Implementing Uncertainty Analysis in Toll Facility Traffic and Revenue Forecasts

By

Yong Zhao, Ph.D., P.E., AICP
Senior Transportation Planner
URS Corporation
9400 Amberglen Boulevard
Austin, TX 78729
Email: yong_zhao@urscorp.com
Phone: 512-217-8360

And

Huimin Zhao, Ph.D., P.E., AICP
Independent Transportation Consultant
10808 Enclave Vista Cave
Austin, TX 78730
Email: hzhao88@gmail.com
Phone: 512-273-5188

Submitted for Presentation at
The 3rd Conference on Innovations in Travel Modeling (ITM2010)

Abstract

As a unique practice in travel demand modeling, traffic and revenue forecasting for toll facilities require highly accurate traffic projection. The questionable point estimations of traffic volume may mislead long-term facility plan and investments. Explicit and rigorous statistical recognition of uncertainty in traffic demand modeling will reduce the risk in planning and designing the toll facilities and provide a proper direction of future development. This paper discusses the methodology and implementations to adopt risk analysis into forecasting process. The result representation is also illustrated. Empirical examples from North Tarrant Expressway Manage Lane project in Ft Worth, Texas are presented.

Key Words: Risk Analysis, Traffic and Toll Revenue Forecasts, Monte Carlo Simulation

Introduction

Uncertainty in travel demand forecasting has been long recognized (e.g., Alonso, 1968), but has been recently received some attention (Krishnamurthy and Kockelman, 2003; Pravinvongvuth, Chootinan, and Chen, 2003; Rodier and Johnston, 2002, Zhao and Kockelman, 2002). As Mahmassani (1984) pointed out, the difficulty to quantify the uncertainty is that not all sources of uncertainty are suitable (and available) for empirical studies. Although most Metropolitan Planning Organizations (MPOs) have not adopted standard uncertainty analysis procedures in their current modeling process, they feel the importance and emergency of this issue[1].

Toll facilities are particularly vulnerable to uncertainty in the transportation industry since they are related to high initial costs and uncertain revenue returns, especially the often low returns in the early years. Toll facility investments closely depend on traffic and toll revenue projections. Therefore, one of the fundamental aspects of the traffic forecasts is to recognize a wide range of possible outcomes in order to explore the implication for the revenues. For this reason, handful hypothetical scenarios can merely suggest an undefined level of risk. A full range of risk analysis should be undertaken to identify the probability of the revenues at a particular confidence level to support financial evaluation. Additionally, explicit and careful recognition of uncertainty will also reduce the risk in planning, designing, and financing the toll facilities.

One financial institute's report (Standard and Poor, 2004) compares first year forecast traffic versus actual traffic volumes for 82 projects and shows the traffic predictions averaged about 76% of their predicted values, with a standard deviation of 0.26. It will be wise that the traffic forecasts can provide and evaluate such variations before the numbers being realized. This paper discusses the methodology to adopt risk analysis into forecasting process and illustrates the steps to implement the procedure and present the results in a simple, straight-forward way. An empirical example is presented from a traffic and revenue study for North Tarrant Expressway Manage Lane project in Ft Worth, Texas.

Methodology

"Uncertainty" in this study refers to the statistical variations and co-variations of the point estimates of variables in the conventional transportation models. Statistical uncertainty is widely adopted in risk analysis for business evaluation and management. Risk analysis usually employs simulation technique because one can simulate risk and uncertainty from a variety of sources simultaneously and impose correlation across inputs.

There are many factors that can cause the uncertain in future revenue streams. For example, economic growth and land use development, recession and inflation, gas price,

---

[1] See the Travel Model Improvement Program Newsletters, *TMIP Connection*, spring and summer issues of 2005.

transit and ITS development, telecommunication, and safety issues. However, based on findings from previous studies (e.g., Zhao and Kockelman, 2002; Rodier and Johnston, 2002), the following factors, as primary variables in traffic demand models are considered in this study:

- Population and employment, directly reflecting economic growth and land use development
- Value of time (VOT), implying household income and traveler's toll road preference

These factors can be treated as random variables and regarded as the source of statistical uncertainty in transportation and toll models. Their statistical uncertainty is adopted in this risk analysis. To track the stochastic errors, Monte Carlo simulation and sensitivity analysis are the primary tools to be used. However, large number runs of simulation in travel demand forecasting are costly and impractical. In this study, a modified simulation method was developed to obtain a valid estimation of the revenue ranges within reasonable computation efforts.

These uncertainties are simulated by first specifying their distributions and then randomly generating values from these distributions. To impose sign constraints on many of these variables (for example, population of a zone cannot be less than the existing level), lognormal distributions are generally used.

Toll revenue forecast output is the major focus of this work, and its variability is due solely to input and parameter uncertainties. Based on results' statistical moments (i.e., means and standard deviations), appropriate probability or confident intervals can be assigned to the final results. The modeling procedure includes the following steps:

<u>Step 1: Examine the distribution and uncertainty of studying factors</u>

First, demographic variations from historical data are collected and each variable's variation was examined using best-fit distribution. Most MPOs utilize land use models to develop the demographic forecasts on population and employment over a long period. It is worthwhile to collect these historical forecasts and compare them with the real Census data to find how the forecast variation exists among different time intervals and geographic areas. The comparison is not to validate the current land use forecasting procedures; rather, it is to determine the level of uncertainty associated with these forecasts. Once the level of uncertainty is revealed, the associated distributions are constructed for simulation. These distributions were then validated by comparing simulation samples against the historical data. Some reasonableness thresholds have been imposed during the validation. For example, household density has a maximum limit in suburban area. This maximum limit was developed using existing data. Finally, demographic forecast ranges were produced using Monte Carlo simulation.

Second, the distribution of VOT is directly derived from the state preference (SP) survey identifying the travelers' preference toward the toll facilities. Usually a multinomial or mixed logit model is developed to evaluate traveler's preference of toll route. In the case of mixed logit model, the ranges of value of time can be produced using Monte Carlo simulation.

Step 2: Reveal the relationship between traffic revenue forecasts and inputs

Representative demographic and VOT points from the forecast range (mean, lower and upper limit) are selected and the respective traffic and revenue points are calculated through the full travel demand model runs. Based on the representative points, traffic revenue and input relationships are established using a linear or nonlinear regression model.

Step 3: Developed the toll traffic forecast distribution

Traffic and revenue distributions are estimated using the regression models developed in Step 2 with Monte Carlo simulation. The risk probabilities for traffic forecast points based on input variable distributions are evaluated. Possible revenue growth trends are also analyzed.

In this study, the random numbers for the study variables were generated using Excel with an add-on, @risk (Palisade Corporation).

Empirical Results

The case study presented in this paper is based on the preliminary findings of a Level 3 Traffic and Toll Revenue study for proposed managed lanes (MLs) along the North Tarrant Express (NTE) corridor in Northeast Dallas-Fort Worth area. This study is to develop a 50-year annual toll revenue forecasts and support a private developer's proposal to enter the comprehensive development agreement (CDA) with Texas Department of Transportation (TxDOT). The project assumes a 2012 opening year and focuses on Segment 1 of IH 820 between IH 35W and IH 820 East Loop, which covers a distance of approximately 6.4 miles, as shown in Figure 1.

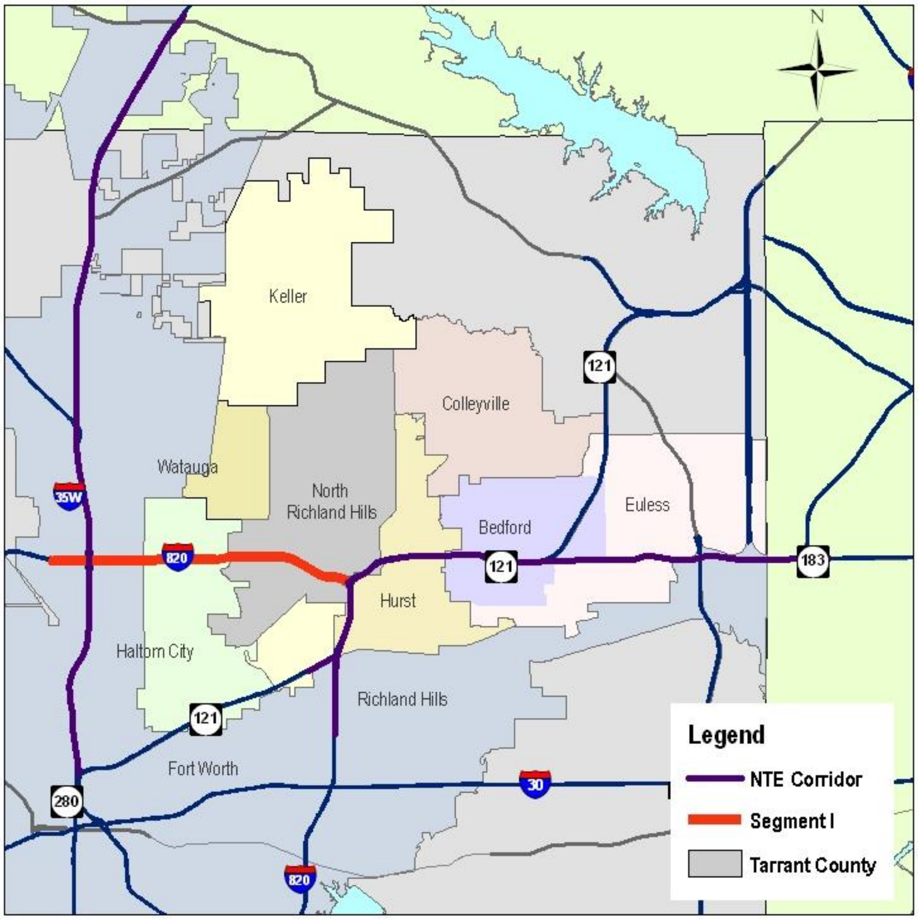

Figure 1. Segment I of the NTE Managed Lanes Corridor

Two demographic variables were examined in this study: population and employment, which reflect the economic growth and land-use development. To develop the uncertainty level associated with these factors, historical North Central Texas Council of Governments (NCTCOG) data were collected and compared. NCTCOG has developed and updated its Mobility Plans every two to three years since 1985. The horizon years were 2000, 2010, 2020, and 2025. In each plan, there were land-use forecasts for year 2000. A comparison of 2000 land use forecasts against the actual land use data from the census and other surveys were essential to establish the variation of demographic forecasts in the DFW area. The NCTCOG social-economic data includes the following variables for each traffic analysis planning zone (TAP).

- Household
- Population
- Basic Employment
- Retail Employment
- Service Employment

Because population is correlated with household, it was only used as a reference in this study.

There are a total of 4,874 TAP zones in the DFW area. Because of the size of TAP zones, the magnitudes of these variables are quite different. Therefore, the percentage variations were modeled as an indirect measure of the variables' uncertainties. The TAP zones are categorized into five area types:

- Central Business District,
- Outer Business Districts,
- Urban Residential,
- Suburban Residential, and
- Rural.

There are at least 200 zones for each area type, so at least 200 observed variations for each area type were collected. To develop the demographic forecast uncertainties, the variance (percentage difference) between the previous projections of the variable and the census data was calculated. For each area type, the distribution of the variance was drawn and the best-fit distribution was found.

The majority area type of the zones in the NTE corridor study area is 4 and 5. Figures 2 and 3 illustrated the household and total (basic, retail, and service) employment distributions for area type 4 (suburban residential). These distributions were the best-fit in illustrating the variation of the historical projections for year 2000. That is, by comparing the observed census data, the previous forecasts can overestimate or underestimate the demographic variables for a specific zone. The total variations yield a range for the demographic variables. In general, the historical projections should contain ranges with possibility distributions, which are demonstrated in Figures 2 and 3.

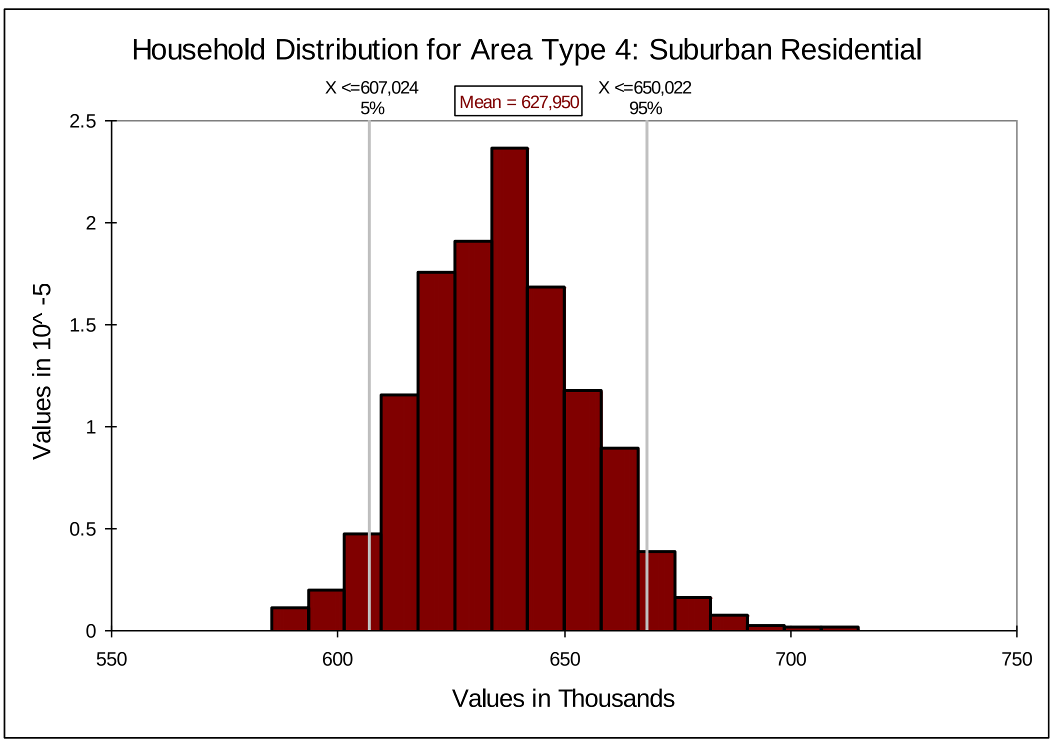

Figure 2. Household Distribution for Area Type 4: Suburban Residential

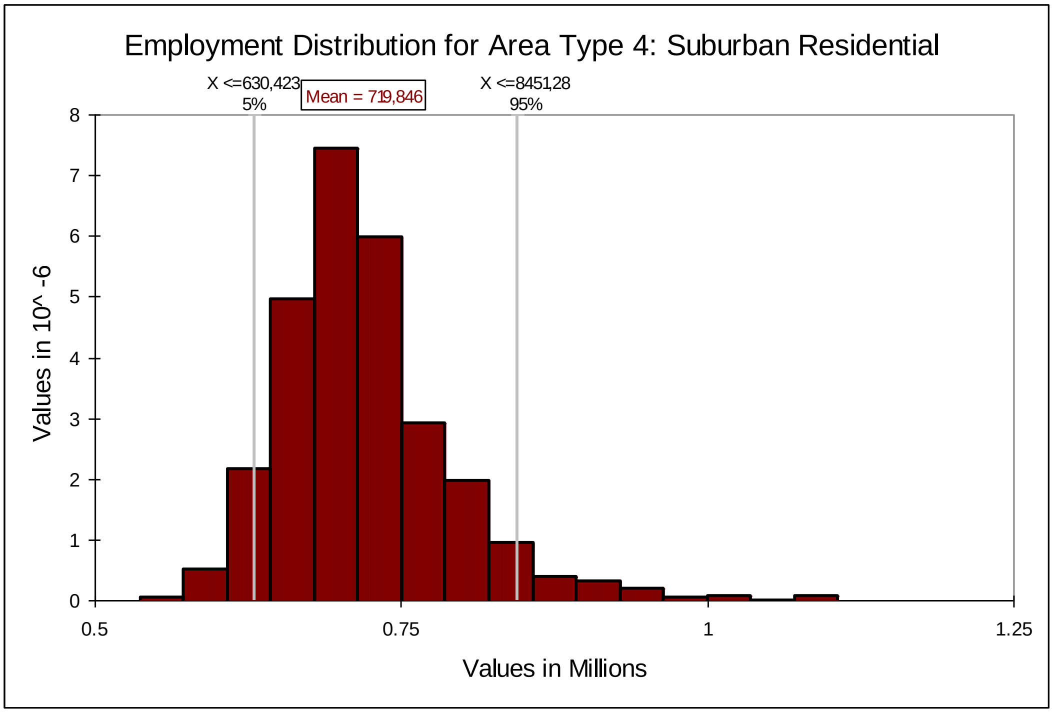

Figure 3. Employment Distributions for Area Type 4: Suburban Residential

The VOT distributions were developed directly from the mixed logit model from the SP survey conducted in 2005, as shown in Figure 4. The average peak period and off-peak VOTs are $10.77 and $9.60, respectively.

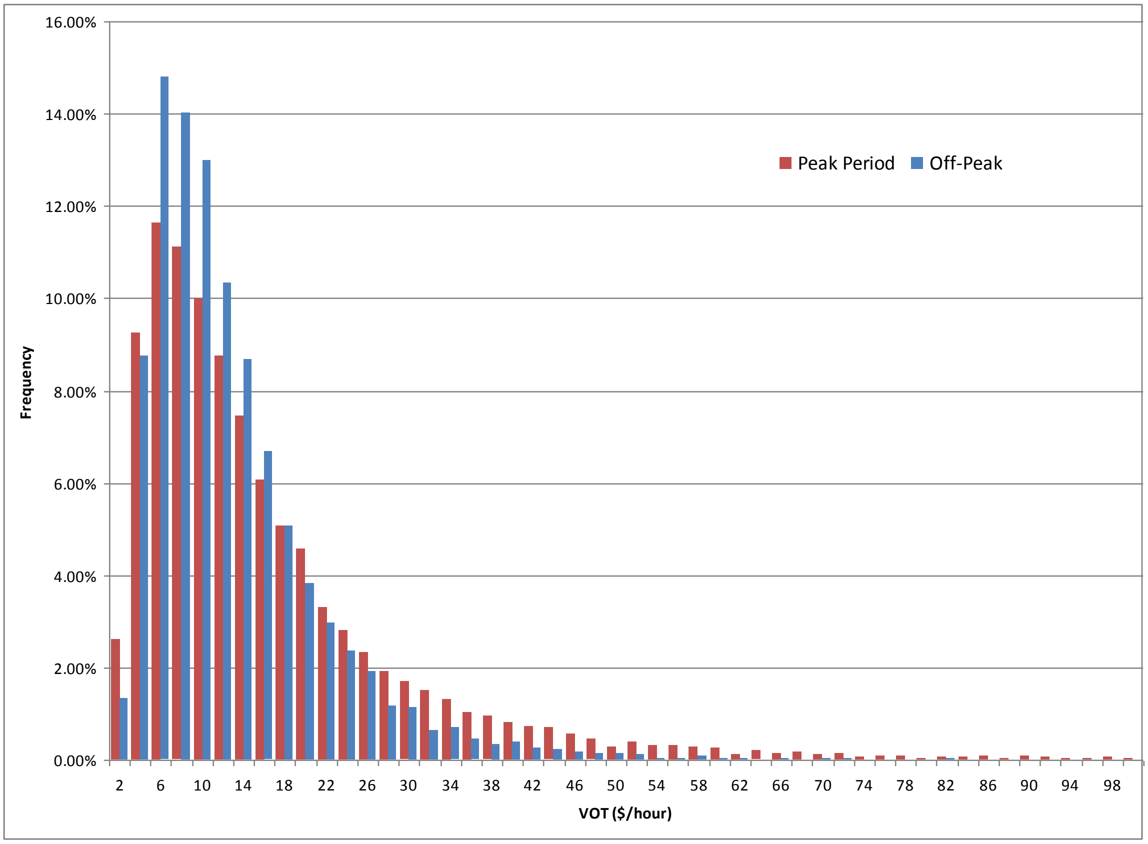

Figure 4. VOT Distributions

Once the distributions for the key variables had been developed, the statistical characteristics of them were evaluated. The mean case was viewed as the most likely case. The 5% and 95% probability points represent the lower bound and the upper bound of the whole range and can be considered as the "worst" and the "best" cases respectively. Basically, there were three cases each for household, total employment, and value of time. Table 1 shows the input variables' ranges for year 2012 and 2030.

Table 1. Input Variations (Base Case Input as 100)

Year 2012

Variable	Lower Bound (5%)	Mean	Upper Bound (95%)
Household	91	103	111
Total Employment	89	102	115
Value of time	91	100	108

Year 2030

Variable	Lower Bound (5%)	Mean	Upper Bound (95%)
Household	80	103	117
Total Employment	79	110	121
Value of time	90	100	110

All of the representative scenarios were then examined to avoid illogical results. For example, the population of a TAZ is less likely to be reduced to less than the existing level. On the other hand, the land use density for residential and employments should not exceed the current maximum. Also, the ratio between household and employment at the regional level should maintain a reasonable range.

Given the representative scenarios for years 2012 and 2030, the full travel demand model and revenue calculations were performed. The input and output data points were sufficient to develop linear regression models to describe the relationship between the demographic and value of time inputs and the revenue outputs. Figure 5 show the relationships.

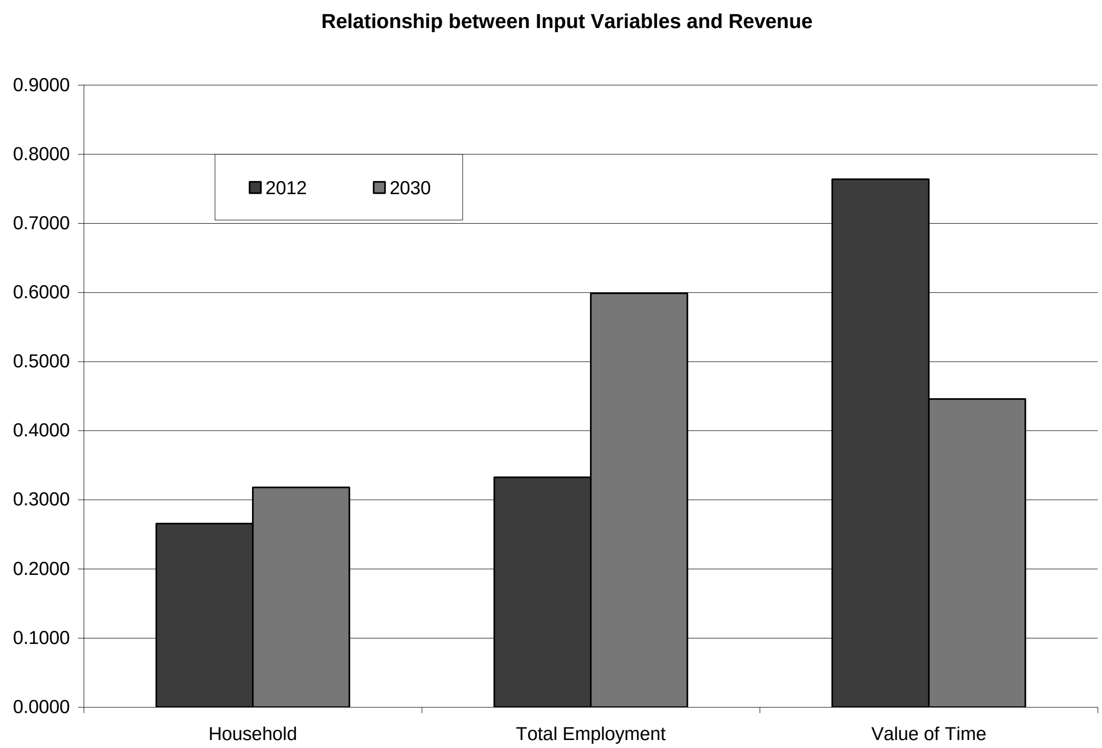

Figure 5. Relationship between Input Variables and Revenue

Using the input distributions and the regression relationships developed, the final revenue ranges for forecasting years can be estimated via Monte Carlo simulation. In this study, 10,000 runs have been conducted for each forecasting year. By drawing from their distributions based on frequency, the confident intervals of revenue estimations were calculated. These intervals were compared against the current revenue results for validation.

The toll revenue developed in using the mean input variable was used as the base cases for years 2012 and 2030. Using the base cases revenues as 100, the simulation results are summarized in Table 2.

Table 2. Revenue Uncertainties (Base Case as 100)

Year	Lower Bound (5%)	Mean	Upper Bound (95%)
2012	77	100	121
2030	59	94	136

The interpretation of the results should be undertaken with care. The lower bound revenue of 2012 suggests that there is only a 5% possibility that the 2012 revenue will be less than 77% of the base case forecast, or there is 95% possibility that the 2012 revenue will be higher than 77% of the base case forecast.

The cumulative probability distribution for year 2012 is illustrated in Figure 6, where the X-axis is the possible revenue in terms of the ratio to the year 2012 base case revenue and the Y-axis is the probability of the revenue estimation larger or equal to the ratio. For example, the 80% probability line suggests that there is an 80% chance that the actual revenues in 2012 will be at least 88% of the base case forecasts. The similar cumulative probability for year 2030 is shown in Figure 7.

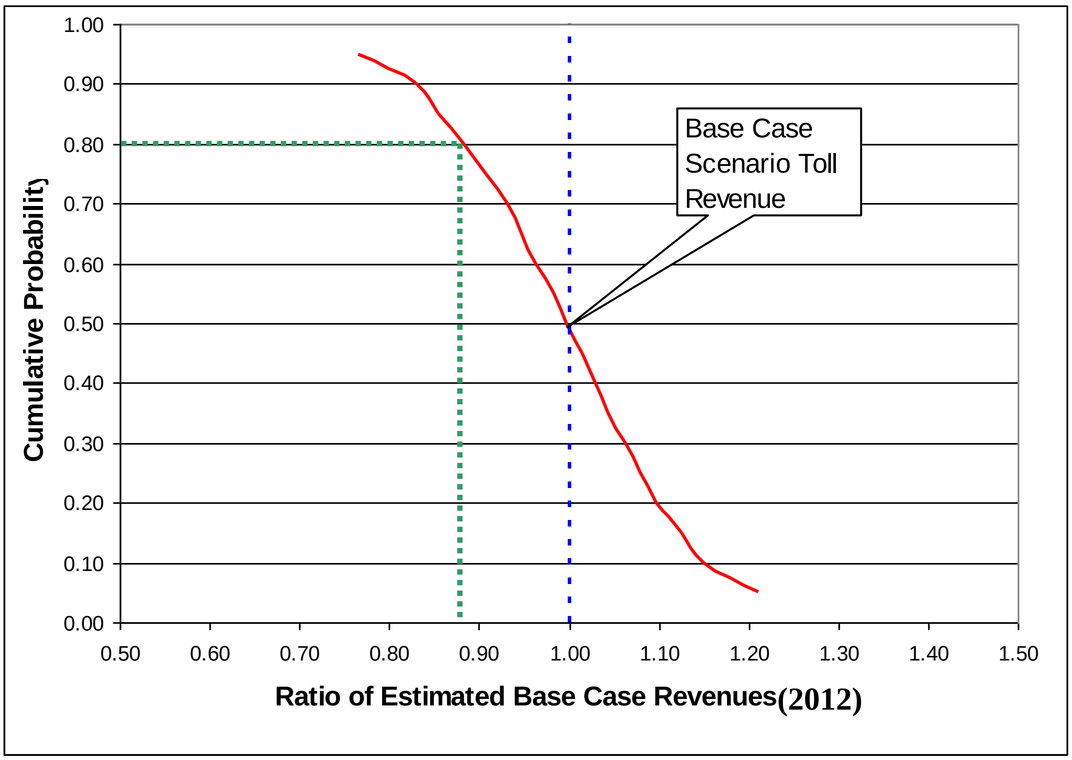

Figure 6. 2012 Revenue Ranges with Cumulative Probability

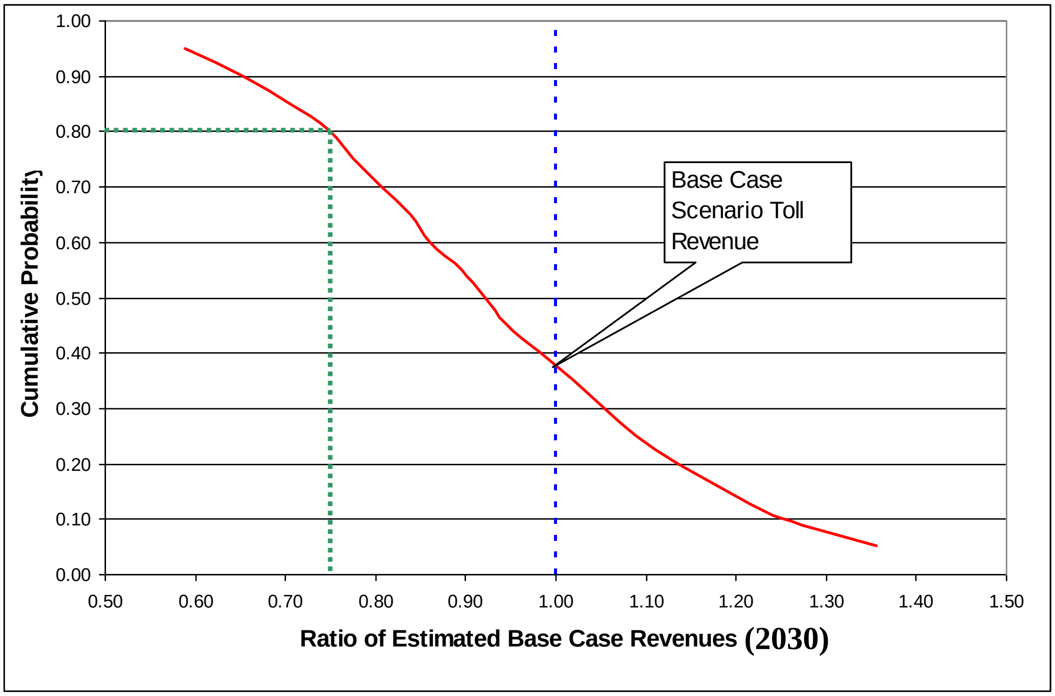

Figure 7. 2030 Revenue Ranges with Cumulative Probability

Conclusions

This study proposed a practical method in synthesizing the uncertainties in travel demand model input data. Using these input data profiles and the relationship to output data will allow travel demand models to appreciate the future uncertainties and provide the base of evaluating the risk of major investments. A case study and the result representation are also illustrated.

Although transportation investments and privatizations require the proper allocation of risk among project stakeholders, traffic risk cannot easily be managed by most stakeholders. Practically, traffic demand is very difficult to predict, as it primarily focuses on the performance of the economy, the decision making process of the users, and the interaction and competition of transportation means. As this study provides an initial step towards the risk recognizing procedure, more detailed analysis is required for future research and practices.

REFERENCES

Alonso, W. 1968. Predicting Best with Imperfect Data. *Journal of the American Institute of Planners*. pp. 248-255.

Krishnamurthy, S. and K. M. Kockelman. 2003. Propagation of Uncertainty in Transportation-Land Use Models: An Investigation of DRAM-EMPAL and UTPP Predictions in Austin, Texas. *Transportation Research Record* No.1831, pp. 219-229.

Mahmassani, H. S. 1984. Uncertainty in Transportation Systems Evaluation: Issues and Approaches, *Transportation Planning and Technology,* pp. 1-12.

Pravinvongvuth, S., P. Chootinan, and A. Chen. 2003. Simulating Variability and Uncertainty of a Combined Travel Demand Model, *Journal of Eastern Asia Society for Transportation Studies*, pp.1324-1339.

Rodier, C. J., and R. A. Johnston. 2002. Uncertain socioeconomic travel demand model and emissions models: could plausible nonconformity? *Transportation Research A*, pp.613-631.

Standards and Poor. 2004. Traffic Forecasting Risk: Study Update 2004.

Zhao, Y., and K. M. Kockelman. 2002. The Propagation of Uncertainty Through Travel Demand Models: An Exploratory Analysis, *Annals of Regional Science* 36, pp. 145-163.