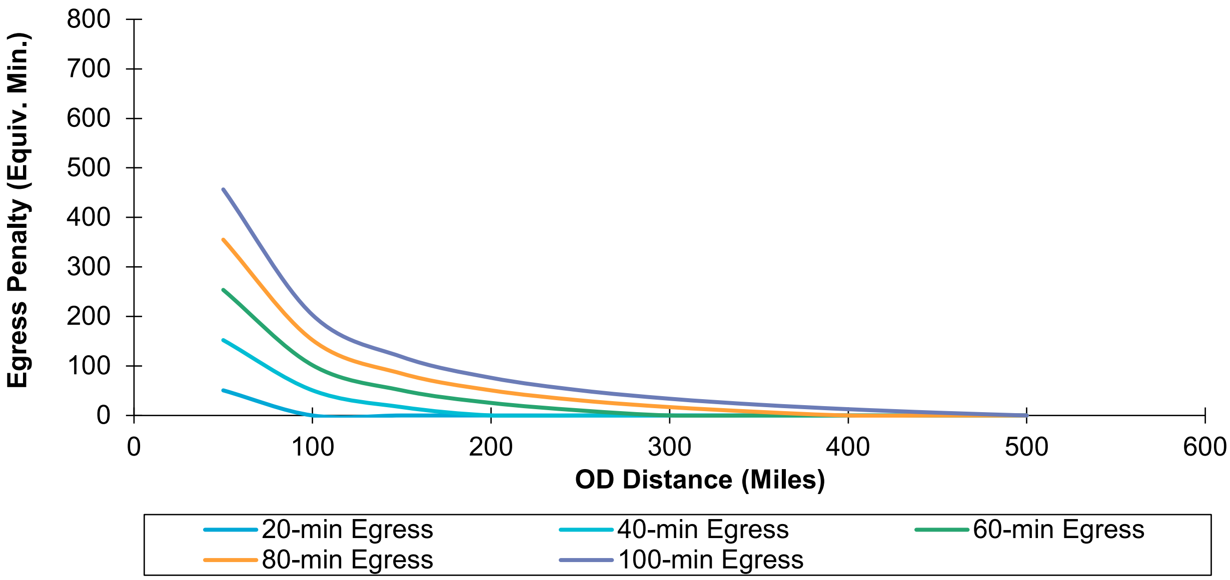
The main advantage of the second option is that this differentiation would naturally occur between local transit and longer CVR or HSR bus connections. Since local transit connections would typically be very short distance and CVR or HSR bus may be short or long distance, the "penalty" associated with transit access/egress would reflect the access/egress mode's overall share of the total trip length. The second option is more appropriate for the risk analysis. The uncertainty associated with the variable is only applied for the HSR main mode (i.e., not air or CVR).

E.2 Development of the Range in the Risk Variable Parameters

Figure E.1 and Figure E.2 show the variable's effect under the current model specification (in terms of equivalent minutes of travel time [39]) for the recreation/other purpose (the results are very similar for business and commute trip purposes). The first plots penalty versus OD distance for constant egress times, and the second plots penalty versus egress time for constant OD distance values. The same concepts apply to the access end of trips. The egress end is shown only as an example; the access time graph is identical.

In both figures, certain regions of the graphs suggest very high penalties for certain types of trips. For instance, Figure E.1 shows very high penalties for the 100-minute egress line when OD distance is less than 100 miles. Likewise, Figure E.2 shows very high penalties for the 50-mile OD distance line when egress time is high.[40] Travelers typically do not make trips of this nature, since other main modes would be highly favored, so these penalty values are very unlikely to be applied in practice.

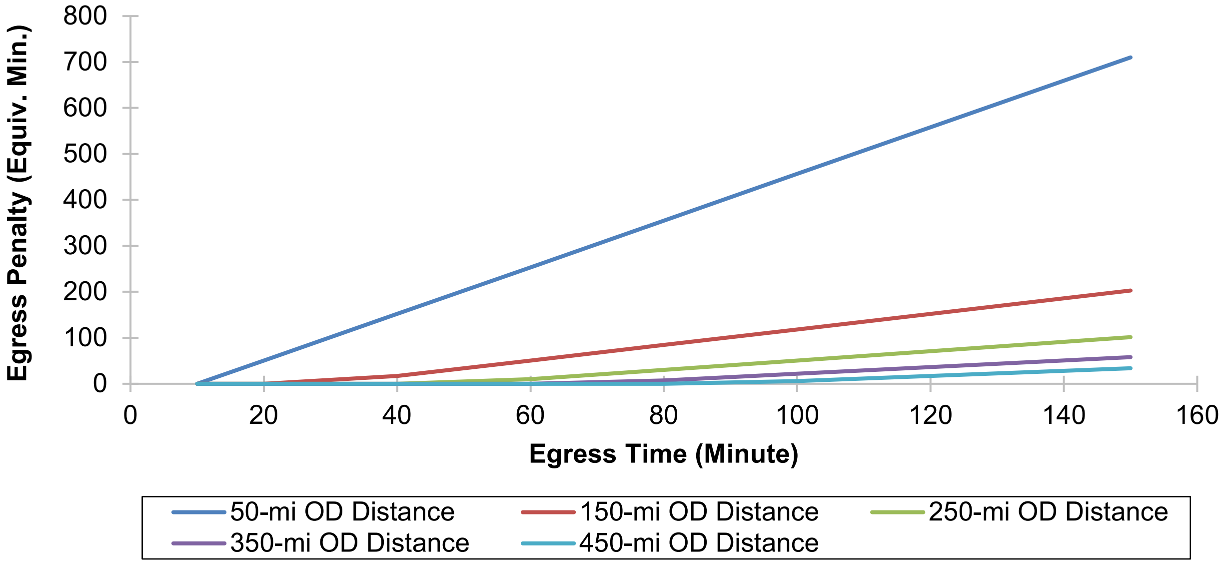
Figure E.1 Penalty versus Origin-Destination Distance for Constant Egress Times



[39] "Equivalent minutes of travel time" is estimated by dividing a constant or a variable by the coefficient associated with travel time. Equivalent minutes of travel time provides a convenient way to measure the magnitude of "unexplained variation" of a model constant using an understandable metric, and to compare values among different models. Equivalent minutes of travel time is a derived measure that can be computed for any model variable. So, for example, a $72 HSR fare (2005 dollars) for an interchange in the recreation/other mode choice model would equate to 337 equivalent minutes of travel time, while the implied equivalent minutes of travel time savings for group travel in an auto for the interchange would equate to a savings of 619 equivalent minutes of travel time. Note, however, these variables are important for their contributions to the mode choice utility function, not as direct measures of travel time.

[40] A chart of penalty versus OD distance for constant access time would look identical to the chart for egress time.

Figure E.2 Penalty versus Egress Time for Constant Origin-Destination Distance Values



The variable parameter range was developed using a French HSR experience as a guide. In 1981, SNCF (the French railway company) transitioned a 350-mile direct CVR route between Paris and Grenoble to a HSR trip between Paris and Lyon and then a CVR access/egress trip between Lyon and Grenoble. The original CVR trip took 300 minutes, but the new HSR trip required 120 minutes of travel on HSR and 90 minutes of access/egress time on CVR. [41] This change saved travelers approximately 90 minutes of travel time, but did not increase ridership between Paris and Grenoble.

We proposed a 90-minute penalty as a rough benchmark for determining a lower bound on the model parameters since ridership did not increase in the French example even with the 90-minute time savings. Using an OD distance of 350 miles, an egress time of 90 minutes, and the aforementioned 90-minute savings to estimate a 90-minute penalty, several approaches were tested to achieve an appropriate lower bound for the variable. The approaches were based on the ways the variable could be affected by uncertainty. The first is the effect on the coefficient associated with the variable. The second is the effect on the threshold variable which, in the BPM-V3, is set such that the variable takes a value of zero when the ratio of access/egress time to distance is less than 0.2. The threshold value was set in model estimation by trial and error. The value of 0.2 was selected for the BPM-V3 because it fit the data better than other potential values.

Several options were considered for setting lower bounds for the threshold variable. Based on a review of potential options, threshold values of 0.05 and 0.10 were tested. In both cases, the coefficient on the variable was selected so that the value of the penalty was about 60 minutes for a case similar to the French example. A 60-minute penalty was used instead of the 90-minute penalty observed in the French

[41] The OD distance and egress times cited for the French experience are approximate, as it is based on Google maps and train timetables. While the network distance is about 350 miles between Paris and Grenoble, the straight-line distance is only 300. And, some egress train options took longer than 90 minutes, up to and over 120 minutes.

experience, because it offered more reasonable model behavior overall, and it was not desirable to change the long-distance models in unreasonable ways to match a single observed data point. Figure E.3 and Figure E.4 plot the penalty versus egress time to OD distance ratios for baseline, drive access/egress variables, transit access/egress variables with threshold value of 0.05, and transit access/egress variables with threshold value of 0.10. Figure E.3 shows the results for the business/commute purpose, and Figure E.4 shows results for the recreation/other purpose. The drive access/egress variable is plotted for comparison purposes only, and has no bearing on the variable discussed in this section. It applies when the access/egress mode is an auto mode (rather than transit).

Figure E.3 Business/Commute Penalty versus Egress Time to Origin-Destination Distance Ratios for Baseline, Drive Access/Egress Variables, and Transit Access/Egress Variable Options

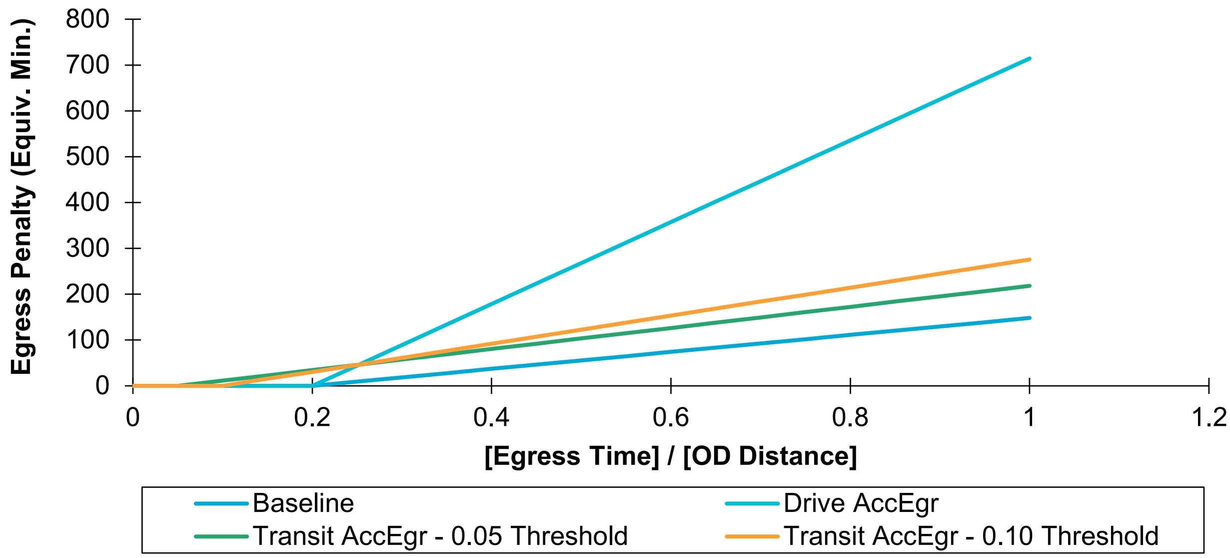
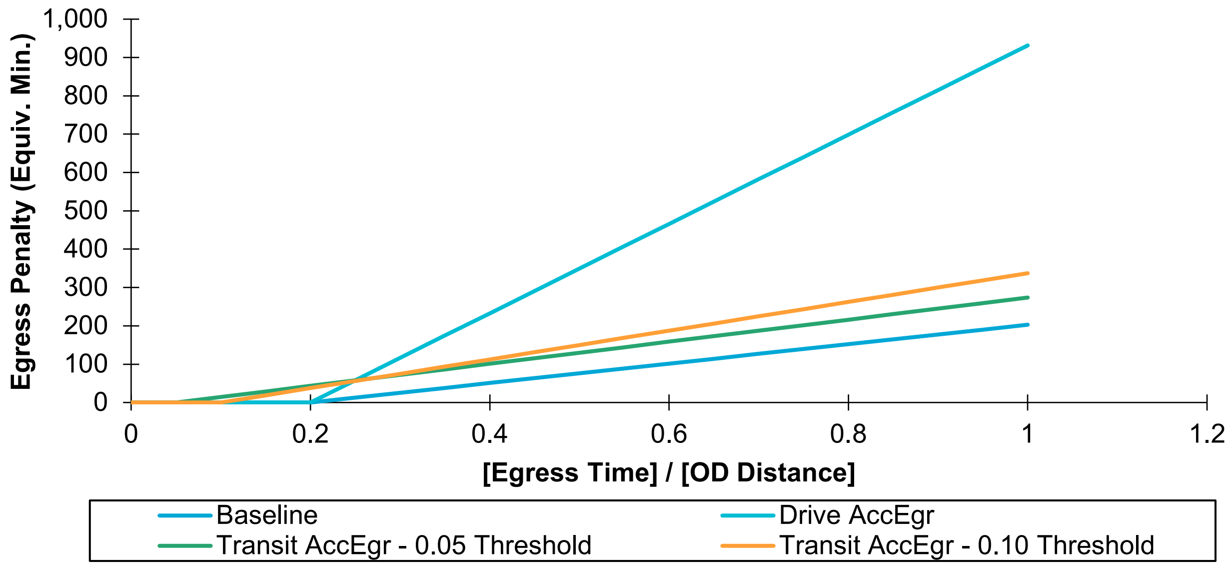


Figure E.4 Recreation/Other Penalty versus Egress Time to Origin-Destination Distance Ratios for Baseline, Drive Access/Egress Variables, and Transit Access/Egress Variable Options



E.3 Range of Coefficient on Transit Access-Egress Time/Auto Distance Variable

The transit access/egress variable with threshold value of 0.10 was chosen as the low scenario. This threshold was chosen over 0.05, because it causes less disruption to the relationships between the drive and transit access/egress variables for trips with shorter access/egress (e.g., when the ratio of access or egress time to OD distance is around 0.1 to 0.2). The minimum coefficient value is set to -2.0 for business/commute purpose and -1.3 for recreation/other purpose. These are set to achieve penalty values of 51 and 66 minutes. These penalty value benchmarks come from the penalties the model suggests for the French scenario for drive access/egress modes. The lower bound on the transit penalty should not exceed the penalty suggested by the model for drive access/egress modes. A 51-minute and 66-minute penalty was used instead of the 90-minute penalty observed in the French experience because it offered more reasonable model behavior overall, and it was not desirable to change the long-distance models in unreasonable ways to match a single observed data point. The coefficient and threshold values vary in parallel (i.e., perfect correlation) for the full model runs and Monte Carlo simulation.

The maximum threshold and coefficient values are set to be identical to the calibrated base/most likely values since there is no evidence to suggest that the penalty to transfer from transit to HSR should be less than the penalty used for CVR and air that was developed based on observed data.

Appendix F. Quantifying the Effects of Autonomous and Shared Use Vehicles on Year 2040 Risk Variables

By 2040, it is likely that autonomous vehicles (AV) and shared-use vehicles will compose some share of automobile travel. AVs could have important features that change the auto mode's perception among travelers, while the increase in shared-use vehicles could directly affect the auto operating cost of travelers, which may impact HSR ridership and revenue. The risk analysis framework considers three key features of the auto mode that might change due to AVs and shared-use vehicles: 1) auto in-vehicle time, 2) auto operating costs, and 3) travelers' perceptions of the disutility of travel time in AVs.

F.1 Autonomous Vehicle Background and Research

AVs represent three potential risks to the revenue and ridership forecasts. First, AVs may reduce auto travel times. Second, AVs may decrease the operating cost of autos. And third, AVs may decrease the disutility of in-vehicle time because travelers are able to focus on activities other than driving.

One potential impact of AV technology is the reduction of travel times and improvement of travel speeds by connecting vehicles, allowing them to travel much closer to one another at high speeds, thus effectively increasing capacity and reducing congestion. The bulk of the travel time benefits from AVs require AVs to make up a majority of all cars on the road, with peak benefits achieved only after AV market penetration reaches about 75 percent. It is also possible that AVs could contribute to congestion in the near term, depending on the programs that control them and how well they are able to interact with non-AVs.[42] Also, especially in urban areas, increased congestion could be caused by 0-occupant AVs traveling to pick up passengers or returning to remote parking locations. Auto travel time was included as a risk variable in the 2016 Business Plan, but its effect on high-speed rail ridership was minimal and thus was not ultimately included in the 2018 Business Plan.

AVs may decrease auto operating costs via better gas mileage, lower insurance premiums if crashes can be reduced and reduced parking costs, as AVs could potentially drop a passenger off and find free or cheaper parking. The possibility of increased vehicle miles traveled (VMT) due to travel with no passenger (e.g., to park) could effectively increase operating costs, though this would require it being legal for AVs to travel without an operator, which could be further into the future than 2040.

Because travelers will be able to engage in other activities in AVs (e.g., checking email, reading, or even sleeping), AVs offer the possibility that being in one's car may be less onerous. This will be accommodated in the model by adjusting the in-vehicle time coefficient associated with the auto mode, as described in Section 3.10.

[42] Litman, Todd. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. February 27, 2015. Victoria Transport Policy.

F.2 Autonomous Vehicle Market Penetration Assumptions

AV market penetration is a key risk variable that informs both the uncertainty in auto in-vehicle coefficient and the uncertainty in auto operating costs. For instance, if market penetration of AVs is 0 percent, then we expect no change to auto in-vehicle coefficient or operating costs. However, if market penetration is 50 percent, we expect a less onerous automobile experience for those using AVs and some effect on operating costs. Several recent studies and papers have provided a range of suggestions:

- In 2014, it was suggested that under the right circumstances, AVs could represent 50 to 75 percent of the auto market by 2035 to 2045[43];
- A 2015 forecast suggested that AVs will have 30 percent market penetration in the 2040s (but 40 percent of all travel), 50 percent market penetration in the 2050s, and 75 percent market penetration will occur sometime after 2060[44]; and
- An alternative 2015 forecast suggested that market penetration will be between 1 percent and 11 percent by 2030 and 7 percent and 61 percent in 2050, depending on a number of factors.[45]

Based on this research, it is assumed that the market penetration of autonomous vehicles among the owned vehicle market in 2040 is a triangular distribution with minimum 10 percent, maximum 75 percent, and most likely 35 percent.

F.3 Shared-Use Vehicle Market Penetration Assumptions

The shared-use market penetration was calculated using a series of assumptions.[46] It was asserted that the long-distance trip shared-use market would vary by area type of the household, with households in denser areas being more likely to use shared-use vehicles. For each area type, a low, most likely, and high value of shared-use vehicle usage was asserted based on professional judgment. From those assertions, a weighted low, most likely, and high value was computed based on long-distance trip shares, as shown in Table F.1. The market penetration share is assumed to have a triangular distribution.

[43] Bierstedt, J., A. Gooze, C. Gray, J. Peterman, L. Raykin, and J. Walters, 2014. Effects of Next Generation Vehicles on Travel Demand and Highway Capacity by FP Think Working Group Members. FP Think Working Group.

[44] Littman, T., 2015. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. February 27, 2015. Victoria Transport Policy.

[45] Milakis, D., M. Snelder, B. van Arem, B. van Wee, and G. Correia. 2015. Development of Automated Vehicles in the Netherlands: Scenarios for 2030 and 2050. Delft, The Netherlands: Delft University of Technology.

[46] The market penetration rates used in this analysis are in addition to the year 2010 (i.e., model calibration year) market penetration for shared-use vehicles used for long-distance trips, such as rental cars.

Table F.1 Shared-Use Market Penetration by Area Type

			Long-Distance Auto Trips Using Shared-Use Vehicles in Year 2040		
Area Type	Long-Distance Trips	Long-Distance Trip Share	Low	Most Likely	High
CBD-Bay Area	74,684	3%	0.2	0.3	0.5
Urban-Bay Area	10,0619	5%	0.1	0.2	0.45
CBD-Other	71,068	3%	0.1	0.2	0.45
Urban-Other	226,623	11%	0.05	0.1	0.35
Small Urban	157,145	7%	0	0.1	0.2
Suburban	1,055,053	49%	0	0	0.1
Rural	461,141	21%	0	0	0.1
Weighted Total			**2%**	**5%**	**20%**

F.4 Development of Autonomous Vehicle Auto Operating Cost Uncertainty

As discussed in Appendix D, the auto operating cost for privately owned non-AVs comprises different components. These components are treated together as one auto operating cost, which is referred to as OC_{Base}. Additional uncertainty was added to pertinent subcomponents, representing the uncertainty in auto operating costs due to AV adoption and shared-use vehicles. Key variables we considered were the level of AV market penetration and shared-use market share.

Because the marginal cost of trips made by shared-use vehicle will include additional costs over and above typical operating costs, those additional costs were considered. For instance, shared-use trips will be charged a surcharge, similar to a toll either based on the amount of time the vehicle is used or distance traveled. To keep the surcharge in the same units as auto operating cost, it is assumed the surcharge is based on distance traveled and would incur a charge per mile traveled. Ranges for cost per mile are predicted by Litman to be between $0.60 and $1.00 per mile, though this seems high given that current shared-use costs are on the order of $0.15 to $0.60 per mile.[47] Given that AVs may dominate this market and might have higher purchase prices, it is conceivable that the costs will be higher by 2040, but probably not as high as forecast by Litman. It is, therefore, assumed that shared-use cost per mile is a uniform distribution with minimum $0.18 and maximum $0.85 (2014 dollars).

AVs are predicted to drive in a more energy efficient manner compared to non-AV drivers due to a decrease in stop-and-go tendencies. Fuel economy could increase by as much as 23 to 39 percent.[48] It is

[47] Litman, T. 2015. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. February 27, 2015. Victoria Transport Policy.

[48] Eno Center for Transportation. 2013. Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations, Eno Center for Transportation, October 2013.

assumed that fuel economy improvements of AVs are uniform distribution with minimum 10 percent and maximum 50 percent. For the purpose of the risk analysis, the 50-percent improvement is a more conservative assumption than the Eno prediction. As discussed in Appendix D, fuel costs represent approximately 40 percent of the low auto operating costs, 50 percent of the base auto operating costs, and 66 percent of the high auto operating costs. It is assumed that only fuel costs would be affected by fuel economy improvements resulting from AV use.

The overall average auto operating cost is computed as a blended average for each market as follows:

$$OC_{avg} = OC_{nonAVnonSV} + OC_{AVnonSV} + OC_{nonAVSV} + OC_{AVSV}$$

Here, $OC_{nonAVnonSV}$ is the portion of operating costs attributable to owned non-AVs and is computed as:

$$OC_{nonAVnonSV} = (1 - S_{SV}) * (1 - S_{AV}) * OC_{Base}$$

S_{SV} is the market share of long-distance trips that use a shared vehicle (i.e., $(1 - S_{SV})$ is the market share of long-distance trips that use a nonshared vehicle); S_{AV} is the market penetration of AVs among nonshared use vehicles; and OC_{Base} is the base value of operating cost that comes from the distribution described in Appendix D for other model years.

OC_{AV} is the portion of operating costs attributable to owned AVs and is computed as:

$$OC_{AVnonSV} = (1 - S_{SV}) * S_{AV} * OC_{Base} * \left(\left[1 - S_{Base,FC}\right] + \left[\frac{S_{Base,FC}}{1 + FE_{AV}}\right]\right)$$

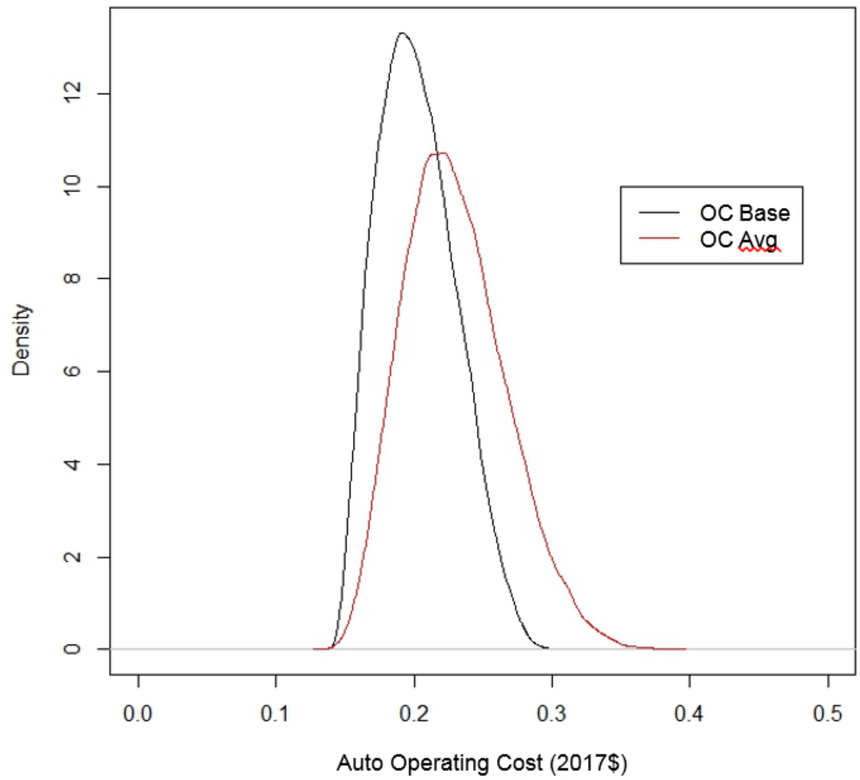
$S_{Base,FC}$ represents the share of base auto operating cost attributable to fuel. FE_{AV} is the fuel economy improvements achieved by AVs, on average. $OC_{sv}OC_{nonAVSV} + OC_{AVSV}$ is the contribution of shared-use vehicles to average auto costs and is computed as follows:

$$OC_{nonAVSV} + OC_{AVSV} = S_{SV} * (1 - S_{AV}) * CPM_{sv} + S_{SV} * S_{AV} * CPM_{sv}$$

CPM_{sv} is the cost per mile surcharge of shared vehicles.

Figure F.1 shows the distribution of auto operating costs in 2040. The black line corresponds to the distribution of the base auto operating costs; and the red line is the overall average auto operating cost distribution, based on the first, blended equation outlined above. The minimum auto operating cost is 14 cents per mile, the most likely is 22 cents per mile, and the maximum is 39 cents per mile (June 2017 Dollars).

Figure F.1 Distribution of Auto Operating Costs in Year 2040



Appendix G. Exceptionally Long Access and Egress: Experience from Japan

Few international high-speed rail (HSR) stations exhibit comparable examples of exceptionally long access and egress as might be possible with California High Speed Rail, particularly in a Silicon Valley to Central Valley scenario. There is generally no significant population center anywhere in the developed world that is more than a three-hour drive from an airport. There is also no significant population center in Western Europe that is more than a three-hour drive from a high-speed rail station. Some cities in China have very long access/egress to the Chinese high-speed rail network, but generally all lines terminate in major cities, providing sufficient demand to saturate rail usage without any long access/egress travel.

The only case we identified with the requisite characteristics is the Hokkaido Shinkansen line in Japan. The Hokkaido Shinkansen in Japan, designed to ultimately connect Tokyo to Sapporo, is currently under construction. This line is partially open for revenue service today. Last year, service began from Tokyo and reached as far north as Sin-Hakodate-Hokuto, a rail station in the south of the Island of Hokkaido (see Figure G.1).

Figure G.1 Hokkaido Shinkansen High-Speed Rail Line in Japan

It is a 3.5-hour, 180-mile drive from central Sapporo (Japan's fifth largest city) to the Hakodate terminal, or about the same travel time via Conventional Rail (CVR). This is slightly farther than the Los Angeles to Bakersfield connection, but comparable to the travel from San Diego to Bakersfield (4 hours, 210 miles).

The City of Hakodate has a population similar in size to Bakersfield. Tokyo has about three times the population of Los Angeles.

Hokkaido (Sapporo) is a separate island from Honshu (Tokyo). Travel between them is only by air, sea, and rail (via tunnel). Driving from Sapporo to Tokyo requires an expensive and relatively slow ferry segment. The air travel market from Tokyo to Sapporo is the second largest city-to-city market in the world by passenger volume, with roughly double the passenger traffic of the Los Angeles to San Francisco market.

The Hokkaido line does not appear to draw many riders, even with the substantial air travel demand between the two locations and an expensive and slow auto alternative. Reports in the media note that, in its first 16 days of operation, it averaged a bit under 6,000 passengers per day (about 2 million annual passengers)’. The Hokkaido Railway Company (an arm of the Japanese government) expects to lose money on this line for at least a decade, until the connection through the mountains to Sapporo is completed.

We have not found any detailed forecasts for the Hokkaido line; however, the ridership numbers for the Shinkansen, or bullet train, appear to be consistent with the theory that the HSR ridership in the Tokyo to Sapporo market is negligible, and the majority of the riders today actually have origins or desinations in or around Hakodate. This finding supports the inclusion of the exceptionally long access/egress risk variable to help determine the impact on ridership requiring long access/egress.

Appendix H. Technical Details for the Application of GPR

Gaussian Process Regression (GPR) represents a well-accepted and widely used method for meta-model analysis of computer simulations. This appendix addresses some of the technical details of the GPR methodology as applied in this risk analysis.

H.1 De-trending

When the output exhibits a clear trend (i.e., can be modeled well with linear regression), it is typical to use both Linear Regression (LR) and GPR models simultaneously. First, an LR model is used to "de-trend" the data, generating a rough prediction of the outcome values at each observation point. Then, the GPR model is constructed from the input values and the residuals of the LR model; not the actual observed output values. By applying GPR on the residuals alone, the GPR is used to explain the localized deviation from the overall (linearly modeled) trend, and not the trend itself.

Applying GPR jointly with LR is advantageous for multiple reasons. First, the particular individual coefficients of the LR model may themselves be of interest, as they indicate the magnitude and direction of the basic high-level relationship between the input and output variables. When employed without an LR, the GPR model does not directly provide individual variable coefficients in the same manner. This prevents a simple evaluation of the face validity of the meta-model results (e.g., does ridership increase when fares decrease?) Second, most GPR kernels (the mathematical form used to express autocorrelation) assume the underlying process does not have a global trend. For predictions of outputs for inputs that are sufficiently distant from any observed data, the GPR expectation will tend to revert to the global mean outcome of the observed data, which may be very different from the trend line outcome, particularly for input values that are near (or beyond) the extreme values of the input factor space.

For these reasons, we have adopted the de-trending approach; whereby, ridership and revenue meta-models comprise an LR model that is refined by a GPR. When using this approach, GPR can be considered an extension and enhancement of LR, rather than a replacement. This approach replaces the LR assumption that errors are independent with the GPR assumption that errors are auto-correlated. Because we know this is the case for deterministic models, including Business Plan Model – Version 3 (BPM-V3), we are guaranteed to achieve a better fitting meta-model using this process than using LR alone.

H.2 Kernel Selection

GPR models must make an assumption about the mathematical form of the correlation between observations. Different assumptions lead to different models. These assumptions are embedded in the selection of the "kernel" function for the regression model. Various kernels make different assumptions about the smoothness and stability of the correlation, as well as the overall level of noise in the model. Most statistical modeling tools for GPR applications include a number of standard kernel functions, such as the radial basis function, Matérn, and others. The selection of a particular kernel function is generally made based on knowledge of the behavior of the underlying model, or through experimentation with the cross-validation of various kernels (see below).

For risk analysis of models where the relevant input variables compose a set measured in different units, it is important to use an anisotropic kernel. This kind of kernel allows for the scale of autocorrelation to vary independently in each dimension in contrast with isotropic kernels, where the scale of autocorrelation is restricted to be the same in all dimensions. Isotropic kernels are appropriate when the dimensions are literal spatial dimensions (e.g., latitudinal and longitudinal distance).

For this risk analysis, we evaluated a variety of kernel functions across multiple forecast years and operational scenarios, and found the variation in model fit across different anisotropic kernels to be generally immaterial. We ultimately developed GPR meta-models for all forecast years and operating scenarios using an anisotropic radial basis function.[49]

H.3 Cross Validation

One feature of using GPR with deterministic main models is that there are no "residuals" for sampled observations used in model estimation: the expected value always passes exactly through every observation point. An important consequence of this feature is that GPR models cannot be evaluated based on traditional "goodness of fit" measures (e.g., R^2) derived from the estimation data. R^2 measures the ratio between explained variation and total variation in outcome measures, but the GPR expectation values always "explains" all of the variation in outcomes; since GPR models have no residuals, this measure is meaningless for such models. Instead of measuring goodness of fit directly based on estimation data, it is necessary to measure fit on a validation data set that is not used for model estimation. Because it is usually expensive to collect additional validation data, it is preferred to conduct K-fold cross validation (CV). For this, the set of observations is randomly partitioned into K subsets (typically 5 to 10; for this risk analysis we have used K=10). The entire GPR model is re-estimated using the same kernel and K-1 subsets of the data (leaving one out). Then a model score is calculated by using the result to predict the outcomes on the remaining held-out subset of observations. The entire process is repeated iteratively holding out each of the K subsets one at a time, and then averaging the resulting scores. The CV score is interpreted in roughly the same manner as R^2 for LR models, such that a score of 1.0 indicates a perfect prediction, and a score of 0.0 is achieved by predicting the global mean of the dependent variable. When the GPR is applied to de-trended data (i.e., on top of a LR model), the resulting CV scores are calculated based on the residuals from the LR model, so they represent the relative improvement in fit over the LR result, and are not directly comparable to the R^2 values.

[49] For mathematical details of the implementation refer to the scikit-learn documentation: http://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.RBF.html.