2024 Business Plan
February 2024

Ridership and Revenue Risk Analysis Report for the 2024 Business Plan

Prepared by:

Steer
800 Wilshire Blvd, Suite 1320,
Los Angeles, CA 90017
United States

+1 (213) 425 0990
www.steergroup.com

Prepared for:

DB E.C.O. North America, Inc.
555 Capitol Mall Suite #1250,
Sacramento,
CA 95814,
United States
https://db-eco.us/en/
Client ref:
Our ref: 23454511

Signature/Approval Sheet

TO: Bill Casey

FROM: Justin Cassarino

SUBJECT: Approval of 2024 Ridership and Revenue Risk Analysis Report

DESCRIPTION OF ENCLOSED DOCUMENT(S): 2024 Ridership and Revenue Risk Analysis Report

REVIEWER	REVIEWER'S INITIALS/DATE:	COMMENTS
Signer #1 Name (Print): Bill Casey CHSRA Chief Operating Officer	SIGNATURE ON FILE	
Signer #2 Name (Print): Brian Annis CHSRA Chief Financial Officer	SIGNATURE ON FILE	
Signer #3 Name (Print): Thomas Meinl ETO Project Director	SIGNATURE ON FILE	
Reviewer #1 Name (Print): Bruce Armistead CHSRA Chief of Rail Operations	SIGNATURE ON FILE	
Reviewer #2 Name (Print): Christian Hicke DB E.C.O. NA V.P. Management Consulting	SIGNATURE ON FILE	
Reviewer #3 Name (Print): Justin Cassarino ETO Consultant	SIGNATURE ON FILE	
Author #1 Name (Print): Alex Mitrani Steer Associate	SIGNATURE ON FILE	
Author #2 Name (Print): Masroor Hasan Steer Director	SIGNATURE ON FILE	

Contents

Figures

Tables

1 Introduction

1.1 The main objective of the risk analysis is to generate a range of probable ridership and revenue forecasts that are the results of variations in specific variable assumptions of the California Rail Ridership Model (CRMM).

1.2 The methodology to perform this analysis includes:

1. Identify a specific set of factors from the full set of inputs that are expected to have significant impact on ridership and revenue.
2. Specify a range of variation for each factor based on reviewing previous studies as well as updated research and information, where pertinent.
3. Based on steps 1 and 2, prepare—as best as possible—a representative sample where each observation is the outcome of a complete model-run for a given vector of input values.
4. Based on the size and structure of the sample data prepared, apply a statistically robust technique or model to estimate the relationship between ridership or revenue with the set of key risk variables identified in step 1. This is the meta-model[1].
5. Evaluate model diagnostics and cross-validate to determine the best model specification.
6. Once a preferred model is identified, use it as part of a Monte Carlo simulation to generate a distribution of ridership and revenue estimates by varying the inputs over continuous ranges.
7. Perform either point-specific or range-specific sensitivity tests based on the distribution generated.

1.3 The proposed methodology is broadly comparable with that used in earlier work by Steer (Steer, 2022a, p. 199) and by Cambridge Systematics (Cambridge Systematics, Inc., 2020).

1.4 The set of variables to be included in the risk analysis has been revised, as have the definitions of the distributions to be used. Many of the key risk factors used in the previous studies mentioned above have been retained here, such as auto operating costs, airfares, HSR fares, and HSR service frequency. A few changes have been proposed based on the results of the pervious sensitivity analysis, a review of the modeling methodology, and the relevant literature. These modifications are outlined in the following chapters.

[1] Note that we are constrained in the order of magnitude of feasible model runs, such that the possible models that can be applied are limited.

1.5 The scope of this risk analysis covers six project scenarios:

- 2030: Valley to Valley (V2V) – EMME supply scenario v2v
- 2040: Valley to Valley (V2V) – EMME supply scenario v2v
- 2050: Valley to Valley (V2V) – EMME supply scenario v2v
- 2030: Phase 1 (PH1) – EMME supply scenario ph1
- 2040: Phase 1 (PH1) – EMME supply scenario ph1
- 2050: Phase 1 (PH1) – EMME supply scenario ph1

1.6 Results for the same EMME supply scenario in different future years were obtained using growth factors – data was only collected with model runs for 2040 for each supply scenario:

- 2040: Valley to Valley (V2V) – EMME supply scenario v2v
- 2040: Phase 1 (PH1) – EMME supply scenario ph1

1.7 The data for 2030 and 2050 were estimated based on the data collected for 2040, with growth factors applied to represent the expected changes in trip volumes. The 2030 and 2050 base case model runs were compared to corresponding 2040 base case model runs to obtain the growth factors.

1.8 The revenue forecasts in this report are displayed in Year of Expenditure (YOE) dollars to reflect monetary values during future years. The methodology to convert revenue from Base Year to YOE dollars consists of two components:

1. Converting values from Base Year 2018$ to June 2023$ using the observed California Consumer Price Index.
2. Converting values from June 2023$ to YOE$ using the California Consumer Price Index (California Department of Finance) and the United States Federal Reserve Inflation Target.

2 Review of Previous Work

Introduction

2.1 This chapter includes a brief literature review, a review of the previous Steer sensitivity analysis for Phase 1 in 2030 (Steer, 2022b), and a review of relevant aspects of the earlier sensitivity analysis undertaken by Cambridge Systematics (Cambridge Systematics, Inc., 2020).

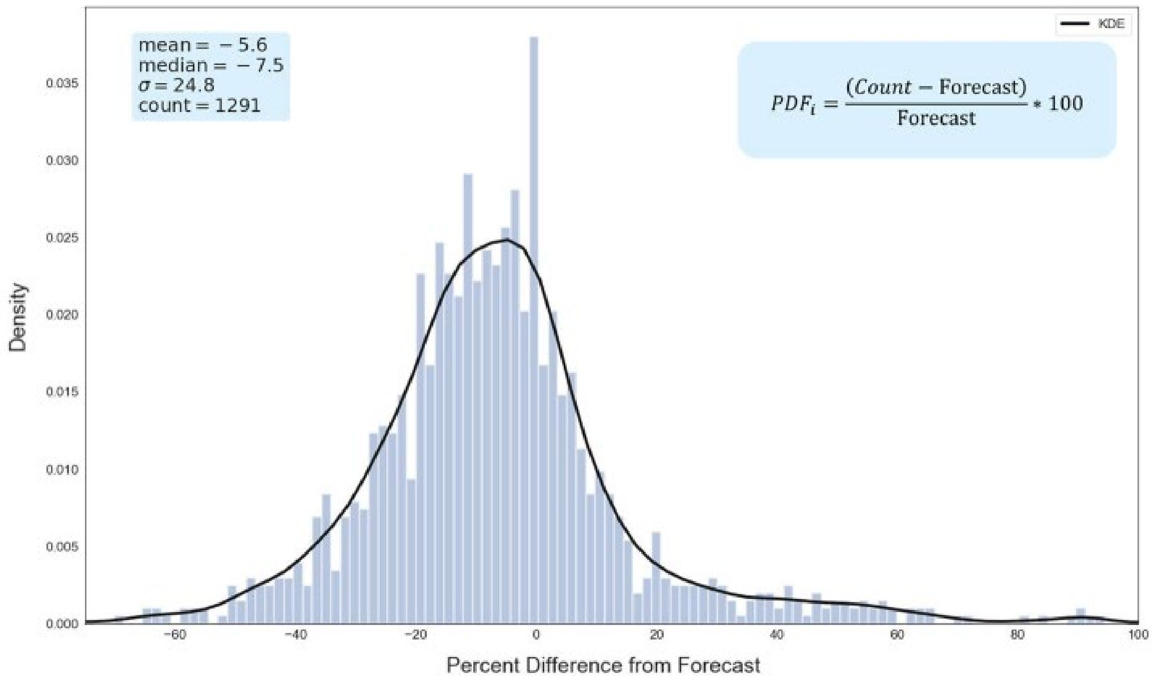
Literature Review

2.2 The literature review focused on publications with topics related to High-Speed Rail, risk analysis or ex-post evaluation. The objective was to broaden the scope beyond the present project and consider relevant evidence from other projects and other places.

Ex-Post Evaluations of Demand Forecast Accuracy

2.3 There are a few recent reviews of forecast accuracy that are based exclusively on road projects but contain so many projects that they are worth considering here. In "The changing accuracy of traffic forecasts" (Hoque et al., 2022) the authors review the accuracy of demand forecasts for 1291 road projects from the USA and Europe. They found that on average the forecasts underestimated the demand by 6%, but they would have over-estimated the demand by 1% on average had it not been for the impact of the post-2008 recession on employment. In "Estimating the uncertainty of traffic forecasts from their historical accuracy" (Hoque et al., 2021), the authors use the same data to demonstrate how to construct an "uncertainty window" for the expected range of observed traffic in relation to the forecasts, using quantile regression based on the available data. Figure 2.1 below shows the distribution of forecast inaccuracy in the data used for these articles.

Figure 2.1: Distribution of percent difference between observed and forecast demand



Source: "The changing accuracy of traffic forecasts" (Hoque et al., 2022, p. 8)

2.4 The article "Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review" (Nicolaisen & Driscoll, 2014) is a review of previous studies of demand forecast accuracy, which covers 12 previous reviews, 6 of which included rail projects. Figure 2.2 below summarizes the data on forecast inaccuracy in terms of the means and standard deviations. A negative number for the mean indicates that observed demand tended to be lower than the forecast demand, as seems to have been the case for rail in all these studies.

Figure 2.2: Comparison of means and standard deviations for observed demand forecast inaccuracy

Author(s)	Sample[a]	Mean	Standard deviation
Mackinder and Evans (1981)	Road: 44	−7%[b]	N/A
NAO (1988)	Road: 128	+8%	43
Pickrell (1990)	Rail: 9	−65%	17
Flyvbjerg et al. (2006)	Road: 183	+10%	44
	Rail: 27	−40%	52
DoT (2007)	Rail: 19	−37%	31
DoT (2008)	Rail: 18	−16%	59
Bain (2009)	Toll: 104	−23%	26
Button et al. (2010)	Rail: 44[c]	−21%	58
Parthasarathi and Levinson (2010)	Road: 108	+6%	41
HA (2011)	Road: 62	+3	21
Welde and Odeck (2011)	Toll: 25	−3%	22
	Road: 25	+19%	21
Nicolaisen (2012)	Road: 146	+11%	35
	Rail: 31	−18%	33

Source: “Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review” (Nicolaisen & Driscoll, 2014, p. 7)

2.5 Regarding the likely sources of forecast errors, the authors write:

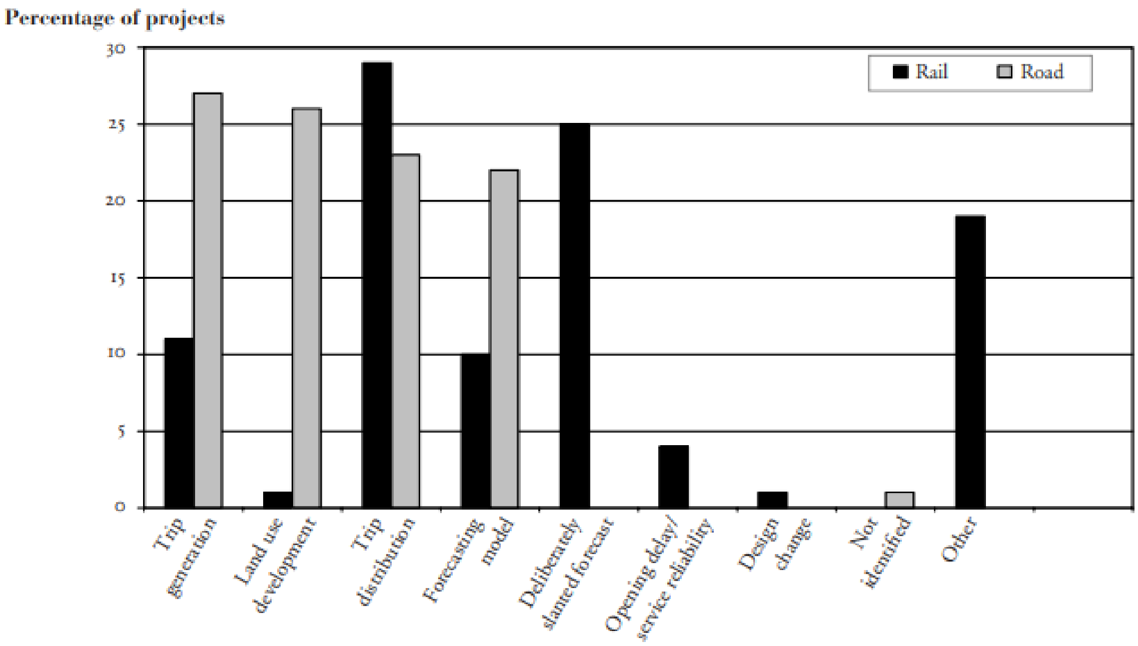
> The most important source of inaccuracy for demand forecasts appears to be auxiliary forecasts of exogenous variables, where inaccurate forecasts of economic growth, car ownership and migration patterns propagate into demand forecasts.

Source: “Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review” (Nicolaisen & Driscoll, 2014)

2.6 The article “How (in) accurate are demand forecasts in public works projects?: The case of transportation” (Flyvbjerg et al., 2005) includes data on forecast and observed demand for 27 rail projects. The average forecast inaccuracy for the rail projects was -51.4% indicating that on average the observed demand was about half of the forecast demand.

2.7 The authors try to identify the causes of demand forecast errors. Figure 2.3 below shows the distributions of stated reasons of forecast inaccuracy for road and rail projects. These “stated reasons” were based on a review of the explanations of the project managers and researchers regarding the inaccuracy of the forecast in each case.

Figure 2.3: Stated causes of inaccuracies in traffic forecasts (N=26 rail projects and 208 road projects)



Source: "How (in) accurate are demand forecasts in public works projects?: The case of transportation" (Flyvbjerg et al., 2005, p. 10)

Ex-Post Evaluations of Population Forecast Accuracy

2.8 The article "An Evaluation of Population Projections by Age" (Smith & Tayman, 2003) includes estimates of the Mean Absolute Percentage Error of two sets of state-level population forecasts with a 20-year horizon, with target years of 1990 and 2000. The Mean Absolute Percentage Errors were as shown in Table 2.1 below.

Table 2.1: Mean Absolute Percentage Errors (MAPEs) of 20-year state-level population forecasts

Base year	Target year	Mean Absolute Percentage Error (MAPE)
1970	1990	6.0
1980	2000	11.2

Source: (Smith & Tayman, 2003, p. 7)

2.9 Emerging results from the 2020 census also suggest that the error in long-term population forecasts can be substantial.

Table 2.2: Error in 20-year forecast of California population based on preliminary analysis of the 2020 census

Variable	Value	Source
Resident population according to the 2020 census	39,538,223	"A Preliminary Analysis of U.S. and State-Level Results From the 2020 Census" (Hartley et al., 2021)
Resident population according to forecasts based on 2000 census	42,206,743	"State Population Projections 2004-2030 Results" (U.S. Census Bureau, 2005)
Percentage Error	-6.3	

Source: Steer based on sources noted in table.

Ex-Post Evaluations of Induced Travel Related to HSR Projects

2.10 "A Review of Ex-Post Evidence for Mode Substitution and Induced Demand Following the Introduction of High-Speed Rail" (Givoni & Dobruszkes, 2013) reviews the available evidence for over 14 HSR projects in several countries. The authors found that on average induced travel was about 20% of total demand, with individual estimates ranging from 6% to 45%. This range is equivalent to demand factors varying between 1.06 and 1.82 with an average of 1.25. The base assumption of 1.08 of the CRRM is close to the lower bound of this range.

Review of Phase 1 in 2030 Preliminary Sensitivity Analysis

2.11 This work involved the following sequential steps (Steer, 2022a, p. 199):

- Selection of sensitivity factors
- Sampling distributions
- Latin hypercube sampling
  - Variables that were inputs to the full model runs only
- Data collection
  - Full model runs
  - Off-model tests
- Meta-modeling
  - Ridership
  - Revenue
- Monte Carlo Simulation
  - Ridership
  - Revenue

Selection of Sensitivity Factors and Their Sampling Distributions

2.12 There were two sets of risk variables that were dealt with differently:

- Variables that fed into the full CRRM model runs
  - The Latin Hypercube sampling was limited to these variables

- Variables that were tested off-model using the CRRM model outputs including induced travel

The results of testing these two sets of variables were combined and used as inputs to the meta-modeling.

2.13 The sensitivity analysis included the risk variables detailed in Table 2.3 below.

Table 2.3: Risk variables and the distributions used for CRRM tests

Index	Variable	Codename	Base	Min.	Mode	Max.	Distribution	Shape	CRRM
1	High-speed rail constant difference factor	hsrcodifa	0.5	0	0.5	1	PERT	4	1
2	Business / commute trip generation factor	bctrgefa	1	0.76	1	1.31	PERT	4	1
3	Recreation / other trip generation factor	rotrgefa	1	0.88	1	1.14	PERT	4	1
4	Auto operating cost factor	autopcofa	1	0.79	1	1.33	PERT	5	1
5	High-speed rail fare factor	hsrfafa	1	0.74	1	1.42	Triangular		1
6	High-speed rail frequency factor	hsrfqfa	1	0.45	1	1.55	Triangular		1
7	Air fare factor	airfafa	1	0.7	1	1.3	Triangular		1
8	Air service frequency	airfrfa	1	0.4	1	1	Triangular		1
9	Access / egress time parameter factor	acegtipafa	1	0.5	1	1.5	PERT	4	1
10	Population and households forecast factor	pohofowefa	1	0.85	1	1.15	PERT	4	1

Index	Variable	Codename	Base	Min.	Mode	Max.	Distribution	Shape	CRRM
11	Employment forecast factor	emfowefa	1	0.85	1	1.15	PERT	4	1
12	Auto in-vehicle time parameter factor	autivtpafa	1	0.6	1	1.4	Triangular		1
13	Employed cost parameter factor	emcopafa	2	1.5	2	3	Triangular		1
14	Long access / egress trips	exloaceg	1	0	-	2	Uniform		0
15	Non-resident trips	visitrav	1	0	-	2	Uniform		0
16	Induced trips	Indutrav	1	0.5	-	1.5	Uniform		0

Notes: “CRRM” is a dummy variable indicating whether each variable was tested as an input to the CRRM full model runs (1 for yes, 0 for no). Source: “Wider sensitivity analysis” (Steer, 2022a, p. 199)

2.14 Most of the distributions of the risk variables were defined using a modified PERT distribution (Vose, 2008, p. 407). A PERT distribution (Special Projects Office, 1958, p. 11) is a type of probability distribution widely used in risk analysis when dealing with few—in particular, three—subjective estimates that a random variable can take. The PERT distribution, or family of distributions, is a variation of the well-known Beta distribution in statistics, and where the mean of the distribution is calculated based on the minimum, maximum and most likely values that the random variable can take. These three estimates are subjective but provide enough information to develop a probability distribution of the random variable, and thus make probability statements.

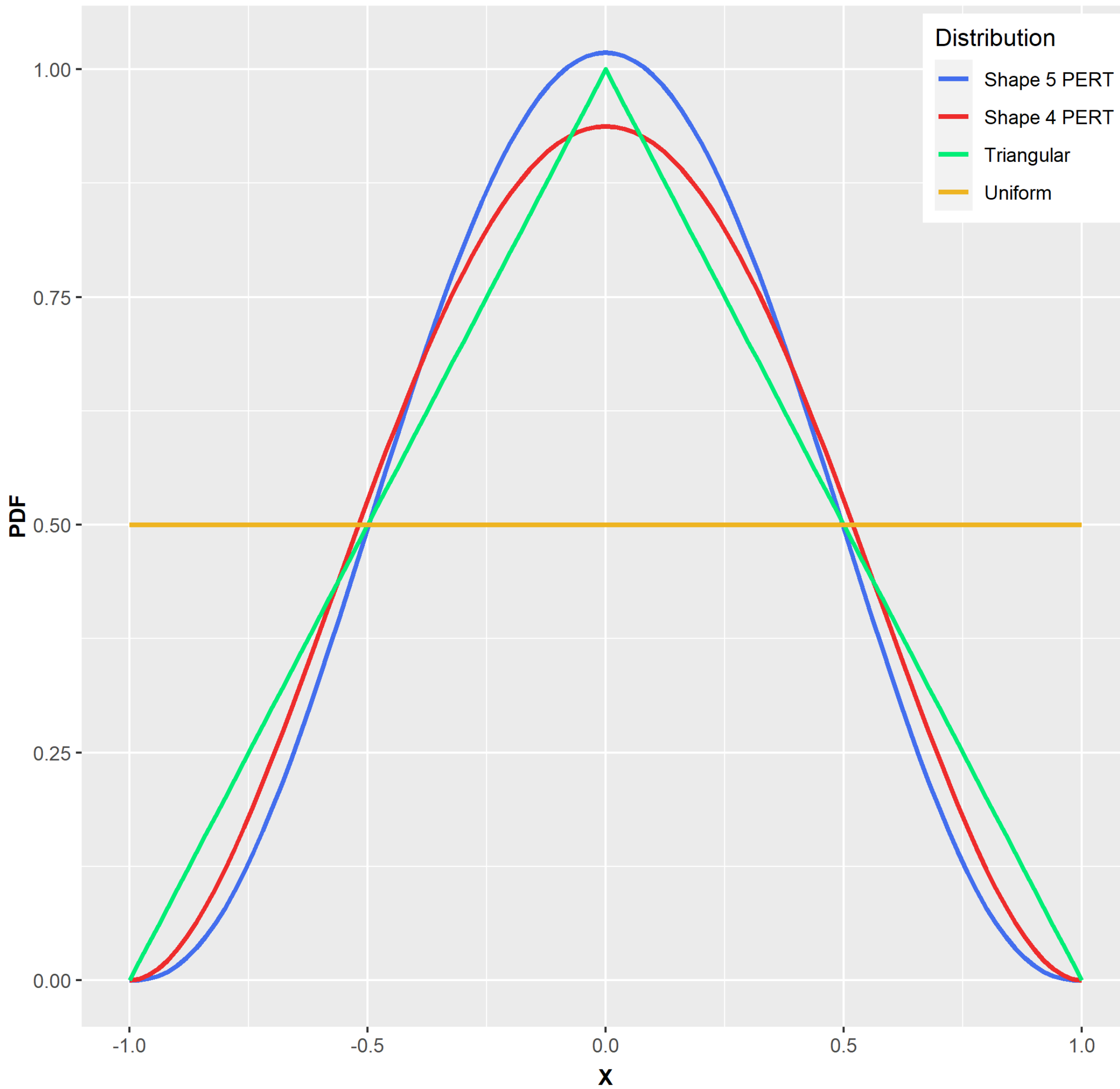
2.15 The mean of the modified PERT distribution is defined as:

$$\mu = \frac{a+kb+c}{k+2} \tag{1}$$

Where a, b, and c represent the minimum, most likely, and maximum values of the user-defined random variable, respectively, and k is the weight assigned to the most likely value. k also defines the shape of the PERT distribution, with larger values of k making the distribution more deterministic, and smaller values increasing its spread. Therefore, k reflects the certainty in the most likely option. The PERT distributions are smooth functions, and thus consistent with what a natural distribution of values could be expected to look like. The triangular distribution, which uses the same subjective estimates to construct a distribution lacks this feature, making it less reliable when the extreme values used to define it are poorly estimated or defined.

2.16 Figure 2.4 presents the types of risk variable distributions considered in the CRRM sensitivity analysis. The curves are the probability density functions of a random variable X that is assumed to vary over a range [-1, +1], with mean (for all except the uniform distribution) at zero.

Figure 2.4: Probability distributions used to define risk variable uncertainty



Source: Steer

2.17 The distributions used for the risk variables as inputs to the model tests do not necessarily have to be the same as the distributions used with the meta models as part of the Monte-Carlo simulation, but they should cover the same range or a broader range in each case.

Latin Hypercube Sampling and Data Collection

2.18 Latin hypercube sampling was used to generate the inputs needed for the CRRM model runs. Considering 10 model runs for each of the 13 input variables this resulted in 130 full model runs.

These tests were undertaken using the version of the base case that did not include induced travel.

2.19 Off-model tests were used for 3 variables - Exceptionally Long Access and Egress, Visitor Travel, and Induced Travel. These were done using the version of the base case that included induced travel. In each case, the variable to be tested was varied from the minimum value to the maximum value at fixed increments (10 tests) while keeping all other variables at their base values.

2.20 To combine the ridership and revenue results of the off-model tests with the ridership and revenue results of the full model runs, scaling factors were applied to the results of the off-model tests to correct for the difference in scale of each variable between the base case without induced travel and the base case with induced travel. These scaling factors (SF) were defined as follows:

$$SF_{var} = \frac{var_{base\ case\ no\ induced\ travel}}{var_{base\ case\ with\ induced\ travel}} \qquad (2)$$

Where: SF = Scaling Factor, var = variable.

2.21 This resulted in the following values for the scaling factors:

Table 2.4: Derivation of scaling factors to correct for differences between the base cases used for model runs and off-model tests

Variable	Comment	Annual Ridership (millions)	Annual Revenue (millions)
Run 259	With induced travel, used for off-model tests	25.53	1,737.34
Run 260	Without induced travel, used for full model tests	23.60	1,544.04
Scaling Factor		0.924	0.889

Source: Steer

Meta-Modeling

2.22 The meta modeling methodology was broadly comparable to that applied in the earlier work done by Cambridge Systematics (Cambridge Systematics, Inc., 2020, p. 40).

2.23 Ridership and revenue were modeled separately, resulting in two metamodels:

- A ridership model; and
- A revenue model.

2.24 Each model used two complementary techniques:

- An ordinary least squares (OLS) regression model that was fitted to the data points generated by the full model runs; and
- A Gaussian process model (GPM) that was fitted to the residuals of the OLS model (Cambridge Systematics, Inc., 2018, p. 105).

2.25 The models were estimated using all the data combined:

- Base model run (1 observation)
- Full model run tests (130 observations)
- Off-model tests (30 observations).

2.26 The data set used in model estimation consisted of 161 observations for each model (ridership and revenue).

Monte Carlo Simulation

2.27 The Monte Carlo simulation produced substantially narrower distributions of ridership and revenue than the earlier work by Cambridge Systematics. The two sets of results are compared in Table 2.5 below.

Table 2.5: Comparison of results of Cambridge Systematics and Steer Monte Carlo simulation, Phase 1

Percentile or reference point	CS ridership index (WRT base value)	CS revenue index (WRT base value)	Steer ridership index (WRT base value)	Steer revenue index (WRT base value)
base run	1.00	1.00	1.00	1.00
p0	0.24	0.30	0.44	0.41
p1	0.41	0.47	0.63	0.61
p10	0.59	0.67	0.77	0.76
p25	0.76	0.85	0.86	0.86
p50	1.00	1.10	0.98	1.00
p75	1.29	1.40	1.13	1.15
p90	1.60	1.70	1.27	1.31
p99	2.13	2.23	1.54	1.60
p100	3.08	3.12	2.24	2.28

Notes: The Cambridge Systematics results are for 2033 and the Steer results are for 2030. Sources: "California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis" (Cambridge Systematics, Inc., 2020, pp. 50, 53), "CA Ridership Modeling" (Steer, 2022a, p. 199).

2.28 There could be several potential reasons for the differences in the two sets of results:

- Differences in the underlying demand model, for instance different choice model parameters
- Differences in the set of variables included in the meta models. The sets of variables are similar but not the same.
- Differences in the set of variables used in the Monte Carlo Simulation.
- Differences in the ranges and distributions used in the Monte Carlo simulation. For example,
  - The Steer work used triangular distributions for "Long access / egress trips" and "Non-resident trips" whereas the CS work used uniform distributions.
  - The CS work used a broader range of values (factors from 0 to 2) for induced travel than the range considered in the Steer analysis (factors from 0.9 to 1.1).
- Differences in the correlation matrix used in the Monte Carlo simulation.

- In the Cambridge Systematics methodology, visitor travel and induced travel were represented as external risk factors. The corresponding components of trips and revenue were added on to the results of the meta models. In contrast, in the Steer approach these variables were included in the meta models.

3 Selection of Sensitivity Factors

Review of Sensitivity Factors Tested Previously

3.1 The sensitivity of ridership and revenue to each factor was evaluated in two ways:

- OLS model results
- Correlation coefficients

OLS Model Results

3.2 The OLS model parameters for Phase 1 in 2030 and the associated t-statistics are shown in Table 3.1 below. These results are useful to check whether each independent variable had a significant effect on the dependent variable or not.

Table 3.1: Summary of OLS model parameters for ridership and revenue models

Variable	Ridership parameter	Ridership t value	Revenue parameter	Revenue t value
Constant	8.968	238.49	12.022	182.97
Access / egress time parameter factor	-0.506	-101.63	-0.496	-56.99
High-speed rail constant difference factor	0.468	94.14	0.616	71.01
High-speed rail fare factor	-0.598	-89.28	0.184	15.73
Auto in-vehicle time parameter factor	0.486	84.52	0.699	69.58
Population and households forecast factor	0.849	50.3	0.811	27.51
Recreation / other trip generation factor	0.766	39.67	0.723	21.42
High-speed rail frequency factor	0.107	25.49	0.134	18.25
Long access / egress trips	0.118	23.46	0.117	13.35
Auto operating cost factor	0.181	18.06	0.257	14.71
Employment forecast factor	0.270	16.29	0.230	7.92
Air fare factor	0.114	14.95	0.163	12.18
Business / commute trip generation factor	0.130	14.33	0.103	6.49
Air service frequency	-0.051	-8.95	-0.011	-1.13
Non-resident trips	0.044	8.74	0.042	4.81
Induced trips	0.076	7.52	0.111	6.33
Employed cost parameter factor	-0.015	-5.03	-0.020	-3.88

Notes: N=161. Source: Steer

3.3 All the independent variables had statistically significant effects on ridership, revenue, or both. The effect of “Air service frequency” on revenue was insignificant at the 95% level (t=-1.13) although this same variable had a significant effect on ridership (t=-8.95).

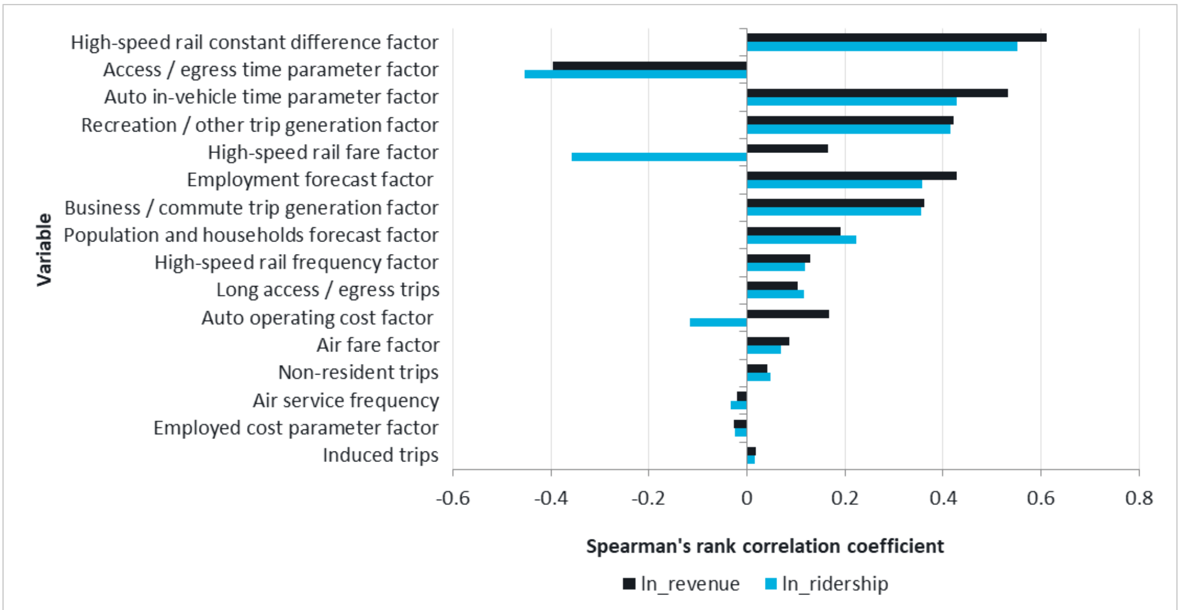
3.4 The OLS model parameters are not very useful for understanding the relative impacts of each variable. For this purpose, correlation coefficients are more useful as they have standardized ranges on the interval [-1, 1] which facilitates their comparison.

Correlation Coefficients

3.5 An analysis of the correlations between the dependent and independent variables was undertaken to evaluate the relative sensitivity of the dependent variables to each independent variable.

3.6 Figure 3.1 and Table 3.2 below show the correlation coefficients obtained with the Monte Carlo simulation data.

Figure 3.1: Correlations between independent and dependent variables in Monte-Carlo simulation data



Notes: N=100,000. Source: Steer

Table 3.2: Correlations between independent and dependent variables in Monte Carlo simulation data

Variable	Correlation with ln_ridership	Correlation with ln_revenue
High-speed rail constant difference factor	0.5521	0.6109
Access / egress time parameter factor	-0.4542	-0.396
Auto in-vehicle time parameter factor	0.4282	0.5332

Variable	Correlation with ln_ridership	Correlation with ln_revenue
Recreation / other trip generation factor	0.4143	0.4219
High-speed rail fare factor	-0.3578	0.1652
Employment forecast factor	0.3571	0.4276
Business / commute trip generation factor	0.3546	0.3611
Population and households forecast factor	0.2239	0.1912
High-speed rail frequency factor	0.1175	0.1289
Long access / egress trips	0.1173	0.1031
Auto operating cost factor	-0.1154	0.167
Air fare factor	0.0688	0.0864
Non-resident trips	0.0487	0.0409
Air service frequency	-0.034	-0.0195
Employed cost parameter factor	-0.0246	-0.0256
Induced trips	0.0152	0.0189

Notes: Spearman's rank correlation coefficient is used. N=100,000. Source: Steer

3.7 The impact of the "Induced trips" variable is probably understated in the Monte Carlo simulation data because this variable was modeled with limited variability of only +/- 10%.

3.8 Considering all the above evidence, most of the variables included in the sensitivity analysis had a significant effect and seemed worth retaining.

3.9 There were only a few variables that were candidates for being dropped or being reformulated:

- The "Employed cost parameter factor" had a low impact compared to the other variables. The purpose of this variable was to test the sensitivity to an assumption affecting the value of time, but it impacted only a fraction of the trips. This variable could be omitted.
- The "Air service frequency" variable had a modest but significant effect on HSR ridership. The effect it represents seems likely to occur and therefore it seemed preferable to retain this variable in the model. It was considered beneficial to use a uniform distribution instead of a triangular distribution to increase the spread of the sampled values. A triangular distribution with mode equal to the base assumption would tend to result in a significant density of sampled values close to the base value.

Other Variables

3.10 The access/egress nest parameter is used to prevent the presence of several access/egress combinations from unduly increasing the demand for the public transport modes (rail, flight, bus and HSR). A nest parameter of 1 would imply that each combination of access and egress modes is perceived as a completely different mode from the others, which is not going to be the case. Therefore, this parameter must necessarily be lower than 1 and it is expected to be 0.5 or lower.

In the base case, the value of the nest parameter is assumed to be 0.5. The industry practice is to treat 0.5 as the most likely value and to also test values between 0.5 and 0.1.

3.11 The potential variability in the assumption of the transfer penalty values prompted us to include it in the sensitivity analysis, to determine its impact on the forecasts.

Selection of Variables

3.12 Table 3.3 shows the variables that were used in the sensitivity analysis for the two 2040 supply scenarios:

- Valley to Valley (V2V)
- Phase 1 (PH1)

Table 3.3: Variables included in the sensitivity analysis

Variable	2040 V2V	2040 PH1	Off-Model Tests
High-speed rail constant difference factor	1	1	0
Business / commute trip generation factor	1	1	0
Recreation / other trip generation factor	1	1	0
Auto operating cost factor	1	1	0
High-speed rail fare factor	1	1	0
High-speed rail frequency factor	1	1	0
Air fare factor	1	1	0
Air service frequency	1	1	0
Access / egress time parameter factor	1	1	0
Population and households forecast factor	1	1	0
Employment forecast factor	1	1	0
Auto in-vehicle time parameter factor	1	1	0
Nest parameter factor	1	1	0
Transfer penalty	1	1	0
Long access / egress trips	1	1	1
Non-resident trips	1	1	1
Induced trips	1	1	1
Total	**17**	**17**	**3**

Source: Steer (2023)

3.13 The changes compared to the previous sensitivity analysis were as follows:

- To be added:
  - Nest parameter factor
  - Transfer penalty
- To be adjusted:
  - Use uniform distribution instead of triangular distribution for “Air service frequency”

- To be dropped:
  - Employed cost parameter factor

4 Sampling Distributions

Introduction

4.1 A few guiding principles were used to inform the recommendations regarding sampling distributions:

- Ranges should be broad
- Distributions should be uncorrelated

In the Monte Carlo simulation, the ranges can be narrowed but they should not be made broader than the ranges that were sampled in the data collection. Similarly, correlations can be introduced in the Monte Carlo simulation; but to estimate good meta models the distributions used in the data collection should be uncorrelated.

4.2 The recommended sampling distributions are summarized in Table 4.1 below. They have been normalized in relation to the central assumption used in the base case model run so that the base values of the factors are all 1.

Table 4.1: Risk variables and distributions

Index	Variable	Codename	Base	Minimum	Mode	Maximum	Distribution	Shape	CRRM
1	Population and households forecast factor	pohofofa	1	0.85	1	1.15	PERT	4	1
2	Employment forecast factor	emfofa	1	0.85	1	1.15	PERT	4	1
3	Business / commute trip generation factor	bctrgefa	1	0.75	1	1.25	PERT	4	1
4	Recreation / other trip generation factor	rotrgefa	1	0.75	1	1.25	PERT	4	1

Index	Variable	Codename	Base	Minimum	Mode	Maximum	Distribution	Shape	CRRM
5	Auto operating cost factor	autopcofa	1	0.70	1	1.30	PERT	5	1
6	High-speed rail fare factor	hsrfafa	1	0.60	1	1.40	Triangular		1
7	High-speed rail frequency factor	hsrfqfa	1	0.45	1	1.55	Triangular		1
8	Air fare factor	airfafa	1	0.7	1	1.3	Triangular		1
9	Air service frequency factor	airfrfa	1	0.4		1	Uniform		1
10	High-speed rail constant difference factor	hsrcodifa	1	0	1	2	PERT	4	1
11	Access / egress time parameter factor	acegtipafa	1	0.5	1	1.5	PERT	4	1
12	Auto in-vehicle time parameter factor	ivtpafa	1	0.6	1	1.4	Triangular		1
13	Nest parameter factor	nestpafa	1	0.2		1	Uniform		1
14	Transfer penalty factor	transpenfa	1	0.3		2	Uniform		1
15	Induced trips factor	indutripsfa	1	0		2	Uniform		0
16	Long access / egress trips factor	exloacegfa	1	0		2	Uniform		0

Index	Variable	Codename	Base	Minimum	Mode	Maximum	Distribution	Shape	CRRM
17	Non-resident trips factor	visitravfa	1	0		2	Uniform		0

Notes: "CRRM" is a dummy variable indicating whether each variable should be tested as an input to the CRRM full model runs (1 for yes, 0 for no). Source: Steer.

4.3 The base case used for the data collection with the CRRM model should correspond to a model run not including induced travel, to make it possible for all the relevant factors to be tested in relation to a consistent reference case.

4.4 Each of the factors to be tested will now be commented on in further detail. The sampling distributions have been ordered in terms of the following thematic groups:

- Trip generation
- Levels of service
- Choice model parameters
- Induced demand and related factors

Trip Generation

Population and Household Forecast Factor

4.5 The risk associated with population and household forecasts will be modeled by applying a factor to the forecasts. This is a departure from the Cambridge Systematics methodology which blended low, central and high forecasts (Cambridge Systematics, Inc., 2020, p. 31). Table 4.2 summarizes the 2040 forecasts used by Cambridge Systematics at state level and gives an indication of the variability of the forecasts data.

Table 4.2: Summary of state-level population and households forecasts for 2040

Variable	Minimum	Mode	Maximum
Population	42,802,960	45,865,590	48,106,929
Population index	0.93	1.00	1.05
Households	15,183,596	15,842,187	16,684,409
Household index	0.96	1.00	1.05
Average people per household	2.82	2.90	2.88

Source: Steer using data from "California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis" (Cambridge Systematics, Inc., 2020, p. 31)

4.6 While the above data suggests a modest degree of variability in the forecasts, the variability in forecasts is not necessarily indicative of how much error there might be in long-term population forecasts when these are compared with census data. Considering the substantial error in long-term population forecasts discussed earlier (see the section "Ex-Post Evaluations of Population Forecast Accuracy", in Chapter 2), the ranges used in the Cambridge Systematics work may be too narrow. The naïve average of the two MAPE values is 8.6, and the observed error based on the most recent census results indicates a forecasting error of 6.3 percent. A PERT shape 4 distribution with minimum 0.85, mode 1, and maximum 1.15 seemed appropriate.

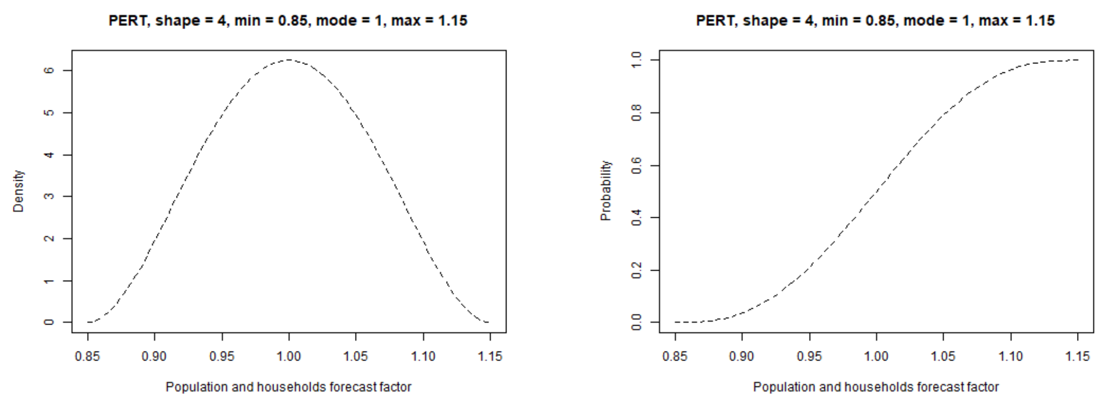
Example of Implementation

$$P' = P*x \tag{3}$$

$$H' = H*x$$

Where P is population, H is households, and x is a simulated random number with the distribution illustrated in Figure 4.1.

Figure 4.1: Sampling distribution of population and households forecast weighting factor



Source: Steer

Employment Forecast Factor

4.7 Variation in employment forecasts will be modeled independently from variation in population and households. Table 4.3 summarizes the 2040 forecasts used by Cambridge Systematics at state level and gives an indication of the variability of the forecasts data.

Table 4.3: Summary of state-level employment forecasts for 2040

Variable	Minimum	Mode	Maximum
Employment	20,582,689	21,375,786	22,099,964
Employment index	0.96	1.00	1.03

Source: Steer using data from “California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis” (Cambridge Systematics, Inc., 2020, p. 31)

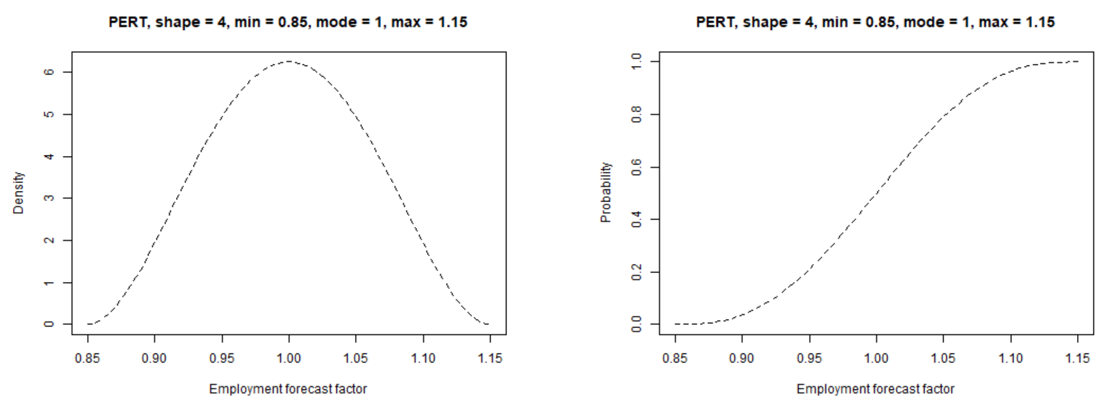
4.8 Evidence on the accuracy of employment forecasts when compared with census data (U.S. Bureau of Labor Statistics, 2020) indicates a Mean Absolute Percentage error of 6.4% at the national level for forecasts with a 10-year horizon. State-level forecasts would be expected to have more errors than national forecasts. Longer-term forecasts would be expected to have more errors than shorter-term forecasts. Therefore, the seven percentage-point range implied by the table above seems inadequate. As with the population and household forecasts, a PERT shape 4 distribution with minimum 0.85, mode 1, and maximum 1.15 seemed appropriate.

Example of Implementation

$$E' = E*x \tag{4}$$

Where E is employment and x is a simulated random number with the distribution illustrated in Figure 4.2.

Figure 4.2: Sampling distribution of employment forecast factor



Source: Steer

Trip Generation Rate Factors

4.9 The CRRM model base case assumption is that trip rates for all purposes remain constant for future years (Steer, 2022a, p. 14). This is unlikely to occur in practice therefore it is appropriate to vary these trip generation factors as part of the sensitivity analysis.

Cambridge Systematics Assumptions

4.10 As the sensitivity analysis included employment as a separate variable, the trips generation factors were limited to the second component – unexplained variation. The bounds were based on maintaining specific trip rates:

- Minimum: The lower bound is constructed by setting values at each model year to produce the same number for each trip purpose as in the model base year. That is, even as population increases, the trips generated within each purpose remain unchanged.
- Base Case: The calculated factors.
- Maximum: The upper bound is defined by simply mirroring the constant.

4.11 Values were created for each model year that would maintain the same total trip rate, and a corresponding factor change away from the base was used to define the upper bound. The year 2010 as the reference year, and 2040 as the future year (Cambridge Systematics, Inc., 2020, p. 70).

Table 4.4: Unexplained variation of trip frequency constants—implied annual long-distance round trips per capita

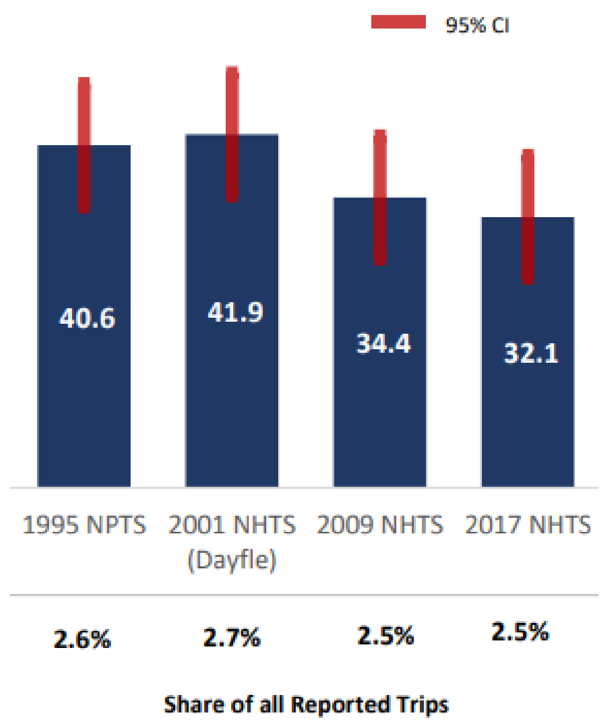
Variable	Units	Minimum	Base Case	Maximum
Business / Commute	Annual trips per capita	1.87	2.46	3.23
Recreation / Other	Annual trips per capita	5.5	6.27	7.14
Index of Business / Commute	dimensionless	0.76	1.00	1.31
Index of Recreation / Other	dimensionless	0.88	1.00	1.14

Source: (Cambridge Systematics, Inc., 2020, p. 70)

Other Sources

4.12 The article "Can We Use the NHTS to Estimate Long-Distance Travel?" (McGuckin, 2018) includes a useful summary of annual trip rates for trips of 50 miles or longer based on NHTS data – see Figure 4.3.

Figure 4.3: Annual per capita trips and share of all trips that are 50 miles or more in length, people 18 and older, with 95% confidence intervals



Source: "Can We Use the NHTS to Estimate Long-Distance Travel?" (McGuckin, 2018)

4.13 The confidence intervals are of the order of +/- 20%. The variations between surveys are slightly larger – approximately +/- 25%.

Recommended Distribution Parameters

4.14 A PERT shape 4 distribution with a minimum of 0.75, a mode of 1.0 and a maximum of 1.25 was used, maintaining the two trip purpose categories (business / commute and recreation / other), with independent distributions be used for each. There was no evidence to substantiate a difference in the distribution assumptions for the two trip purpose categories.

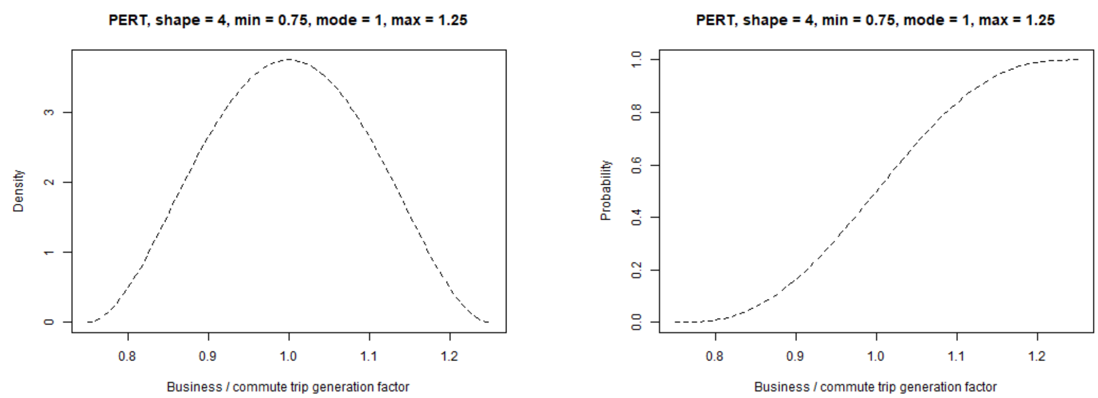
Example of Implementation

$$F_{BC}' = F_{BC} * x \qquad (5)$$

$$F_{RO}' = F_{RO} * y$$

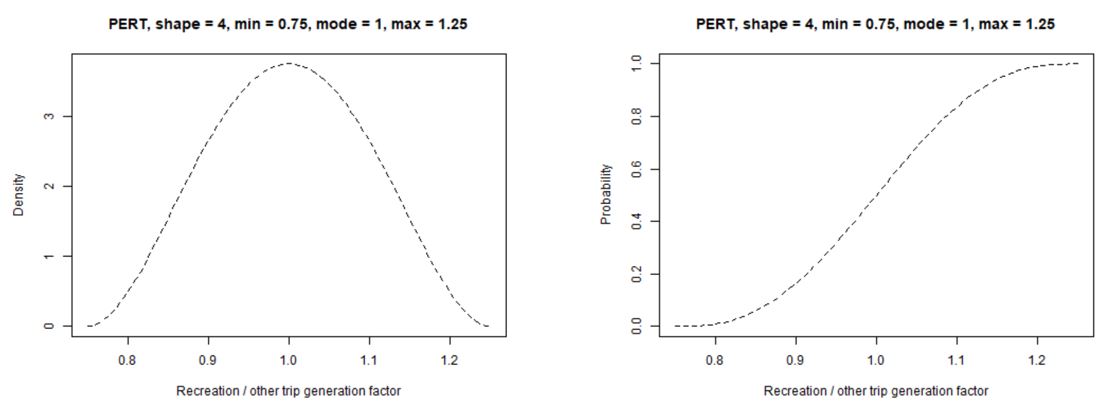
Where F_{BC} is the factor for business / commute, F_{RO} is the factor for recreation / other, and x and y are simulated random numbers with the distributions illustrated in Figure 4.4 and Figure 4.5.

Figure 4.4: Sampling distribution of business / commute trip generation factor



Source: Steer

Figure 4.5: Sampling distribution of recreation / other trip generation factor



Source: Steer

Levels of Service

Auto Operating Cost Factor

4.15 The CRRM model includes estimates of operating costs for privately owned vehicles and shared use vehicles (TNC[2] & taxi). The primary resource used in the model is updated as follows:

- The historical California all grades all formulations retail gasoline prices (EIA)
- The key indicators of the energy consumption in the transportation sector (EIA)
- Revised non-gasoline operating cost (AAA)

4.16 The forecast of privately owned vehicle operating costs includes fuel price, fuel efficiency, and non-fuel operating cost estimates. The EIA offers the estimation for these components with different scenarios. For the maximum operating cost, Steer applied the fuel efficiency under the "AEO2018 without clean power plan" scheme. The fuel efficiency estimated under the "high oil and gas resource and technology" scenario was taken for calculating the minimum operating cost.

Table 4.5: Range of auto operating cost for each forecast year by auto operating cost component (June 2018 dollars)

Year	Auto Operating Cost Component	Minimum	Base Case	Maximum
2030	Auto Operating Cost ($/mile)	0.17	0.23	0.25
2040	Auto Operating Cost ($/mile)	0.17	0.23	0.24
2050	Auto Operating Cost ($/mile)	0.17	0.23	0.24
2030	Fuel efficiency (mpg)	38.39	30.67	31.65
2040	Fuel efficiency (mpg)	39.53	35.10	36.61
2050	Fuel efficiency (mpg)	39.81	36.86	38.18
2030	Fuel price ($/gallon)	3.06	4.53	5.14
2040	Fuel price ($/gallon)	3.17	4.91	5.71
2050	Fuel price ($/gallon)	3.19	5.29	6.04
2030	Maintenance, repair, tires (per mile)	0.09	0.09	0.09
2040	Maintenance, repair, tires (per mile)	0.09	0.09	0.09
2050	Maintenance, repair, tires (per mile)	0.09	0.09	0.09

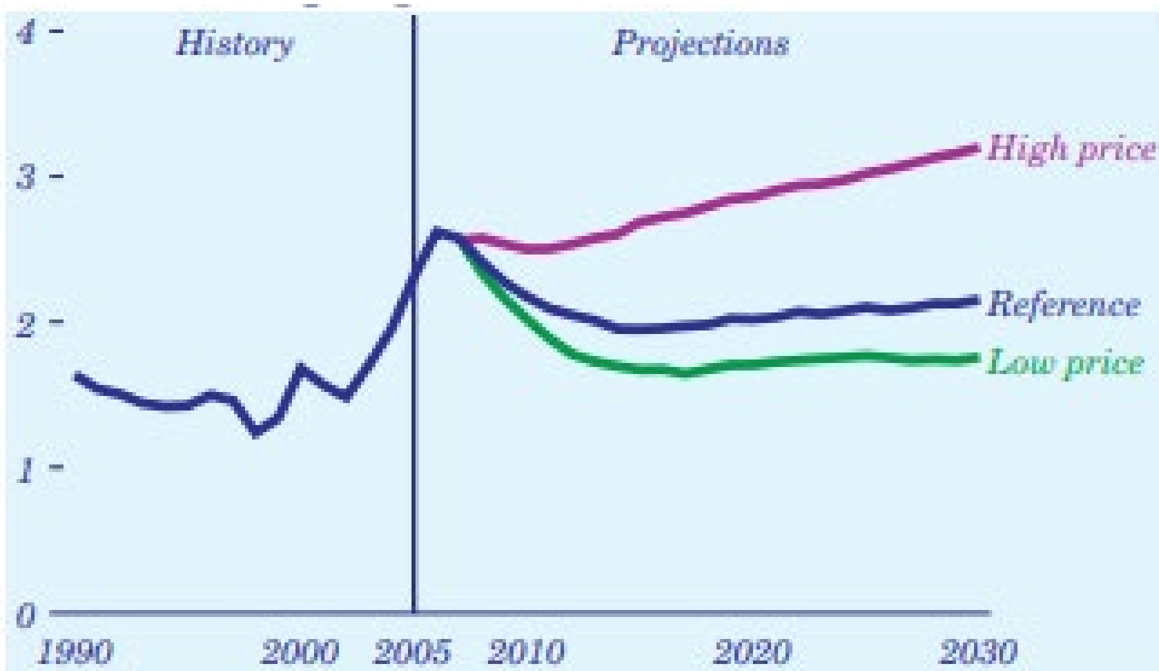
Source: Steer using data from U.S. Energy Information Administration

4.17 The variability in the above forecasts does not indicate how much error there might be in gasoline price forecasts when these are compared with historical data. Past forecasting errors are worth considering as a way of dimensioning possible future forecasting errors.

[2] TNC = Transportation Network Company

4.18 In the annual energy outlook 2020 report, EIA predicted the future gasoline price between 2007 and 2020, as shown in Figure 4.6. The high, reference and low prices suggested that distribution should be asymmetric – there seemed to be more potential for prices to rise than for them to fall.

Figure 4.6: Average U.S. delivered prices for motor gasoline, 1990-2030 (2005 dollars per gallon)



Source: "Annual Energy Outlook 2007" (EIA, 2007, p. 106)

4.19 Comparison of the forecast gasoline prices, and the observed annual average prices indicates an approximate Mean Absolute Percentage Error (MAPE) of 15% - see Table 4.6 for details of the calculation.

Table 4.6: Calculation of forecast errors in 2005 EIA forecasts for 2010 and 2020 retail gasoline prices

Base year	Forecast Year	Forecast price / gallon (2005 prices)	Observed price / gallon (nominal)	Observed price / gallon (2005 prices)	Error (%)
2005	2010	2.2	2.835	2.54	15.5%
2005	2020	2.0	2.258	1.71	-15%

Sources: "Annual Energy Outlook 2007" (EIA, 2007, p. 106), "U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)" (EIA, 2022)

Recommended Distribution Parameters

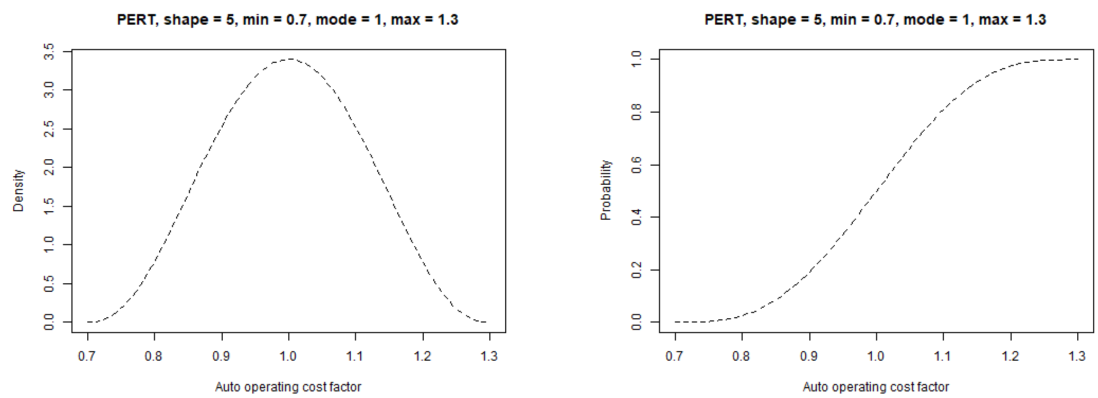
4.20 A PERT shape 5 with a minimum of 0.70, mode of 1 and maximum of 1.30 seemed appropriate. These parameters were chosen to cover a somewhat broader range than that indicated by the consideration of forecasting errors and forecasting scenarios – a broader range is necessary for the observed forecasting errors to have a reasonable chance of being sampled.

Example of Implementation

C' = C*x (6)

Where C is the operating cost and x will be a simulated random number with the distribution illustrated in Figure 4.7.

Figure 4.7: Sampling distribution of auto operating cost factor



Source: Steer

High-Speed Rail (HSR) Fare Factor

4.21 Potential variability in HSR fares was tested using a factor that has been normalized around the most likely value. 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 are used to guide the development of the maximum HSR fares.

Recommended Distribution Parameters

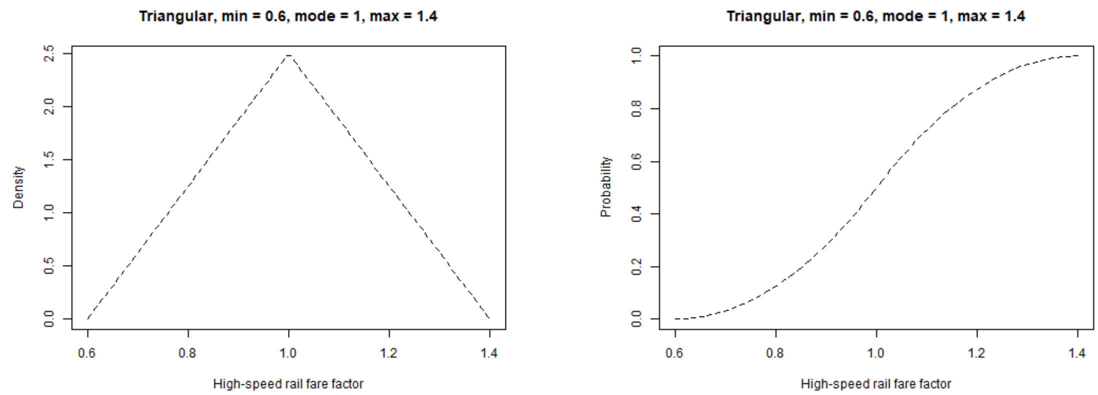
4.22 A triangular distribution was used with minimum of 0.6, mode of 1 and maximum of 1.4. This was a slightly broader range than the distribution used in the previous work (Steer, 2022a, p. 215).

Example of Implementation

Fare' = Fare*x (7)

Where x will be a simulated random number with the distribution illustrated in Figure 4.8.

Figure 4.8: Sampling distribution of the high-speed rail fare factor



Source: Steer

High-Speed Rail Frequency Factor

4.23 The assumptions for the distribution of the high-speed rail frequency factor were the same as those used in previous sensitivity analysis (Steer, 2022a, p. 216).

Recommended Distribution Parameters

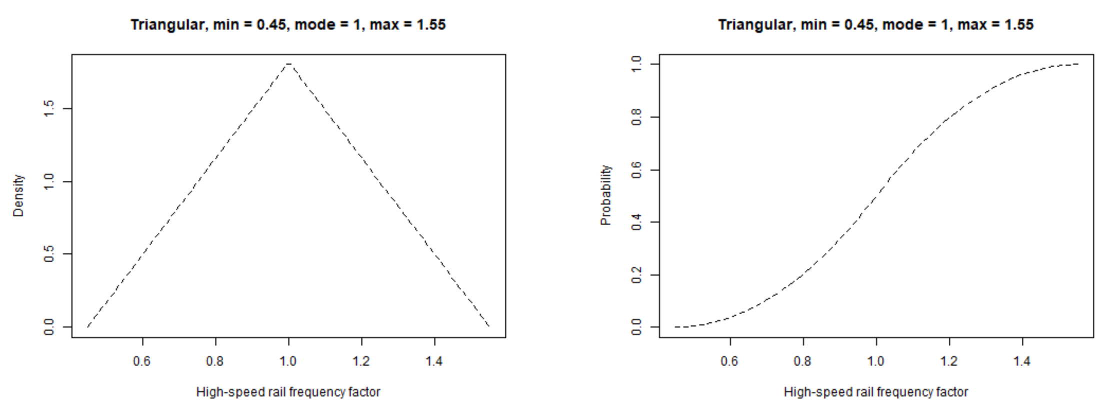
4.24 A triangular distribution was used with minimum of 0.45, mode of 1 and maximum of 1.55.

Example of Implementation

Frequency' = Frequency*x (8)

Where x will be a simulated random number with the distribution illustrated in Figure 4.9.

Figure 4.9: Sampling distribution of the high-speed rail frequency factor



Source: Steer

Air Fare Factor

4.25 Potential variability in HSR service frequencies was tested using a factor that had been normalized around the most likely value. The assumptions for the distribution of this factor were initially based on the values in "California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis" (Cambridge Systematics, Inc., 2020, p. 29). However, after reviewing recent research on air-rail competition (Zhang et al., 2019) it was concluded that the risk variable should allow for the possibility of air fare reductions as well as increases. Thus, the lower bound was defined by simply mirroring the constant.

Recommended Distribution Parameters

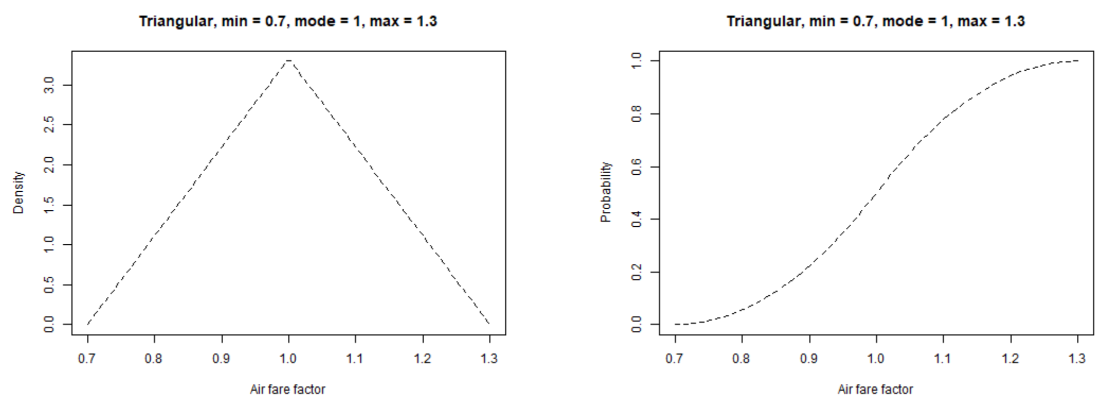
4.26 A triangular distribution was used with minimum of 0.7, mode of 1 and maximum of 1.3.

Example of Implementation

Airfare' = Airfare*x (9)

Where x will be a simulated random number with the distribution illustrated in Figure 4.10.

Figure 4.10: Sampling distribution of the air fare factor



Source: Steer

Air Service Frequency Factor

4.27 Recent research suggested that the introduction of HSR could lead to substantial air service frequency reductions on affected routes, as much as 60% (Zhang et al., 2019). However, there was also research that has found that in some cases frequencies have been maintained while reducing the volume of seats (Albalate et al., 2015). The possibility of air service frequency increasing was considered. The article "Airlines' reaction to high-speed rail entries: Empirical study of the Northeast Asian market" (Wan et al., 2016) mentioned cases where this happened, however only for long-haul journeys of over 800km where air services could compete more effectively by increasing their frequencies, which is not the case for San Francisco – Los Angeles.

Recommended Distribution Parameters

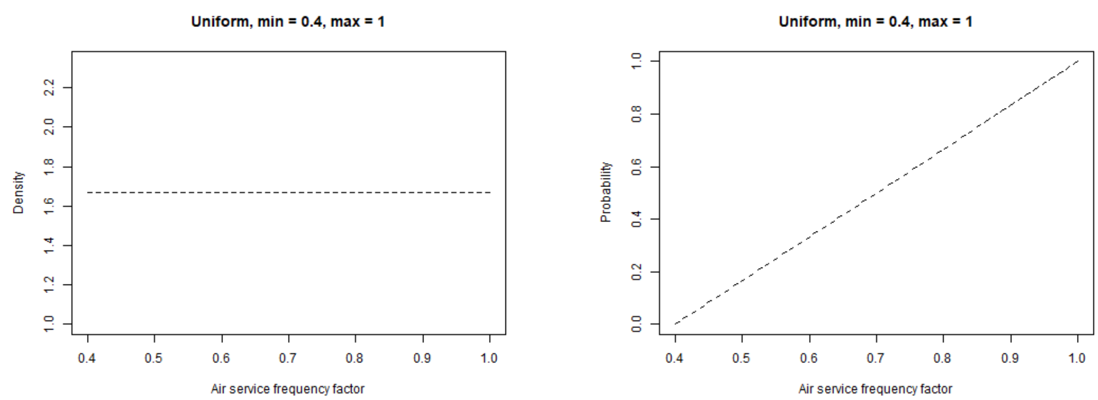
4.28 A uniform distribution with a minimum of 0.4 and a maximum of 1 was used. The uniform distribution was preferred in this case to assure sufficient density of observations across the whole range.

Example of Implementation

$$Airfrequency' = Airfrequency*x \tag{10}$$

Where x will be a simulated random number with the distribution illustrated in Figure 4.11.

Figure 4.11: Sampling distribution of the air frequency factor



Source: Steer

4.29 Additional tests were carried out to evaluate the impact of the air service frequency factor when increased by 10% and 20% for the 2040 base case model runs. Test results are reported in Data Collection chapter (see 5.9 below).

Choice Model Parameters

High-Speed Rail Constant Difference Factor

4.30 The high-speed rail constant accounts for the unobserved attributes of the HSR mode that is not captured directly by the system (e.g., with the observable attributes like fare, travel time and frequency) but also explains the desirability of HSR as a mode. It could include reliability, comfort, convenience, visibility, flexibility, safety, and other factors.

4.31 In the CRRM model, the constant is a mode specific parameter when the direct distance between zones is fixed. It is split into an HSR rural constant and an HSR urban constant. The range for the HSR constant is defined as follows:

- Minimum: The calibrated conventional rail constant is used (C_{CR})
- Base Case: The calibrated HSR constant is used (C_{HSR})
- Maximum: The calibrated HSR constant is used, plus the difference between the calibrated HSR and conventional rail constant rail constant ($2*C_{HSR} - C_{CR}$)

4.32 The above assumes that the difference C_{HSR} - C_{CR} will be positive. A similar process could be used for determining the maximum and minimum for the CRRM case, with some modification as it is expected that at least one of the two constants will vary as a function of distance function of distance instead of a single constant.

4.33 A more general formulation was used, which used the absolute value of the difference between HSR and conventional rail constants, and allowed for both constants potentially being functions of distance:

- Minimum: $C_{HSR,d} - |C_{HSR,d} - C_{CR,d}|$
- Base Case: $C_{HSR,d}$
- Maximum: $C_{HSR,d} + |C_{HSR,d} - C_{CR,d}|$

Recommended Distribution Parameters

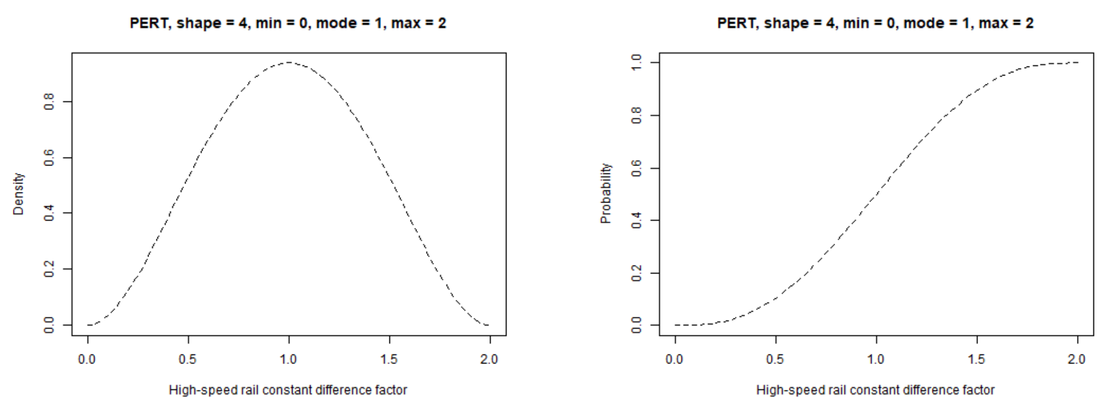
4.34 We modeled the high-speed rail constant difference factor with a PERT shape 4 distribution with a minimum of 0, a mode of 1 and a maximum of 2.

Example of Implementation

$$C_{HSR,d}' = C_{HSR,d} - |C_{HSR,d} - C_{CR,d}| + x*2|C_{HSR,d} - C_{CR,d}| \tag{11}$$

Where x is a simulated random number with the distribution illustrated in Figure 4.12.

Figure 4.12: Sampling distribution of high-speed rail constant difference factor



Source: Steer

Access/Egress Time Parameter Difference Factor

4.35 Access/egress disutility plays a significant role in the choice of travel mode. Especially in the Silicon Valley to Central Valley scenario, individuals who wish to travel primarily by transit will generally have to make one or more transfers before or after traveling on HSR.

Recommended Distribution Parameters

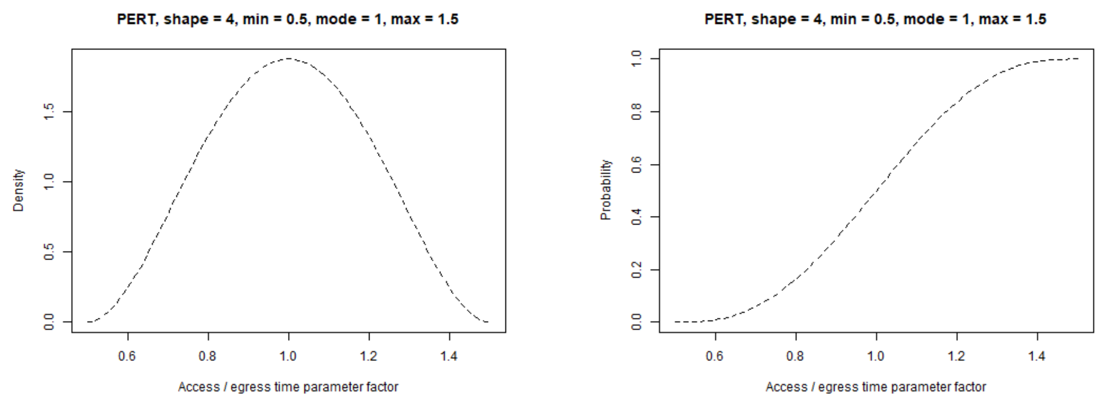
4.36 We applied a PERT shape 4 distribution with minimum of 0.5, mode of 1 and maximum of 1.5.

Example of Implementation

$$\beta' = \beta * x \tag{12}$$

Where x will be a simulated random number with the distribution illustrated in Figure 4.13.

Figure 4.13: Sampling distribution of access / egress time parameter difference factor



Source: Steer

Auto In-vehicle Time Parameter Factor

4.37 Potential variability affecting the value of auto in-vehicle time will be tested using a factor applied to the in-vehicle time parameters in the CRRM mode choice model.

4.38 The CRRM model has different in-vehicle time parameters for business, commute, leisure, and other trip purposes. The same auto in-vehicle time parameter factor would be applied for all these trip purposes.

Recommended Distribution Parameters

4.39 We applied a triangular distribution with minimum of 0.6, mode of 1 and maximum of 1.4. Both higher and lower values were tested as it can be expected that there could be changes in both directions:

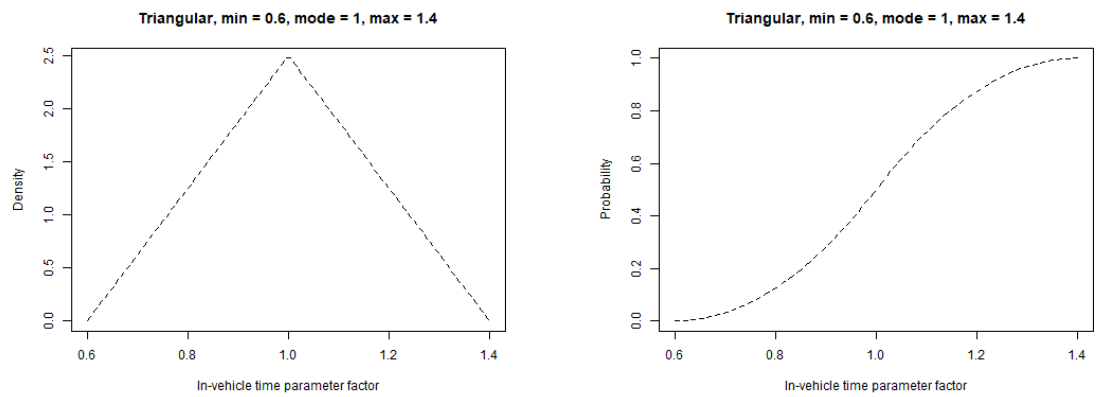
- Increased congestion will tend to increase the disutility of auto in-vehicle time (Wardman & Ibáñez, 2012)
- Increasing prevalence of autonomous vehicles will tend to reduce the disutility of auto in-vehicle time (Harb et al., 2021; Kolarova et al., 2018; Rashidi et al., 2020)

Example of Implementation

$$\beta_{IVT}' = \beta_{IVT} * x \tag{13}$$

Where x will be a simulated random number with the distribution illustrated in Figure 4.14.

Figure 4.14: Sampling distribution of auto in-vehicle time parameter factor



Source: Steer

Nest Parameter Factor

4.40 A nest parameter is used to represent correlations between the perceived utilities of options that are relatively similar, which result in substitution patterns that do not vary depending on whether options are in the same nest or not (Train, 2003, pp. 97–111).

4.41 The CRRM model uses a nest parameter to represent the relative similarities between different access-egress combinations used with public transit main modes (rail, bus, air and HSR).

> In metro areas, the model distinguishes only between “private auto”, “shared auto” and “all transit” access/egress connections to/from airports and stations. This is done at both the origin and the destination ends, allowing for nine possible access/egress mode combinations. (Steer, 2022a, p. 17)

4.42 For reasons related to the formulation of the choice model, there are limits to the possible values this parameter can take. In the case of the CRRM model, the nest parameter must be greater than 0 and less than or equal to 1. Higher values will tend to increase the demand for the public transport main modes, and lower values will tend to reduce their demand. A value of 1 would imply that each modal combination would be viewed as a completely independent mode of transport.

4.43 The base case scenario uses an assumption of 0.5 for this parameter. Given that there will sometimes be up to 9 options for any main mode, it seems highly unlikely that this parameter could be any higher, therefore this will be considered the upper limit for the range to be tested.

Recommended Distribution Parameters

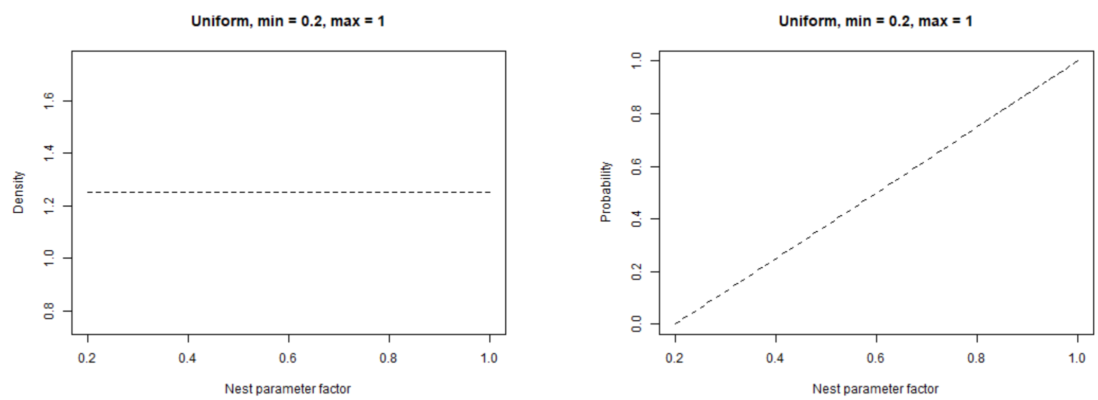
4.44 We applied a uniform distribution with minimum of 0.2 and maximum of 1. Given that the base case assumption is 0.5, this implies that the values tested for the nest parameter in the sensitivity analysis will vary between 0.1 and 0.5.

Example of Implementation

$$N' = N*x \quad (14)$$

Where N is the nest parameter and x is a simulated random number with the distribution illustrated in Figure 4.15.

Figure 4.15: Sampling distribution of nest parameter factor



Source: Steer

Transfer Penalty Factor

4.45 Transfer penalties reflect the perceived disutility and inconvenience of transferring between services, in addition to the actual transfer time (walk time, wait time, or associated transfer cost).

Recommended Distribution Parameters

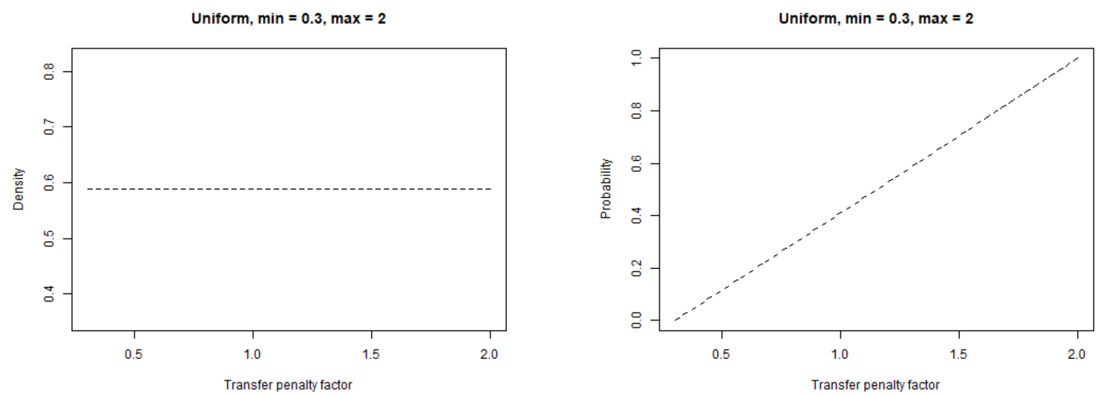
4.46 Considering the range of values used for the transfer penalty on recent projects in California, we applied a uniform distribution with minimum of 0.3 and maximum of 2.

Example of Implementation

$$TP' = TP*x \quad (15)$$

Where TP is the transfer penalty and x is a simulated random number with the distribution illustrated in Figure 4.16.

Figure 4.16: Sampling distribution of transfer penalty factor



Source: Steer

Induced Demand and Related Factors

4.47 Off-model tests were used to collect data on the sensitivity of the model to several variables that could be changed without requiring additional full model runs. These were:

- Long access / egress trips;
- Non-resident trips; and
- Induced trips.

4.48 Unlike the factors used as inputs to the full model runs, these factors were tested using evenly spaced, incremental values as shown in Table 4.7. A base case runs including induced travel and the most likely levels of long access / egress trips and non-resident trips were used to represent the factor levels of 1.0 in the table.

Table 4.7: Tests for induced demand and related factors

Factor	Number of tests	Lower limit of factor variation	Upper limit of factor variation	Values to be tested
long access / egress trips	11	0	2	0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0
non-resident trips	11	0	2	0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0
induced trips	11	0.5	1.5	0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5

Source: Steer

Implementation

4.49 The Latin Hypercube sampling was implemented using R (R Core Team, 2023). The specific packages and functions used are detailed in Table 4.8 below.

Table 4.8: R packages and functions used for Latin Hypercube sampling

Package or function	Summary of functionality	Key dependencies
CASA package (Mitrani, 2023)	R package containing several functions used for sensitivity analysis.	lhs (Carnell, 2021a, 2021b) mc2d (Pouillot et al., 2021)
samplr.R function	Produces a transformed latin hypercube and related distribution graphs, calls hypercuber.R and mc2d.	mc2d (Pouillot et al., 2021)
hypercuber.R function	hypercuber.R calls lhs	lhs (Carnell, 2021a, 2021b)

Source: Steer (2024)

5 Data Collection

5.1 The data collection was done with CRRM model runs for year 2040. While the base case scenarios for the 2040 Business Plan were run with induced demand, the data collection runs were done without induced demand. The impact of induced demand on the results were tested with off-model tests as described in section 4.47 above. Since induced demand was not included, the results from the data collection tests were based on consistent demand, hence this made it possible for all the relevant factors to be tested in relation to a consistent reference case.

5.2 Since the model was run for year 2040, the year 2030 and 2050 results were developed using growth factors from the base case runs for those years. The data collection was done for the following scenarios:

- 2040: Valley to Valley (V2V) – EMME supply scenario v2v1
- 2040: Phase 1, San Francisco to Anaheim (PH1) – EMME supply scenario ph1

5.3 The rule of thumb of 10 tests per variable used in previous sensitivity analysis for 2020 Business Plan (Cambridge Systematics, Inc., 2020; Steer, 2022b) was maintained. Table 5.1 shows the number of model runs conducted by scenario, including the base case scenarios. A total of 292 CRRM model runs were performed, and additional sensitivity tests were performed for air frequency factor (as discussed in section 5.9 below),

Table 5.1: Model runs

Year	Scenario	Risk Variables	Model Runs
2030	Valley to Valley (base case)		1
2030	Valley to Valley (base case including induced travel)		1
2030	Phase 1 (base case)		1
2030	Phase 1 (base case including induced travel)		1
2040	Valley to Valley (base case)	-	1
2040	Valley to Valley (base case including induced travel)	-	1
2040	Phase 1 (base case)	-	1
2040	Phase 1 (base case including induced travel)	-	1
2040	Valley to Valley	14	140
2040	Phase 1	14	140
2050	Valley to Valley (base case)	-	1

Year	Scenario	Risk Variables	Model Runs
2050	Valley to Valley (base case including induced travel)	-	1
2050	Phase 1 (base case)	-	1
2050	Phase 1 (base case including induced travel)	-	1
Total			**292**

Source: Steer (2024)

Model Runs

5.4 An interface between the outputs of the Latin hypercube sampling and the inputs of the model was set up and used to facilitate the model runs.

5.5 Most of the risk factors apply to California High Speed Rail and to the Brightline West as it is a High-Speed Rail mode as well. The exceptions to this are the high-speed rail fare factor and the high-speed rail frequency factor, which were applied only to the California High Speed Rail fares and frequencies.

5.6 The ridership and revenue data used for the sensitivity analysis was filtered to extract only ridership and revenue for California High Speed Rail.

Off-Model Tests

5.7 Off-model tests were conducted to collect data on the sensitivity of the few factors which were tested independently of the model. These factors are:

- Long access/egress trips;
- Non-resident trips; and
- Induced trips.

5.8 The results for the off-model tests are shown in Table 5.2 and Table 5.3 for Phase 1 and Valley to Valley respectively. The tables show the variability of HSR trips and revenue with changes in the off-model factors.

Table 5.2: Details of off-model tests for Phase 1

Factor	Tests	Lower limit of factor variation	Upper limit of factor variation	Lower limit of trips variation	Upper limit of trips variation	Lower limit of revenue variation	Upper limit of revenue variation
Long access / egress trips factor	10	-100%	100%	-13.4%	13.4%	-13.3%	13.3%
Non-resident	10	-100%	100%	-8.7%	8.7%	-8.4%	8.4%

Factor	Tests	Lower limit of factor variation	Upper limit of factor variation	Lower limit of trips variation	Upper limit of trips variation	Lower limit of revenue variation	Upper limit of revenue variation
trips factor							
Induced trips factor	10	-50%	50%	-4.2%	4.2%	-5.9%	5.9%

Source: Steer (2023)

Table 5.3: Details of off-model tests for Valley to Valley

Factor	Tests	Lower limit of factor variation	Upper limit of factor variation	Lower limit of trips variation	Upper limit of trips variation	Lower limit of revenue variation	Upper limit of revenue variation
Long access / egress trips factor	10	-100%	100%	-11.0%	11.0%	-10.8%	10.8%
Non-resident trips factor	10	-100%	100%	-9.2%	9.2%	-9.1%	9.1%
Induced trips factor	10	-50%	50%	-1.6%	1.6%	-2.3%	2.3%

Source: Steer (2023)

Air Service Frequency Factor Tests

5.9 The air service frequency factors were capped at 1 for the data collection considering that with an introduction of a new HSR service, frequency of airline trips will not increase as air is direct competitor for the HSR. However, in order to test the impact of increased air service frequency on HSR ridership and revenue, the sensitivity tests were done with the air fare frequency factor of 1.1 and 1.2 while keeping all other factors at base level. The results of the sensitivities are presented in Table 5.4. With an increase in air service frequency, ridership as well as revenue on HSR reduces.

Table 5.4: Air service frequency factor test results

Year	Scenario	Air service frequency factor	Yearly ridership (million)	Yearly revenue (YOE$ million)	Ridership change	Revenue change
2040	Phase 1	1.0	28.39	3,576.00	-	-

Year	Scenario	Air service frequency factor	Yearly ridership (million)	Yearly revenue (YOE$ million)	Ridership change	Revenue change
2040	Phase 1	1.1	26.69	3,209.25	-6.0%	-10.3%
2040	Phase 1	1.2	26.66	3,203.88	-6.1%	-10.4%
2040	Valley to Valley	1.0	12.22	1,170.80	-	-
2040	Valley to Valley	1.1	11.80	1,113.87	-3.4%	-4.9%
2040	Valley to Valley	1.2	11.79	1,112.27	-3.5%	-5.0%

Source: Steer

5.10 It can be observed that an increase in air service frequency by 10% results in 6% and 3% reduction in ridership in 2040 Phase 1 and 2040 Valley to Valley, respectively. Interestingly when air service frequency is increased to 20%, there is little to no impact compared to the 10% increase. The revenue impact is somewhat larger, probably since the trips being captured by air tend to be longer distance trips.

6 Meta Modeling

Introduction

6.1 The meta modeling methodology reported by Cambridge Systematics (Cambridge Systematics, Inc., 2018, p. 42) and used in the previous sensitivity analysis by Steer (Steer, 2022b) was updated to account for potential heteroskedasticity in the errors of the ordinary least squares (OLS) regression model (White, 1980).

6.2 Ridership and revenue were modeled separately for Phase 1 and Valley to Valley, resulting in four metamodels:

- 2040 Phase 1 ridership model;
- 2040 Phase 1 revenue model;
- 2040 Valley to Valley ridership model; and
- 2040 Valley to Valley revenue model.

6.3 Each model used two complementary techniques:

- An ordinary least squares (OLS) regression model that was fitted to the data points generated by the full model runs; and
- A Gaussian process model (GPM) that was fitted to the residuals of the OLS model (Cambridge Systematics, Inc., 2018, p. 105).

6.4 The GPM work was done using the “kernlab” package for R (Karatzoglou et al., 2004).

6.5 The results were evaluated using k-fold cross-validation, using the “caret” package for R (Kuhn, 2008).

6.6 Table 6.1 below summarizes the CASA package and the R functions that were developed and built within the package for the metamodeling.

Table 6.1: R functions from CASA package used for metamodeling

Function	Summary of functionality	Key dependencies
model_ols	Estimates an OLS model, corrects the standard errors for heteroskedasticity	(R Core Team, 2024), sandwich (Zeileis et al., 2020), lmtest (Zeileis & Hothorn, 2002)
model_gpm	Estimates a Gaussian Process Model (GPM)	kernlab (Karatzoglou et al., 2004)
model_ols_gpm	Model with OLS and GPM components	model_ols, model_gpm, caret (Kuhn, 2008)

Source: Steer

6.7 The models were estimated using all the data combined as follows:

Table 6.2: Data used for metamodeling (2040)

Scenario	Data used
Phase 1	Base model run (1 observation) Full model run tests (140 observations) Off-model tests (30 observations)
Valley to Valley	Base model run (1 observation) Full model run tests (140 observations) Off-model tests (30 observations)

Source: Steer

6.8 The data set used in model estimation consisted of 171 observations for each model.

6.9 The CRRM tests were undertaken without induced travel to ensure a consistent point of reference. The ridership and revenue results of these tests were factored to allow for the expected impacts of induced travel using the results of base case runs with and without induced travel. The induced travel factors are detailed in Table 6.3 below.

Table 6.3: Induced travel factors (2040)

Scenario	Variable	Value without induced travel (million)	Value with induced travel (million)	Induced travel factor
Phase 1	Ridership	25.99	28.39	1.09
Phase 1	Revenue (YOE$)	4,066.21	4,614.76	1.13
Valley to Valley	Ridership	11.83	12.22	1.03
Valley to Valley	Revenue (YOE$)	1,441.04	1,510.90	1.05

Source: Steer (2024)

6.10 Growth factors for ridership and revenue were used to represent changes between the 2040 scenarios and the corresponding scenarios for 2030 and 2050. These factors are provided in Table 6.4, together with the reference values for ridership and revenue used to derive them. Note that in this case revenue is shown in 2023$ (not YOE$) because a constant reference year is needed for the growth factors to be meaningful, and this is how the growth factors were calculated – the conversion of the results to YOE$ was done at a later stage of the process.

Table 6.4: Growth factors

Scenario	Variable	Reference value 2030 (million)	Reference value 2040 (million)	Reference value 2050 (million)	Growth factor 2030	Growth factor 2040	Growth factor 2050
Phase 1	Ridership	27.57	28.39	29.01	0.9713	1.0000	1.0219
Phase 1	Revenue (2023$)	2,390.60	2,456.51	2,504.30	0.9732	1.0000	1.0195

Scenario	Variable	Reference value 2030 (million)	Reference value 2040 (million)	Reference value 2050 (million)	Growth factor 2030	Growth factor 2040	Growth factor 2050
Valley to Valley	Ridership	11.80	12.22	12.54	0.9654	1.0000	1.0264
Valley to Valley	Revenue (2023$)	780.09	804.28	821.46	0.9699	1.0000	1.0214

Source: Steer (2024)

Results

6.11 Table 6.5 through Table 6.8 show the results of the meta models for 2040. Only the 2040 models are reported here, because the 2030 and 2050 models are essentially the same as these, based on the same data with only very minor differences due to the growth factors that were applied.

Table 6.5: OLS model for 2040 Phase 1 ridership

Description	Variable	Estimate	Std. Error (HC1)	t value (HC1)
Constant	constant	8.677	0.0396	219.17
High-speed rail constant difference factor	hsrcodifa	0.262	0.0033	80.13
Business / commute trip generation factor	bctrgefa	0.153	0.0125	12.31
Recreation / other trip generation factor	rotrgefa	0.739	0.0137	54.04
Auto operating cost factor	autopcofa	0.225	0.0127	17.70
High-speed rail fare factor	hsrfafa	-0.600	0.0078	-76.97
High-speed rail frequency factor	hsrfqfa	0.117	0.0057	20.43
Air fare factor	airfafa	0.051	0.0115	4.46
Air service frequency factor	airfrfa	-0.012	0.0064	-1.82
Access / egress time parameter factor	acegtipafa	-0.662	0.0079	-83.87
Population and households forecast factor	pohofofa	0.868	0.0240	36.11
Employment forecast factor	emfofa	0.302	0.0254	11.91
Auto in-vehicle time parameter factor	ivtpafa	0.603	0.0090	66.84
Nest parameter factor	nestpafa	0.268	0.0052	51.78
Transfer penalty factor	transpenfa	-0.040	0.0027	-14.68
Long access/egress trips factor	exloacegfa	0.135	0.0023	59.00
Non-resident trips factor	visitravfa	0.087	0.0006	150.34
Induced trips factor	indutripsfa	0.084	0.0023	37.46

Notes: N = 171. Residual standard error: 0.01412 on 153 degrees of freedom. Multiple R-squared: 0.9963. Adjusted R-squared: 0.9959. F-statistic: 2414 on 17 and 153 DF, p-value: < 2.2e-16. Source: Steer

Table 6.6: OLS model for 2040 Phase 1 revenue

Description	Variable	Estimate	Std. Error (HC1)	t value (HC1)
Constant	constant	11.884	0.0583	203.72
High-speed rail constant difference factor	hsrcodifa	0.326	0.0050	65.62
Business / commute trip generation factor	bctrgefa	0.134	0.0209	6.42
Recreation / other trip generation factor	rotrgefa	0.704	0.0171	41.14
Auto operating cost factor	autopcofa	0.325	0.0139	23.37
High-speed rail fare factor	hsrfafa	0.310	0.0152	20.42
High-speed rail frequency factor	hsrfqfa	0.123	0.0071	17.36
Air fare factor	airfafa	0.088	0.0141	6.28
Air service frequency factor	airfrfa	-0.015	0.0078	-1.86
Access / egress time parameter factor	acegtipafa	-0.647	0.0109	-59.41
Population and households forecast factor	pohofofa	0.773	0.0319	24.24
Employment forecast factor	emfofa	0.290	0.0297	9.76
Auto in-vehicle time parameter factor	ivtpafa	0.860	0.0112	76.99
Nest parameter factor	nestpafa	0.274	0.0063	43.60
Transfer penalty factor	transpenfa	-0.055	0.0037	-14.82
Long access/egress trips factor	exloacegfa	0.133	0.0021	62.42
Non-resident trips factor	visitravfa	0.084	0.0036	23.09
Induced trips factor	indutripsfa	0.119	0.0084	14.08

Notes: N = 171. Residual standard error: 0.01949 on 153 degrees of freedom. Multiple R-squared: 0.9941, Adjusted R-squared: 0.9935. F-statistic: 1520 on 17 and 153 DF, p-value: < 2.2e-16. Source: Steer.

Table 6.7: OLS model for 2040 Valley to Valley ridership

Description	Variable	Estimate	Std. Error (HC1)	t value (HC1)
Constant	constant	8.278	0.0766	108.05
High-speed rail constant difference factor	hsrcodifa	0.099	0.0062	15.89
Business / commute trip generation factor	bctrgefa	0.153	0.0273	5.61
Recreation / other trip generation factor	rotrgefa	0.686	0.0266	25.76
Auto operating cost factor	autopcofa	0.202	0.0209	9.70
High-speed rail fare factor	hsrfafa	-0.468	0.0147	-31.81
High-speed rail frequency factor	hsrfqfa	0.136	0.0127	10.70
Air fare factor	airfafa	0.057	0.0188	3.02
Air service frequency factor	airfrfa	-0.026	0.0101	-2.53
Access / egress time parameter factor	acegtipafa	-0.581	0.0123	-47.41

Description	Variable	Estimate	Std. Error (HC1)	t value (HC1)
Population and households forecast factor	pohofofa	0.851	0.0412	20.66
Employment forecast factor	emfofa	0.320	0.0397	8.04
Auto in-vehicle time parameter factor	ivtpafa	0.395	0.0180	21.92
Nest parameter factor	nestpafa	0.256	0.0076	33.65
Transfer penalty factor	transpenfa	-0.174	0.0056	-31.23
Long access/egress trips factor	exloacegfa	0.110	0.0024	46.19
Non-resident trips factor	visitravfa	0.092	0.0017	54.87
Induced trips factor	indutripsfa	0.032	0.0001	217.00

Notes: N = 171. Residual standard error: 0.02559 on 153 degrees of freedom. Multiple R-squared: 0.9846, Adjusted R-squared: 0.9828. F-statistic: 573.6 on 17 and 153 DF, p-value: < 2.2e-16. Source: Steer

Table 6.8: OLS model for 2040 Valley to Valley revenue

Description	Variable	Estimate	Std. Error (HC1)	t value (HC1)
Constant	constant	11.336	0.1116	101.53
High-speed rail constant difference factor	hsrcodifa	0.097	0.0091	10.65
Business / commute trip generation factor	bctrgefa	0.111	0.0375	2.96
Recreation / other trip generation factor	rotrgefa	0.681	0.0375	18.16
Auto operating cost factor	autopcofa	0.281	0.0301	9.34
High-speed rail fare factor	hsrfafa	0.464	0.0240	19.32
High-speed rail frequency factor	hsrfqfa	0.223	0.0176	12.67
Air fare factor	airfafa	0.055	0.0259	2.13
Air service frequency factor	airfrfa	-0.057	0.0152	-3.76
Access / egress time parameter factor	acegtipafa	-0.568	0.0174	-32.67
Population and households forecast factor	pohofofa	0.825	0.0548	15.07
Employment forecast factor	emfofa	0.301	0.0548	5.50
Auto in-vehicle time parameter factor	ivtpafa	0.606	0.0240	25.23
Nest parameter factor	nestpafa	0.261	0.0118	22.15
Transfer penalty factor	transpenfa	-0.233	0.0076	-30.59
Long access/egress trips factor	exloacegfa	0.109	0.0027	40.05
Non-resident trips factor	visitravfa	0.091	0.0020	44.94
Induced trips factor	indutripsfa	0.048	0.0011	43.38

Notes: N = 171. Includes HSR bus revenue. Residual standard error: 0.03619 on 153 degrees of freedom. Multiple R-squared: 0.9755, Adjusted R-squared: 0.9728. F-statistic: 358.8 on 17 and 153 DF, p-value: < 2.2e-16. Source: Steer.

6.12 It should be noted that the reported standard errors and corresponding t-statistic values of the parameter estimates account for potential heteroskedasticity, therefore avoiding overconfidence in the results.

6.13 All four models have adjusted R-squared values of 0.97 or greater, indicating that the estimated models fit the ridership and revenue data very well.

6.14 The corrected standard errors and corresponding t values of the estimated parameters indicate that they are statistically significant, with the exception the of air service frequency factor being insignificant in the Phase 1 ridership and revenue models at the 95% confidence level – that said, these parameters seem reasonable in magnitude and would be significant at the 90% confidence level, therefore they have been retained.

6.15 The parameters associated with the new variables that were not tested in previous versions of the sensitivity analysis both came out with the expected signs and with t-values indicating that the estimates are highly significant, in all four models. Increasing the transfer penalty tends to reduce HSR ridership and revenue (as expected) and increasing the nest parameter tends to increase HSR ridership and revenue (also as expected). Increasing the nest parameter tends to favour HSR because it represents the available access-egress mode combinations being perceived as more differentiated and therefore appealing to more people.

6.16 The Gaussian process model with a Gaussian Radial Basis kernel function was estimated using kernlab (Karatzoglou et al., 2004). Caret (Kuhn, 2008) was used to train the Gaussian process model using a grid search and K-fold cross-validation, and to estimate the optimum value of the hyperparameter sigma. The optimum value of sigma was found by scanning all candidate values between 0.01 and 10 in increments of 0.01. The optimal value was taken as that which minimized the Root Mean Square Error (RMSE) in the cross validation. A value of 10 was used for K in the K-fold cross-validation. Gaussian process model results are presented in Table 6.9 below.

Table 6.9: Results of Gaussian process model training (2040)

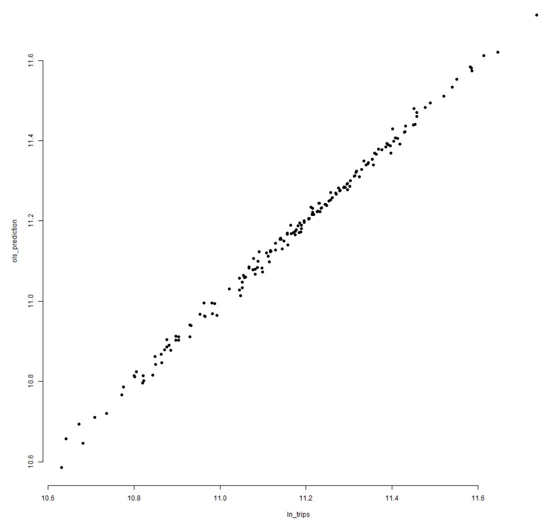
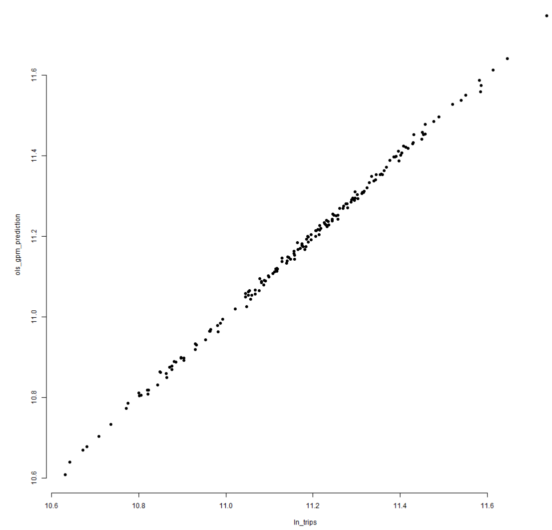
	Phase 1		Valley to Valley	
	Ridership model	**Revenue model**	**Ridership model**	**Revenue model**
Train instances	171	171	171	171
Optimum value of sigma	0.08	0.33	0.07	0.08
Train error	0.287	0.243	0.312	0.297

Source: Steer (2024)

6.17 Scatterplots showing the ridership and revenue data plotted against the model predicted values are presented for the following cases:

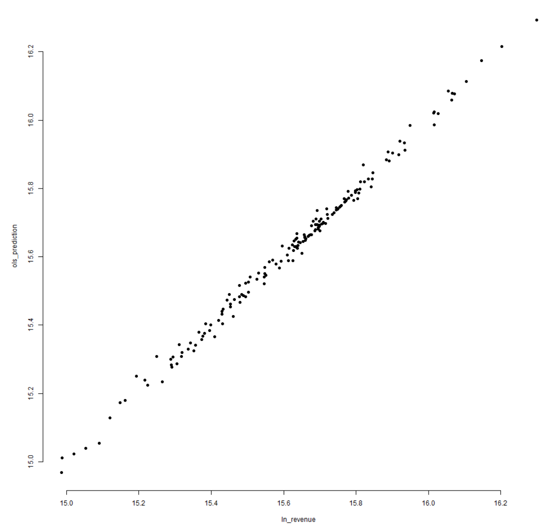
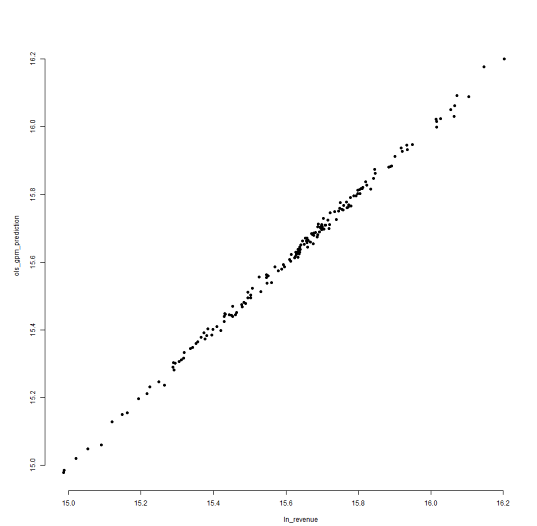
- OLS only estimates; and
- OLS + GPM estimates.

Figure 6.1: Scatterplots of estimated versus observed ridership for 2040 Phase 1 (natural log scale)

OLS Model	OLS + GPM Model
	

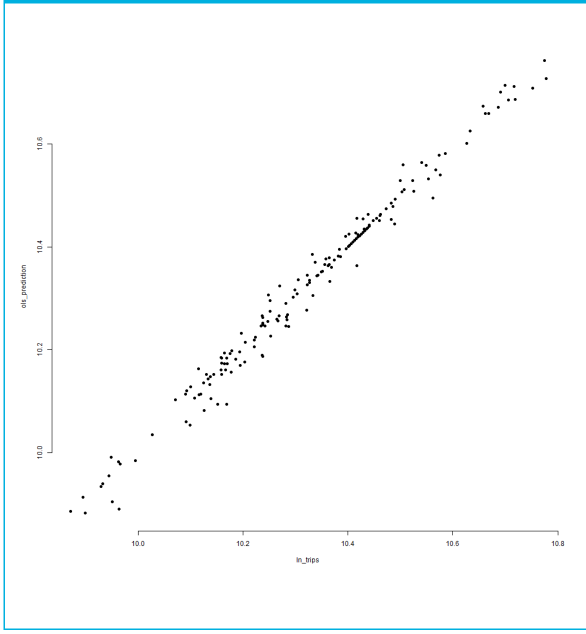
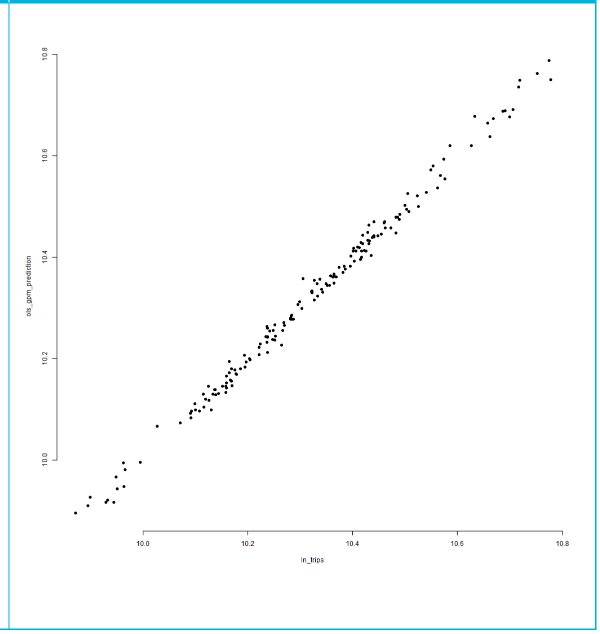
Source: Steer (2024)

Figure 6.2: Scatterplots of estimated versus observed revenue for 2040 Phase 1 (natural log scale)

OLS Model	OLS + GPM Model
	

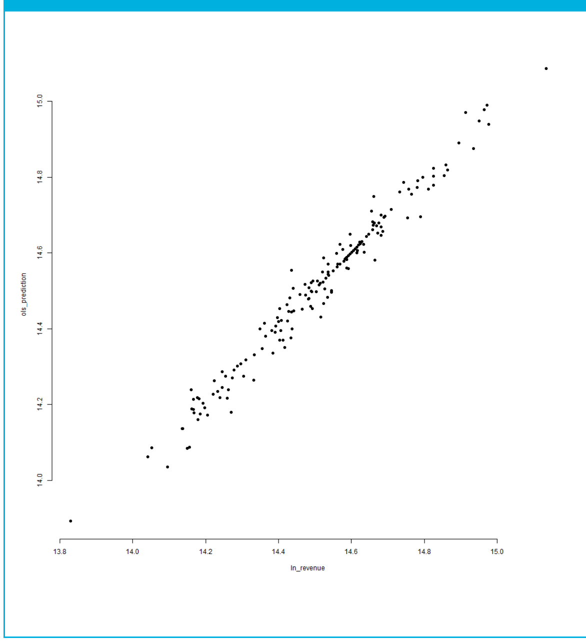
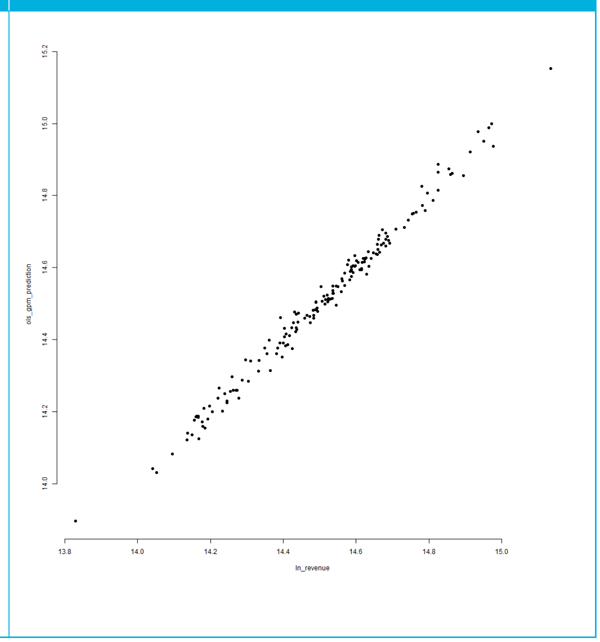
Source: Steer (2024)

Figure 6.3: Scatterplots of estimated versus observed ridership for 2040 Valley to Valley (natural log scale)

OLS Model	OLS + GPM Model
	

Source: Steer (2024)

Figure 6.4: Scatterplots of estimated versus observed revenue for 2040 Valley to Valley (natural log scale)

OLS Model	OLS + GPM Model
	

Source: Steer (2024)

6.18 As expected, the combination of OLS and GPM models results in a significantly better fit compared to using only the OLS model, in all cases.

7 Monte Carlo Simulation

Introduction

7.1 The Monte Carlo simulation involves simulating draws from specific random distributions for each of the variables of interest, considering correlations between the distributions in some cases. The resulting data is then used to apply the OLS and GPM models described in the previous chapter, to estimate corresponding values of ridership and revenue. This process differs from the Meta Modeling in that the models are applied to values that are different from those for which data was collected, and a much larger number of data points are simulated (100,000).

Methodology

7.2 The distributions that were used for the Monte Carlo simulation included all the variables tested in the full model runs, plus the variables included in the off-model tests. Table 7.1 lists the variables and their limits as tested in the risk assessment.

Table 7.1: Risk variable distributions for Monte Carlo simulation (2040 Phase 1, 2040 Valley to Valley)

Index	Description	Variable	Base	Min	Mode	Max	Distribution	Shape
1	Population and households forecast factor	pohofofa	1	0.85	1	1.15	PERT	4
2	Employment forecast factor	emfofa	1	0.85	1	1.15	PERT	4
3	Business / commute trip generation factor	bctrgefa	1	0.75	1	1.25	PERT	4
4	Recreation / other trip generation factor	rotrgefa	1	0.75	1	1.25	PERT	4
5	Auto operating cost factor	autopcofa	1	0.7	1	1.3	PERT	5
6	High-speed rail fare factor	hsrfafa	1	0.6	1	1.4	Triangular	
7	High-speed rail frequency factor	hsrfqfa	1	0.45	1	1.55	Triangular	
8	Air fare factor	airfafa	1	0.7	1	1.3	Triangular	

Index	Description	Variable	Base	Min	Mode	Max	Distribution	Shape
9	Air service frequency factor	airfrfa	1	0.4		1	Uniform	
10	High-speed rail constant difference factor	hsrcodifa	1	0	1	2	PERT	4
11	Access / egress time parameter factor	acegtipafa	1	0.5	1	1.5	PERT	4
12	Auto in-vehicle time parameter factor	ivtpafa	1	0.6	1	1.4	Triangular	
13	Nest parameter factor	nestpafa	1	0.2		1.0	Uniform	
14	Transfer penalty factor	transpenfa	1	0.3		1.7	Uniform	
15	Long access/egress trips factor	exloacegfa	1	0		2	Uniform	
16	Non-resident trips factor	visitravfa	1	0		2	Uniform	
17	Induced trips factor	indutripsfa	1	0		2	Uniform	

Source: Steer

7.3 The upper limit of Transfer penalty factor was changed to 1.7 to ensure that the median value generated from the Monte Carlo simulation would align with the base value of 1. There are two other variables where the tested distributions were asymmetric around the base value of 1 – air service frequency factor and nest parameter factor. It was not possible to center them on 1 without extrapolating beyond the tested ranges, so it can be expected that the median values of ridership and revenue produced by the Monte Carlo Simulation might differ somewhat from the base case values.

7.4 Correlations between some of the distributions were included in the test, as shown in Table 7.2 below. Specifically, there was a 75% correlation between the population & households forecast factor and the employment forecast factor, and there was also a 50% correlation between the business / commute trip generation factor and the recreation / other trip generation factor.

Table 7.2: Correlation matrix tested with Monte Carlo simulation (Both 2040 Phase 1 and 2040 Valley to Valley)

Var.	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
V1	1	0.75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V2	0.75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V3	0	0	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0
V4	0	0	0.5	1	0	0	0	0	0	0	0	0	0	0	0	0	0
V5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
V6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
V7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
V8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
V9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
V10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
V11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
V12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
V13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
V14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
V15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
V16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
V17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Source: Steer

7.5 Table 7.3 summarizes the R functions from the CASA package that were developed for the Monte Carlo simulation.

Table 7.3: R Functions for Monte Carlo simulation

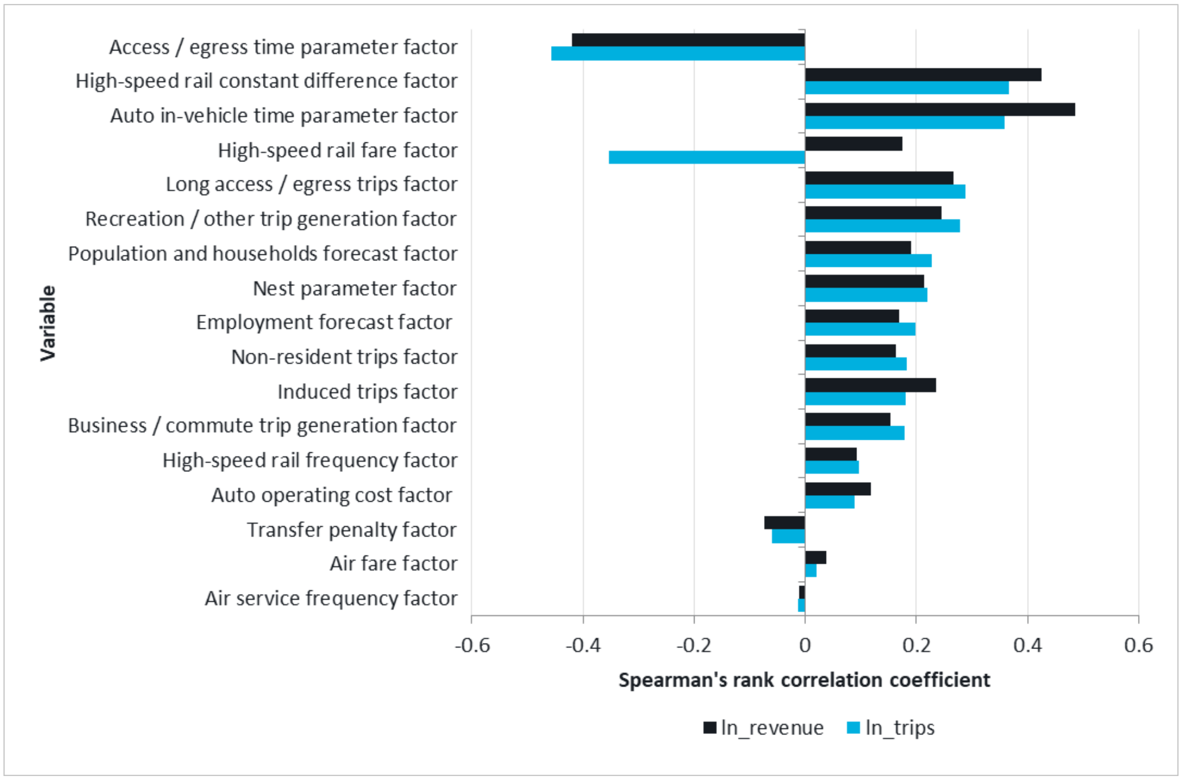
Function	Summary of functionality	Key dependencies
mcr	Monte Carlo simulation using mc2d package	mc2d (Pouillot et al., 2023)
model_ols_gpm	After estimating OLS and GPM models, computes the predictions based on the mcr function generated sample.	model_ols, model_gpm

Source: Steer

7.6 The Monte Carlo simulation was implemented considering the 17 variables and correlation structure detailed above, with 100,000 draws. The process of drawing from the distributions was undertaken in one go, whereas the application of the models was done in 10 chunks of 10,000 observations each, accumulating the results at the end of this process. It was found that this approach made better use of the available computing resources.

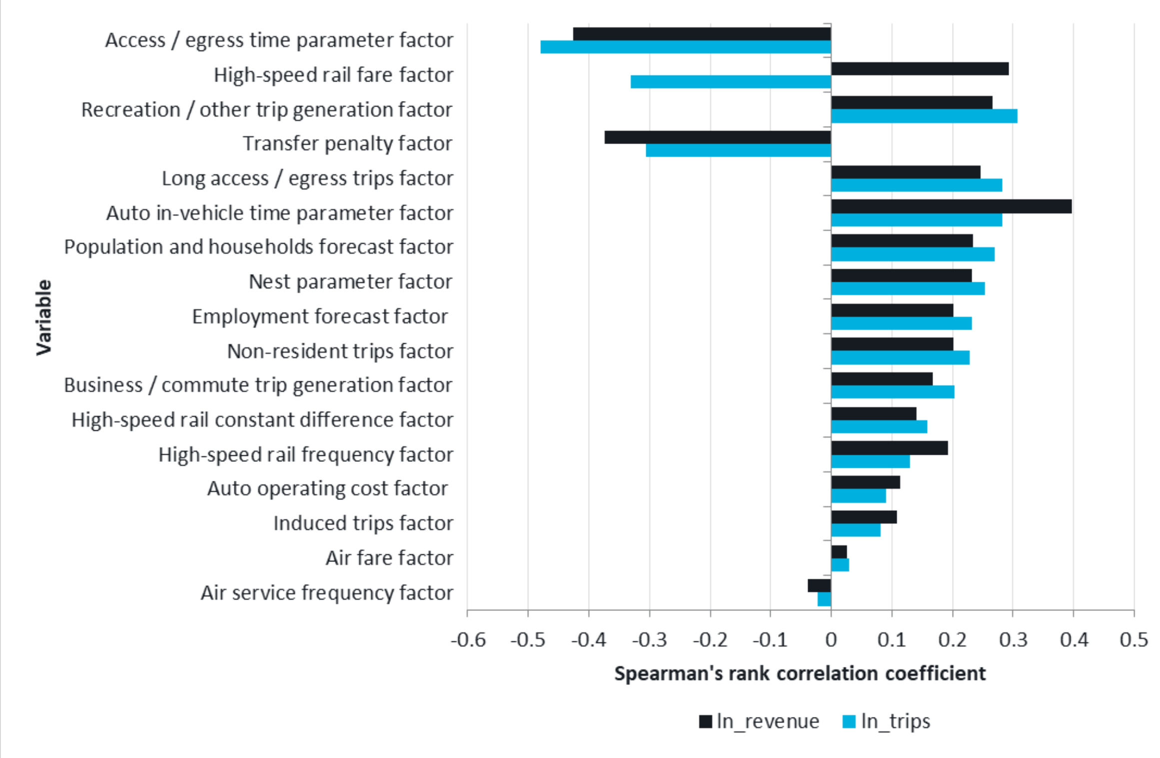
7.7 As a check on the results, correlation coefficients were calculated for the pairs of variables formed by taking each dependent variable and each of the independent variables. The correlation coefficients are defined on a scale from -1 (perfect negative correlation, e.g. when one goes up the other goes down) to +1 (perfect positive correlation), with 0 meaning no correlation. The results of this analysis give an idea of the relative strength and direction of the correlation between each pair of variables and can be seen in Figure 7.1 and Figure 7.2.

Figure 7.1: Correlation coefficients of variables with log ridership and log revenue – 2040 Phase 1



Source: Steer (2024)

Figure 7.2: Correlation coefficients of variables with log ridership and log revenue – 2040 Valley to Valley



Notes: Revenue includes HSR revenue and also HSR bus revenue. Source: Steer (2024)

7.8 The results of this simple analysis show that the relative magnitudes of the correlations look sensible as do the directions of the correlations. The correlations for ridership and revenue are generally similar with the exception of the variable High speed rail fare factor, which is negatively correlated with ridership but positively correlated with revenue – which makes sense, as higher fares tend to produce less ridership but more revenue when demand is inelastic.

Results

7.9 The revenue figures for all tests are presented in Year of Expenditure USD (YOE$). The annual inflation rate assumed for future-year YOE$ are from the California Consumer Price Index (California Department of Finance) and the United States Federal Reserve Inflation Target. This resulted in the index values shown in Table 7.4 for the conversion of future year – originally calculated in 2023$ – into YOE$ values.

Table 7.4: Price index for converting 2023 USD into YOE values for future years

Year	Years from 2023	Price index
2023	0	1.0000
2030	7	1.1942
2040	17	1.4557
2050	27	1.7745

Source: DB ECO based on CPI-U CA Index (All Urban Consumers) for 2024-2026 (California Department of Finance) and FOMC-PCE Inflation for 2027-2050 (United States Federal Reserve).

2030

7.10 Ridership and revenue input data for the 2030 meta model was based on the data collected with the 2040 CRRM model, scaled with growth factors as described in 6.10 above. However, the Monte Carlo simulation itself was performed with an independent set of 100,000 draws from the random variable distributions described above in Table 7.1 and Table 7.2, for each scenario. Therefore, the results of the Monte Carlo simulation may present some differences even in relative terms compared to the corresponding results for the other years.

2030 Phase 1

7.11 The results of the Monte Carlo simulation for 2030 Phase 1 ridership and revenue are summarized in Table 7.5.

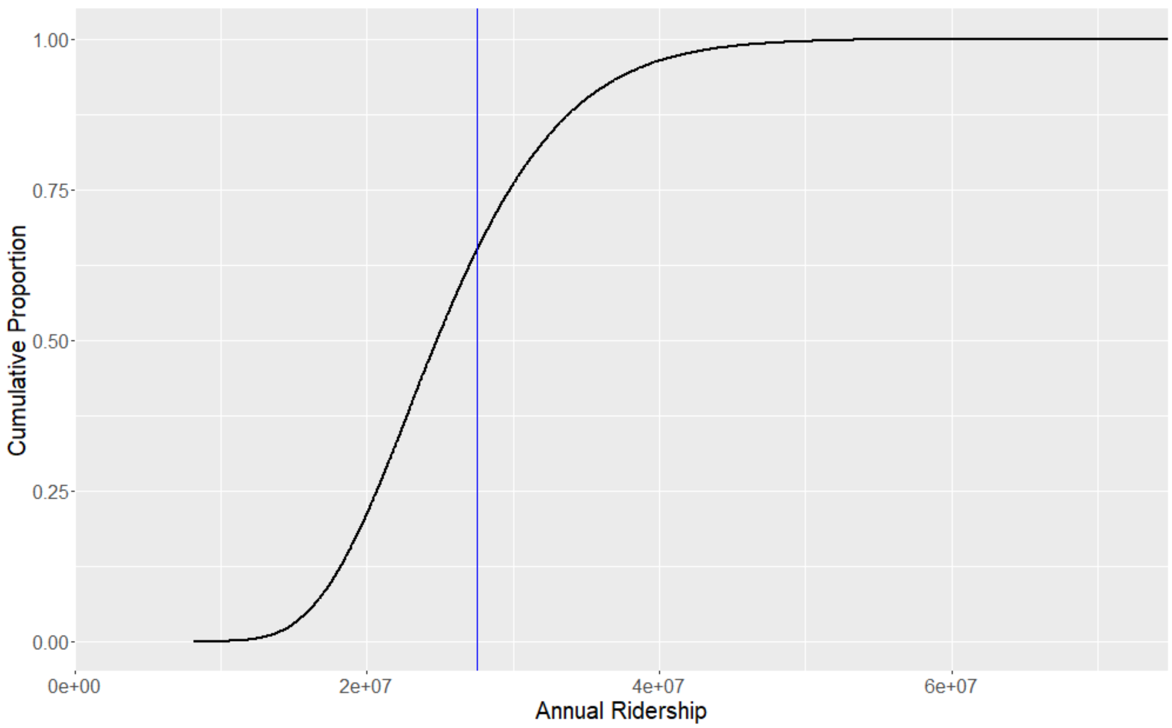
Table 7.5: Summary of Monte Carlo simulation for 2030 Phase 1

Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
Base case	27.57	2,854.85	1.00	1.00
p0	8.14	756.35	0.30	0.26
p1	13.41	1,322.45	0.49	0.46
p10	17.61	1,766.03	0.64	0.62
p25	20.69	2,099.43	0.75	0.74
p50	24.80	2,546.33	0.90	0.89
p75	29.79	3,093.78	1.08	1.08
p90	34.99	3,673.48	1.27	1.29
p99	45.96	4,900.36	1.67	1.72
p100	74.83	8,036.53	2.71	2.82

Source: Steer (2024)

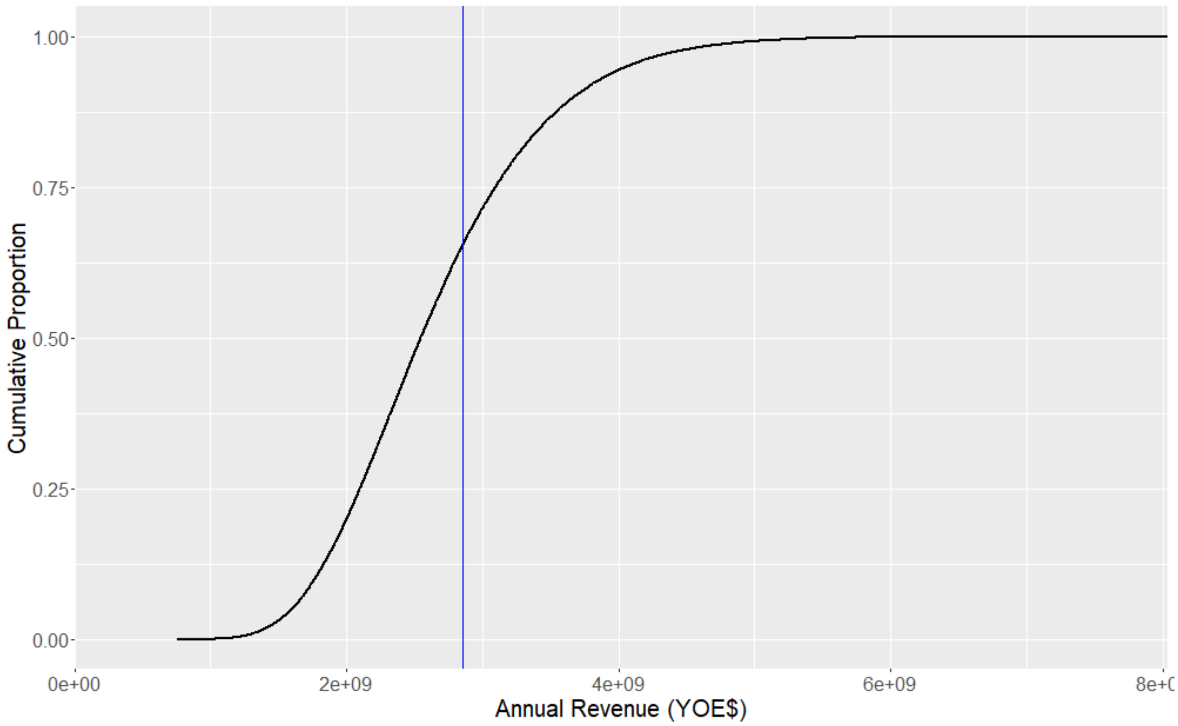
7.12 The median values of ridership and revenue are somewhat lower than the corresponding base case values. This is due to a few variables that were modeled asymmetrically in the sensitivity analysis – the air service frequency factor (varied from 0.4 to 1.0) can be expected to increase HSR demand, and the nest parameter factor (varied from 0.2 to 1.0) can be expected to reduce it. The net effect of these two factors tends to slightly reduce HSR ridership and revenue compared to the base case.

Figure 7.3: Distribution of ridership for 2030 Phase 1 produced by the Monte Carlo simulation



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.4: Distribution of revenue for 2030 Phase 1 produced by the Monte Carlo simulation



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

2030 Valley to Valley

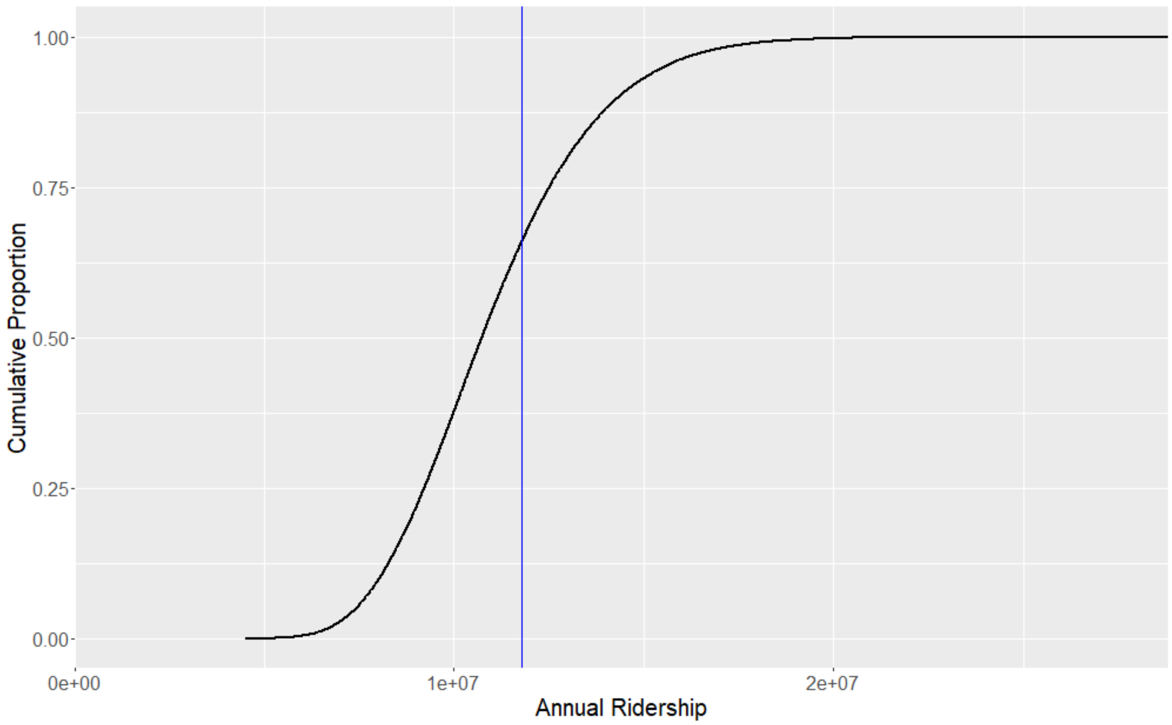
7.13 The results of the Monte Carlo simulation for 2030 Valley to Valley ridership and revenue are summarized in Table 7.6.

Table 7.6: Summary of Monte Carlo simulation for 2030 Valley to Valley

Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
Base case	11.80	960.40	1.00	1.00
p0	4.51	322.17	0.38	0.34
p1	6.41	496.20	0.54	0.52
p10	8.04	640.20	0.68	0.67
p25	9.22	743.39	0.78	0.77
p50	10.73	880.44	0.91	0.92
p75	12.51	1,043.19	1.06	1.09
p90	14.34	1,215.00	1.22	1.27
p99	18.00	1,561.52	1.53	1.63
p100	28.83	2,451.69	2.44	2.55

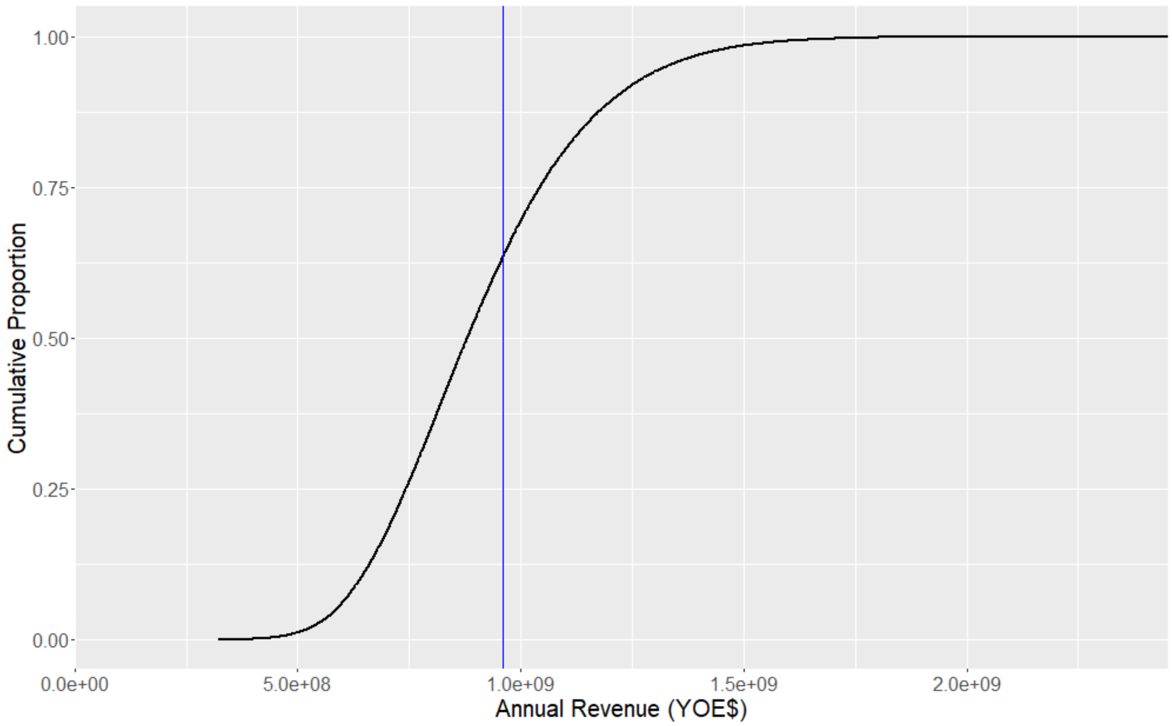
Notes: Revenue includes HSR revenue and also HSR bus revenue. Source: Steer (2024)

Figure 7.5: Distribution of ridership for 2030 Valley to Valley produced by the Monte Carlo simulation



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.6: Distribution of revenue for 2030 Valley to Valley produced by the Monte Carlo simulation



Notes: Revenue includes HSR revenue and HSR BUS revenue. The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

2040

7.14 The following results are based directly on model results for the year 2040.

2040 Phase 1

7.15 The results of the Monte Carlo simulation for 2040 Phase 1 ridership and revenue are summarized in Table 7.7.

Table 7.7: Summary of Monte Carlo simulation for 2040 Phase 1

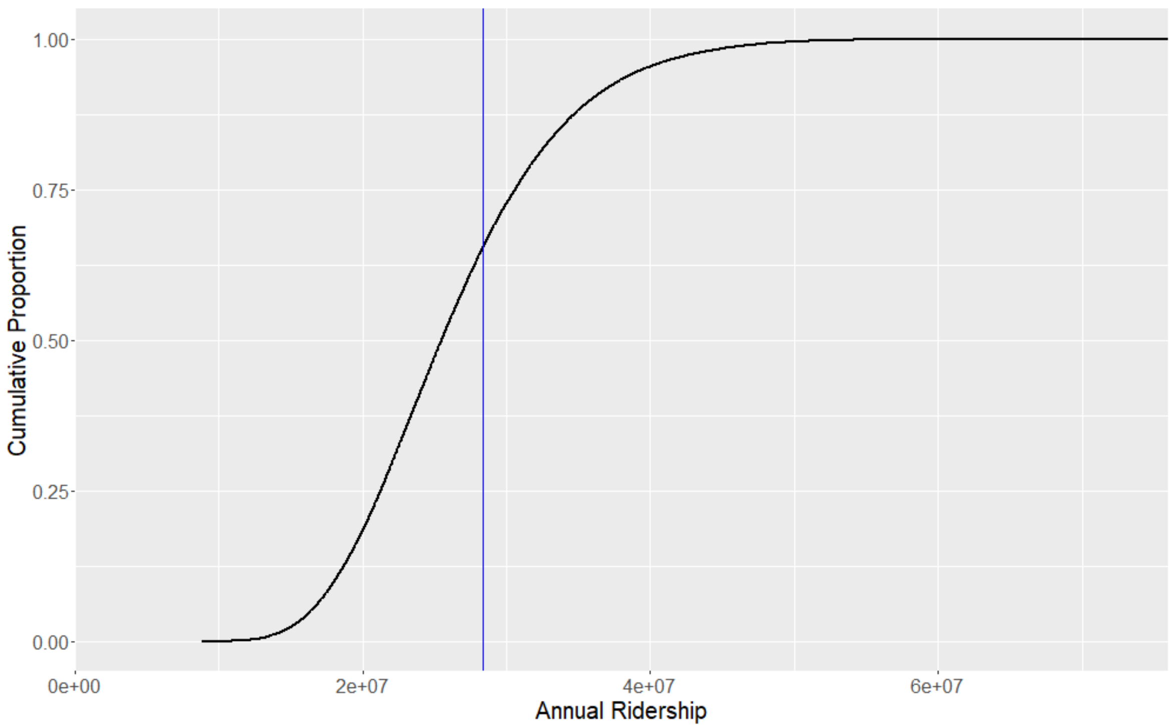
Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
Base case	28.39	3,576.00	1.00	1.00
p0	8.85	971.88	0.31	0.27
p1	13.76	1,648.03	0.48	0.46
p10	18.05	2,202.19	0.64	0.62
p25	21.27	2,621.22	0.75	0.73
p50	25.50	3,183.18	0.90	0.89
p75	30.61	3,869.60	1.08	1.08
p90	35.94	4,595.18	1.27	1.29

Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
p99	47.08	6,147.74	1.66	1.72
p100	75.98	10,298.30	2.68	2.88

Source: Steer (2024)

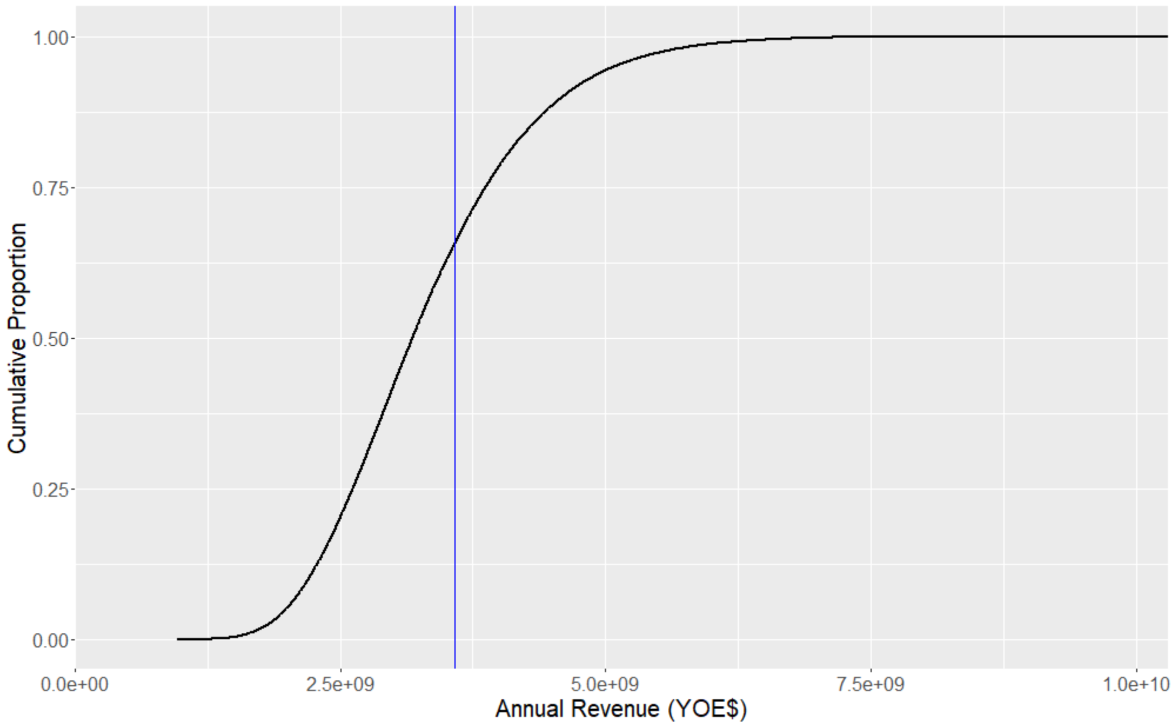
7.16 It can be observed that the p10 and p90 values for 2040 Phase 1 ridership are 64% and 127% of the corresponding base case value. Furthermore, the p10 and p90 values for 2040 Phase 1 revenue are 62% and 129% of the corresponding base case value.

Figure 7.7: Distribution of ridership for 2040 Phase 1 produced by the Monte Carlo simulation



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.8: Distribution of revenue for 2040 Phase 1 produced by the Monte Carlo simulation (YOE$)



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

2040 Valley to Valley

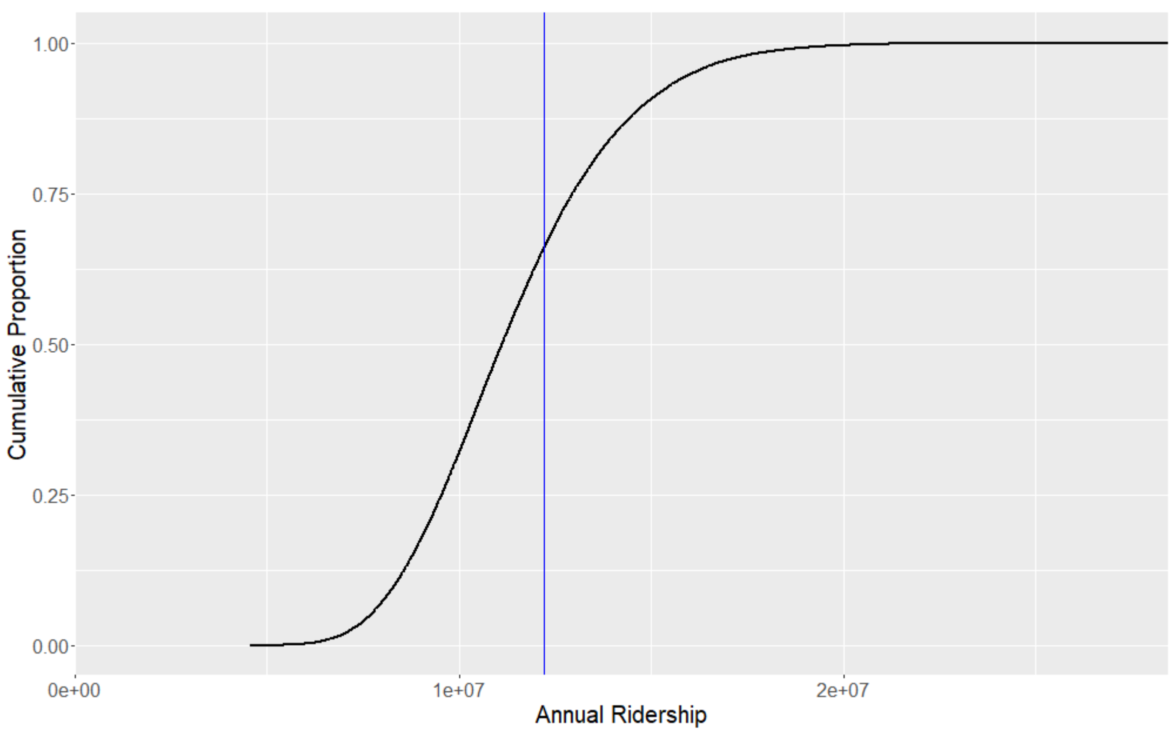
7.17 The results of the Monte Carlo simulation for 2040 Valley to Valley ridership and revenue are summarized in Table 7.8.

Table 7.8: Summary of Monte Carlo simulation for 2040 Valley to Valley

Percentile or reference point	Ridership (in million)	Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
Base case	12.22	1,206.43	1.00	1.00
p0	4.56	384.55	0.37	0.32
p1	6.63	623.89	0.54	0.52
p10	8.33	804.46	0.68	0.67
p25	9.55	934.44	0.78	0.77
p50	11.12	1,106.40	0.91	0.92
p75	12.96	1,310.26	1.06	1.09
p90	14.86	1,524.98	1.22	1.26
p99	18.66	1,961.96	1.53	1.63
p100	28.45	3,131.18	2.33	2.60

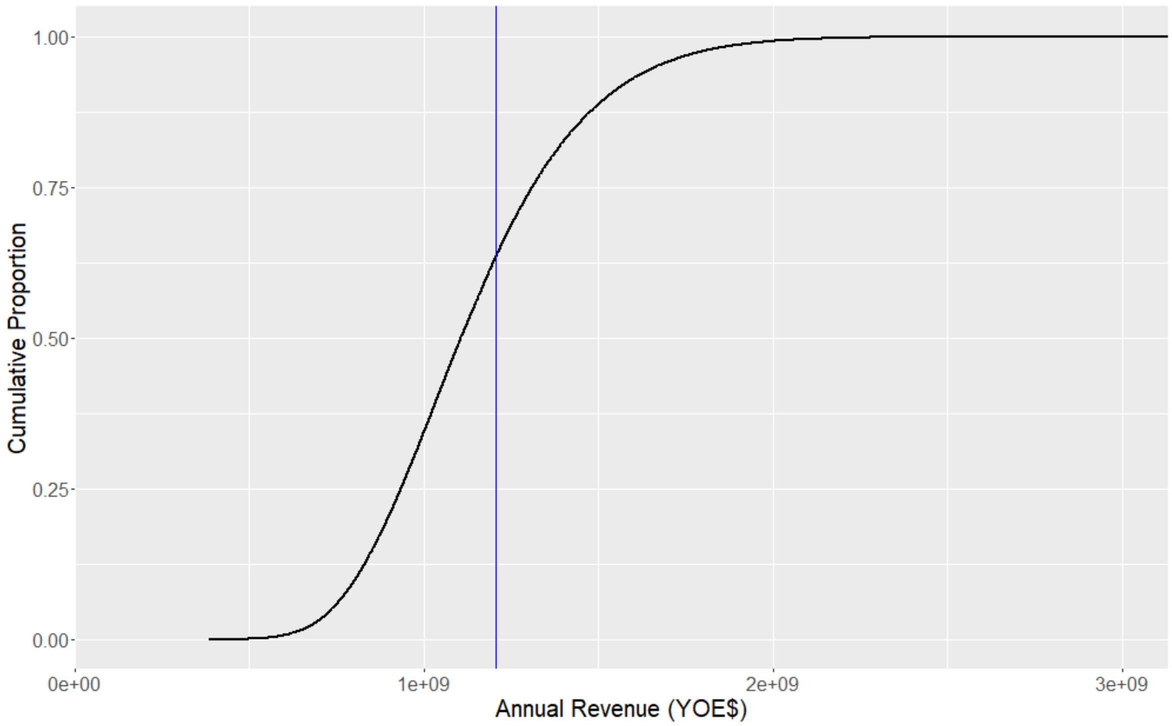
Notes: Revenue includes HSR revenue and also HSR bus revenue. Source: Steer

Figure 7.9: Distribution of ridership for 2040 Valley to Valley produced by the Monte Carlo simulation



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.10: Distribution of revenue for 2040 Valley to Valley produced by the Monte Carlo simulation



Notes: Revenue includes HSR revenue and HSR BUS revenue. The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

2050

7.18 Ridership and revenue input data for the 2050 meta model was based on the data collected with the 2040 CRRM model, scaled with growth factors as described in 6.10 above. However, the Monte Carlo simulation itself was done with an independent set of 100,000 draws from the random variable distributions described above in Table 7.1 and Table 7.2, for each scenario. Therefore, the results of the Monte Carlo simulation may present some differences even in relative terms compared to the corresponding results for the other years.

2050 Phase 1

7.19 The results of the Monte Carlo simulation for 2050 Phase 1 ridership and revenue are summarized in Table 7.9.

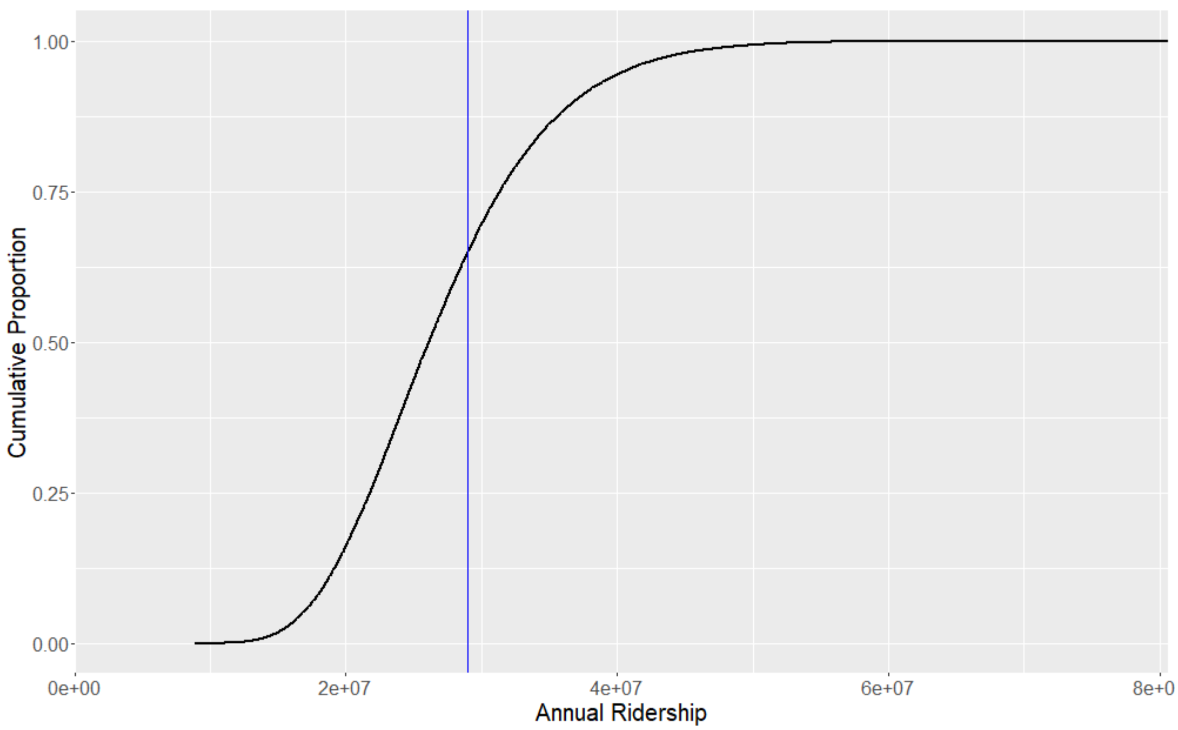
Table 7.9: Summary of Monte Carlo simulation for 2050 Phase 1

Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
base run	29.01	4,443.93	1.00	1.00
p0	8.84	1,245.75	0.30	0.28
p1	14.11	2,060.37	0.49	0.46
p10	18.53	2,745.04	0.64	0.62

Percentile or reference point	Annual Ridership (million)	Annual Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
p25	21.79	3,260.21	0.75	0.73
p50	26.13	3,961.20	0.90	0.89
p75	31.36	4,819.84	1.08	1.08
p90	36.85	5,737.34	1.27	1.29
p99	48.35	7,617.76	1.67	1.71
p100	80.63	12,250.17	2.78	2.76

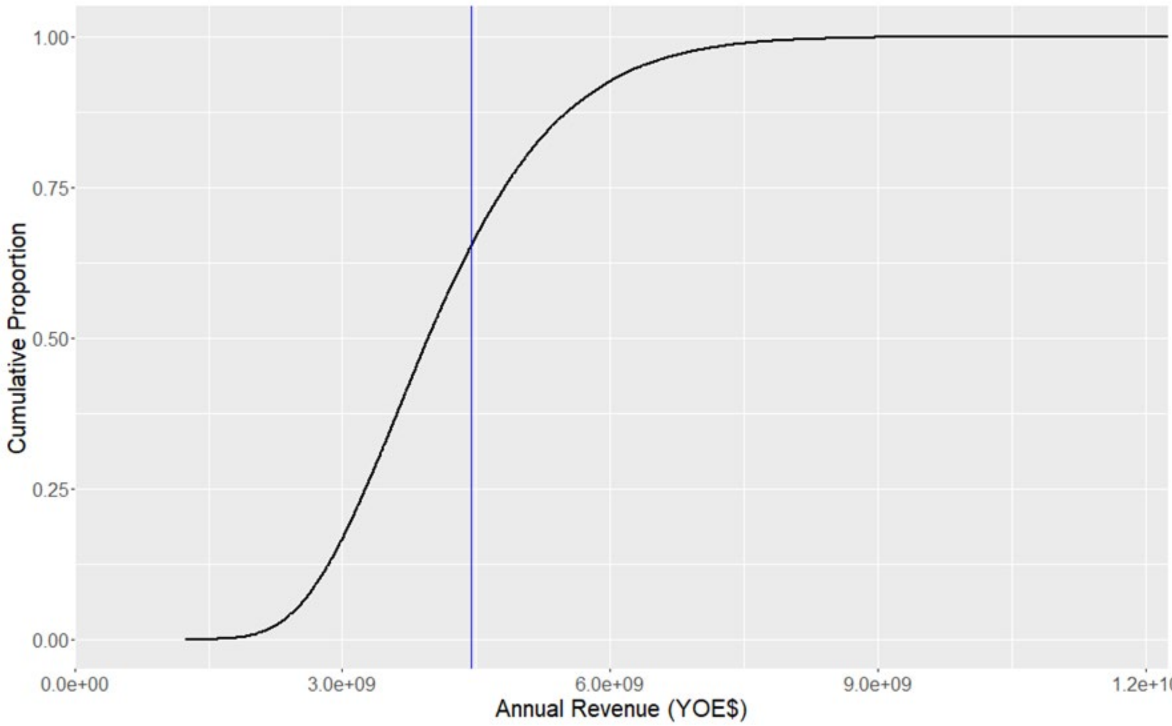
Source: Steer (2024)

Figure 7.11: Distribution of ridership for 2050 Phase 1 produced by the Monte Carlo simulation.



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.12: Distribution of revenue for 2050 Phase 1 produced by the Monte Carlo simulation.



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

2050 Valley to Valley

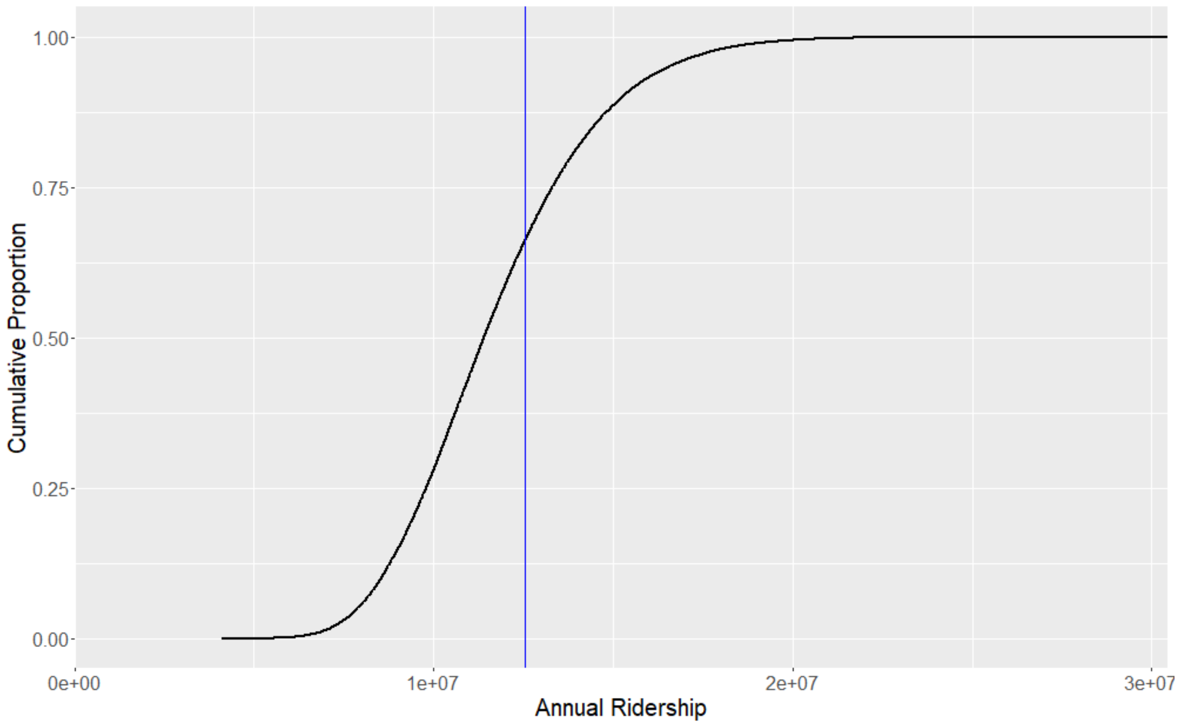
7.20 The results of the Monte Carlo simulation for 2050 Valley to Valley ridership and revenue are summarized in Table 7.10.

Table 7.10: Summary of Monte Carlo simulation for 2050 Valley to Valley

Percentile or reference point	Ridership (in million)	Revenue (YOE$ million)	Ridership index (WRT base value)	Revenue index (WRT base value)
Base case	12.54	1,501.42	1.00	1.00
p0	4.10	499.28	0.33	0.33
p1	6.82	777.00	0.54	0.52
p10	8.54	999.13	0.68	0.67
p25	9.79	1,164.23	0.78	0.78
p50	11.40	1,378.77	0.91	0.92
p75	13.31	1,632.73	1.06	1.09
p90	15.25	1,897.94	1.22	1.26
p99	19.16	2,440.63	1.53	1.63
p100	30.45	4,141.62	2.43	2.76

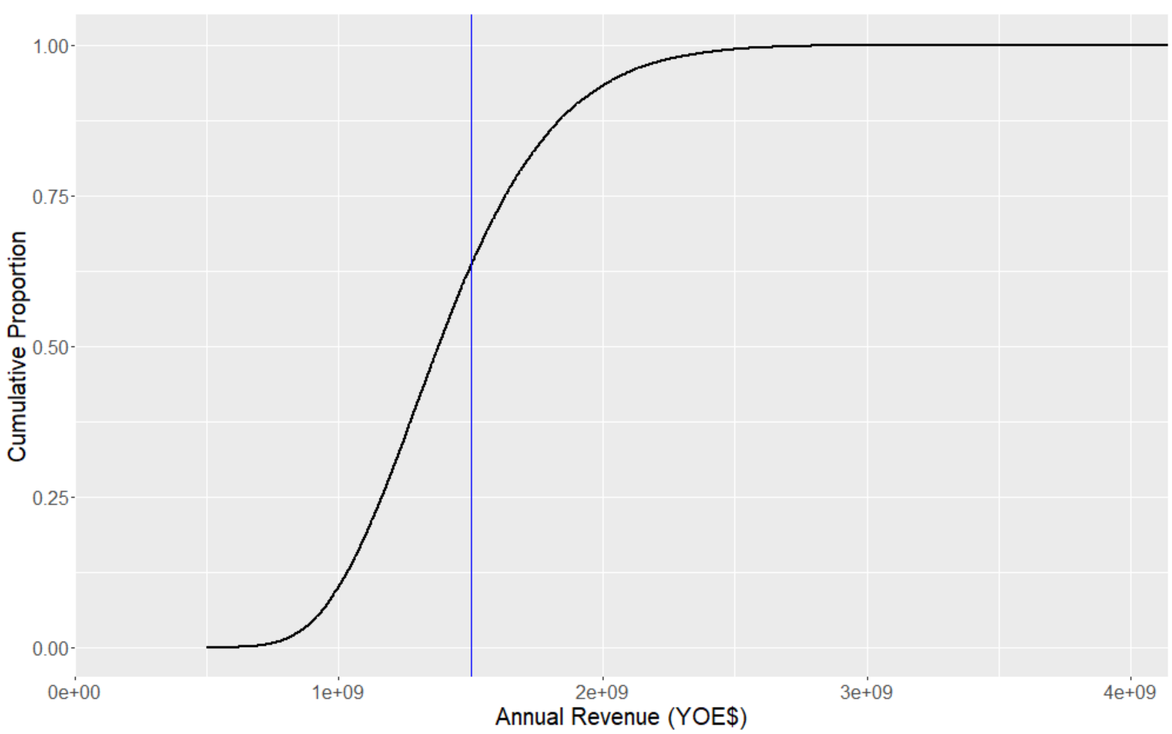
Notes: Revenue includes HSR revenue and also HSR bus revenue. Source: Steer (2024)

Figure 7.13: Distribution of ridership for 2050 Valley to Valley produced by the Monte Carlo simulation.



Note: The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Figure 7.14: Distribution of revenue for 2050 Valley to Valley produced by the Monte Carlo simulation.



Notes: Revenue includes HSR revenue and HSR BUS revenue. The blue vertical line on the graph indicates the corresponding base case value. Source: Steer (2024)

Discussion

7.21 Table 7.11 offers a comparison of the latest results for Phase 1 with the comparable results from the 2022 Steer risk analysis and the 2020 Cambridge Systematics analysis that were presented earlier in Table 2.5.

Table 7.11: Comparison of Monte Carlo simulation results for Phase 1

Percentile or reference point	CS ridership index (WRT base value)	CS revenue index (WRT base value)	Steer 2022 ridership index (WRT base value)	Steer 2022 revenue index (WRT base value)	Steer 2024 ridership index (WRT base value)	Steer 2024 revenue index (WRT base value)
base run	1.00	1.00	1.00	1.00	1.00	1.00
p0	0.24	0.30	0.44	0.41	0.30	0.26
p1	0.41	0.47	0.63	0.61	0.49	0.46
p10	0.59	0.67	0.77	0.76	0.64	0.62
p25	0.76	0.85	0.86	0.86	0.75	0.74
p50	1.00	1.10	0.98	1.00	0.90	0.89
p75	1.29	1.40	1.13	1.15	1.08	1.08
p90	1.60	1.70	1.27	1.31	1.27	1.29
p99	2.13	2.23	1.54	1.6	1.67	1.72
p100	3.08	3.12	2.24	2.28	2.71	2.82

Notes: The Cambridge Systematics results are for 2033 and the Steer results are for 2030. Sources: “California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis” (Cambridge Systematics, Inc., 2020, pp. 50, 53), “CA Ridership Modeling” (Steer, 2022a, p. 199).

7.22 Overall, the latest results look reasonable. The median values being lower than the base case values imply that it is slightly more likely that ridership and revenue will turn out lower than the base case than higher. This apparent discrepancy is healthy considering the nature of risk analysis and it is also consistent with the results of ex-post reviews of forecast demand versus actual demand, which show that more often than not actual demand turns out lower than predicted – see Figure 2.2 and ‘Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review’ (Nicolaisen & Driscoll, 2014).

7.23 The spreads of the latest Monte Carlo simulation results for ridership and revenue are broadly similar at the low end but more conservative at the high end than the earlier results for the 2020 Business Plan risk analysis (Cambridge Systematics, Inc., 2020).

7.24 Compared to the 2022 Steer risk analysis the latest results indicate somewhat lower expectations for ridership and revenue across most of the distribution except the extreme high end where these latest results have higher upper bounds. This difference was to be expected because as part of the methodology review that preceded this latest analysis, we identified variables that were not very significant to be removed from the sensitivity analysis and added 2 new variables (the transfer penalty and the access-egress nest parameter) that were expected to have more of an impact – an expectation which the results of the meta modelling showed to be correct.

7.25 The median values for ridership and revenue in the latest results are approximately 10 percentage points lower than the corresponding base case values. This is due to a few variables that were modeled asymmetrically in the sensitivity analysis because the likely risks in these cases were one-sided. The air service frequency factor (varied from 0.4 to 1.0) can be expected to increase HSR demand, and the access-egress nest parameter factor (varied from 0.2 to 1.0) can be expected to reduce it. The net effect of these two factors is driven by the nest parameter factor which has a greater impact on the results – see Figure 7.1 and Figure 7.2. This tends to slightly reduce HSR ridership and revenue compared to the base case, also shifting the curves towards lower values of ridership and revenue compared to the results of the earlier sensitivity analysis exercises.

7.26 Does this imply that the base case assumption for the access-egress nest parameter should be revised in future? It seems worth considering. The sensitivity analysis is mainly about future risks, which are related to potential future divergence between the situation modeled in the calibration scenario and the future situations which may arise. However, the interpretation of the access-egress nest parameter is not as simple as, for instance, the value of time, and it is not a parameter associated with a specific variable that could be updated to reflect expected changes between the future scenario and the base scenario. The access-egress nest parameter was not estimated as part of the choice model, it was an assumption based on consideration of values obtained in other studies and the CRRM model was calibrated taking into account the assumed value. The calibration process and the pivoting mechanism used to match observed county-county flows by main mode mean that whichever value is assumed for the access-egress nest parameter, the CRRM model will fit the base data reasonably well, but different values may have different implications for modelling future HSR demand. Therefore, the risk associated with this parameter is related to the potential divergence between the assumed value and the ideal value that would best represent actual behavior in the future scenarios.

7.27 The access-egress nest parameter represents the relative similarity of the different combinations of access-egress options for a given main mode, where the smaller the value is the more they are perceived as being similar, and where the larger the value is the more they are perceived as being distinct. At one extreme, a value of 0 for this parameter would imply that all access-egress combinations for a given main mode are perceived as bunched together as if they were a single travel option, not distinct options. At the other extreme, a value of 1 for the access-egress nest parameter would imply that each access-egress combination for a given main mode would be perceived as if it were a completely distinct modal option. Therefore, the lower values of the access-egress nest parameter will tend to attract less demand and higher values will tend to attract more demand.

7.28 If the access-egress nest parameter used in the model for a future update of the base case were revised downwards, the model would probably need to be re-calibrated, and in any case there is a pivot mechanism to ensure that in the base case the predicted demand by main mode should match the corresponding observed values at the county level (Steer, 2023). The re-calibration and pivoting to match observed values would mean that the results for the calibration scenario might not change significantly, but the use of a lower value for this parameter (for instance 0.25) could imply changes to the HSR forecasts and would have the advantage of making it more reasonable to model related risks with a symmetrical distribution centered on the base case value. It is something that seems worth considering for future model updates.

8 References

Albalate, D., Bel, G., & Fageda, X. (2015). Competition and cooperation between high-speed rail and air transportation services in Europe. *Journal of Transport Geography*, *42*, 166–174. https://doi.org/10.1016/j.jtrangeo.2014.07.003

Cambridge Systematics, Inc. (2018). *California High-Speed Rail 2018 Business Plan—Ridership and Revenue Risk Analysis*. https://hsr.ca.gov/wp-content/uploads/docs/about/business_plans/2018_CA_High_Speed_Rail_Business_Plan_Ridership_and_Revenue_Risk_Analysis.pdf

Cambridge Systematics, Inc. (2020). *California High-Speed Rail 2020 Business Plan—Ridership and Revenue Risk Analysis* (p. 96). https://hsr.ca.gov/wp-content/uploads/docs/about/business_plans/2020_Business_Plan_Ridership_and_Revenue_Risk_Analysis.pdf

Carnell, R. (2021a). *Package 'lhs'* [Computer software]. https://cran.r-project.org/web/packages/lhs/lhs.pdf

Carnell, R. (2021b, September 7). *Basic Latin hypercube samples and designs with package lhs*. https://cran.r-project.org/web/packages/lhs/vignettes/lhs_basics.html

EIA. (2007). *Annual Energy Outlook 2007* (p. 243). EIA.

EIA. (2022, September 30). *U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)*. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=emm_epm0_pte_nus_dpg&f=a

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2005). How (in) accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association*, *71*(2), 131–146.

Givoni, M., & Dobruszkes, F. (2013). A Review of Ex-Post Evidence for Mode Substitution and Induced Demand Following the Introduction of High-Speed Rail. *Transport Reviews*, *33*(6), 720–742. https://doi.org/10.1080/01441647.2013.853707

Harb, M., Stathopoulos, A., Shiftan, Y., & Walker, J. L. (2021). What do we (Not) know about our future with automated vehicles? *Transportation Research Part C: Emerging Technologies*, *123*, 102948. https://doi.org/10.1016/j.trc.2020.102948

Hartley, C., Perry, M., & Rogers, L. (2021). *A Preliminary Analysis of U.S. and State-Level Results From the 2020 Census* (p. 8). https://www.census.gov/content/dam/Census/library/working-papers/2021/demo/pop-twps0104.pdf

Hoque, J. M., Erhardt, G. D., Schmitt, D., Chen, M., Chaudhary, A., Wachs, M., & Souleyrette, R. R. (2022). The changing accuracy of traffic forecasts. *Transportation*, *49*(2), 445–466. https://doi.org/10.1007/s11116-021-10182-8

Hoque, J. M., Erhardt, G. D., Schmitt, D., Chen, M., & Wachs, M. (2021). Estimating the uncertainty of traffic forecasts from their historical accuracy. *Transportation Research Part A: Policy and Practice*, *147*, 339–349. https://doi.org/10.1016/j.tra.2021.03.015

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab-an S4 package for kernel methods in R. *Journal of Statistical Software*, *11*, 1–20.

Kolarova, V., Steck, F., Cyganski, R., & Trommer, S. (2018). Estimation of the value of time for automated driving using revealed and stated preference methods. *Transportation Research Procedia*, *31*, 35–46. https://doi.org/10.1016/j.trpro.2018.09.044

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26.

McGuckin, N. (2018). *Can We Use the NHTS to Estimate Long-Distance Travel?* https://doi.org/10.13140/RG.2.2.23563.13607

Mitrani, A. (2023). *Casa* (0.0.0.9002) [R]. Steer. https://github.com/SteerGroup/casa

Nicolaisen, M. S., & Driscoll, P. A. (2014). Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review. *Transport Reviews*, *34*(4), 540–557. https://doi.org/10.1080/01441647.2014.926428

Pouillot, R., Delignette-Muller, M. L., Kelly, D. L., & Denis, J. B. (2023). The mc2d package. *R Package*, *30*. https://cloud.r-project.org/web/packages/mc2d/vignettes/docmcEnglish.pdf

Pouillot, R., Delignette-Muller, M.-L., Kelly, D. L., & Denis, J.-B. (2021). *The mc2d package*. https://cran.r-project.org/web/packages/mc2d/vignettes/docmcEnglish.pdf

R Core Team. (2023). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2024). *Documentation for package ‘stats’ version 4.4.0*. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

Rashidi, T. H., Waller, T., & Axhausen, K. (2020). Reduced value of time for autonomous vehicle users: Myth or reality? *Transport Policy*, *95*, 30–36. https://doi.org/10.1016/j.tranpol.2020.06.003

Smith, S. K., & Tayman, J. (2003). An Evaluation of Population Projections by Age. *Demography*, *40*(4), 17. https://i-mail.bebr.ufl.edu/sites/default/files/Research%20Reports/2003%20Demog%20%28Age%20Proj%29.pdf

Special Projects Office. (1958). *Program Evaluation Research Task (PERT): Summary Report Phase 1*. Bureau of Naval Weapons, Department of the Navy. https://web.archive.org/web/20151112203807/http://www.dtic.mil/dtic/tr/fulltext/u2/735902.pdf

Steer. (2022a). *CA Ridership Modeling*.

Steer. (2022b). *Sensitivity Analysis*. Steer.

Steer. (2023). *Ridership and Revenue Forecasting Inputs to the 2023 Project Update Report* (23454508). Steer.

Train, K. (2003). *Discrete choice methods with simulation*. http://eml.berkeley.edu/books/train1201.pdf

U.S. Bureau of Labor Statistics. (2020, March 16). *Labor Force Projections Evaluation*. https://www.bls.gov/emp/evaluations/2008-2018-labor-force.htm

U.S. Census Bureau. (2005, September). *State Population Projections 2004-2030 Results*. CDC WONDER Online Database. https://wonder.cdc.gov/controller/datarequest/D7

Vose, D. (2008). *Risk Analysis: A Quantitative Guide*. John Wiley & Sons. https://books.google.cl/books?id=9CaoAqaRcVwC&lpg=PP1&pg=PA407#v=onepage&q&f=false

Wan, Y., Ha, H.-K., Yoshida, Y., & Zhang, A. (2016). Airlines’ reaction to high-speed rail entries: Empirical study of the Northeast Asian market. *Transportation Research Part A: Policy and Practice*, *94*, 532–557. https://doi.org/10.1016/j.tra.2016.10.014

Wardman, M., & Ibáñez, N. J. (2012). The congestion multiplier: Variations in motorists’ valuations of travel time with traffic conditions. *Transportation Research Part A: Policy and Practice*, *46*(1), 213–225. https://doi.org/10.1016/j.tra.2011.06.011

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Zeileis, A., & Hothorn, T. (2002). *Diagnostic checking in regression relationships* (Vol. 2). na. https://journal.r-project.org/articles/RN-2002-018/RN-2002-018.pdf

Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, *95*, 1–36.

Zhang, A., Wan, Y., & Yang, H. (2019). Impacts of high-speed rail on airlines, airports and regional economies: A survey of recent research. *Transport Policy*, *81*, A1–A19. https://doi.org/10.1016/j.tranpol.2019.06.010

Control Information

Prepared by	Prepared for
Steer 800 Wilshire Blvd, Suite 1320, Los Angeles, CA 90017 United States +1 (213) 425 0990 www.steergroup.com	DB E.C.O. North America, Inc. 555 Capitol Mall Suite #1250, Sacramento, CA 95814, United States https://db-eco.us/en/
Steer project/proposal number	**Client contract/project number**
23454511	
Author/originator	**Reviewer/approver**
Alex Mitrani	Masroor Hasan
Other contributors	**Distribution**
Shahriar Ahmed Tanu Bansal	Client: Steer:
Version control/issue number	**Date**
1 10	February 2024

Complex questions.
Powerful answers.

Explore **steergroup.com** or visit one of our 21 offices:

Bogotá, Colombia	Los Angeles, USA	Rome, Italy
Bologna, Italy	Manchester, UK	San Juan, Puerto Rico
Boston, USA	Mexico City, Mexico	Santiago, Chile
Brussels, Belgium	Milan, Italy	São Paulo, Brazil
Leeds, UK	New Delhi, India	Toronto, Canada
Lima, Peru	New York, USA	Vancouver, Canada

steer