Evaluating Parameter Uncertainty in Transportation Demand Models

Natalie Mae Gray

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Gregory S. Macfarlane, Chair
Grant G. Schultz
Daniel P. Ames

Department of Civil and Construction Engineering

Brigham Young University

Evaluating Parameter Uncertainty in Transportation Demand Models

Natalie Mae Gray
Department of Civil and Construction Engineering
Master of Science

BYU Engineering

Abstract

NEED The abstract of a thesis should describe the motivation, objective, overall results, and central findings of the thesis. It may have multiple paragraphs if necessary.

Keywords: sensitivity analysis, transportation demand model, transportation planning, latin hypercube sampling, monte carlo simulation

Acknowledgments

NEED Students should acknowledge funding sources. They may also use the acknowledgment page to express appreciation for the committee members, friends or family who provided assistance in research, writing or technical aspects of the dissertation, thesis or selected project. Acknowledgements should be simple and in good taste.

Table of Contents

List of Figures

List of Tables

1 Introduction

2 Literature Review

2.1 Overview

Researchers acknowledge that there exists uncertainty in transportation demand models, though few choose to quantify it. This review looks at the types of uncertainty that exists, and research that has been done to evaluate uncertainty. Then, this review considers methods for value sampling, which has been used in research to create a range of input values or parameters.

2.2 Types of Uncertainty

Uncertainty generally exists from two basic sources: input uncertainty and model uncertainty (Rasouli & Timmermans, 2012). Input uncertainty includes behavioral and socioeconomic data – for instance, the number of jobs and residents in a zone might be coded incorrectly – as well as parameter estimates including automobile operating costs, gasoline costs, and values of time. Model uncertainty can be further divided into specification error and estimation error. Specification error is a failure to estimate the "true" model. Estimation error, on the other hand, is a failure to estimate the correct values of model constants and parameters, or taking the mean estimated value as a single parameter.

2.2.1 Input Uncertainty

2.2.2 Model Uncertainty

2.3 Sampling Methods

There are two popular methods of value sampling, Monte Carlo simulation and Latin hypercube sampling. Monte Carlo simulation draws independently from multiple distributions, while Latin hypercube sampling makes draws that cover the parameter space more efficiently and can capture the joint distribution between two or more parameter values. As a result, Latin hypercube sampling can reduce the number of draws needed to fully re-create the statistical variance in a model, but the amount of reduction is unknown and may not be universal to all problems (Yang et al., 2013).

2.4 Summary

In a four-step travel demand model, most error research to this point has focused on input uncertainty, rather than model uncertainty (Rasouli & Timmermans, 2012). For this study, the estimation error within the model uncertainty is of the most immediate concern.

3 Model Design and Methodology

3.1 Overview

3.2 Transportation Demand Model

To examine the effects of parameter input sensitivity, we developed a trip-based travel model with four steps:

1. trip generation,
2. trip distribution,
3. mode choice, and
4. destination choice.

Trip generation, the first step, was conducted using socioeconomic (SE) data and household trip productions from Roanoke Valley Transportation Planning Organization RVTPO. The trip productions were summarized by household sizes, vehicles, and workers, and the weighted mean of each trip purpose was taken. The three trip purposes used are Home Based Work (HBW), Home Based Other (HBO), and Non-Home Based (NHB). Trip attraction was skipped for this analysis.

The second step, trip distribution, used distance and travel time skims from RVTPO. The skims were simplified to use auto, nonmotorized, and transit modes. Travel time for auto used the single occupancy vehicle peak time, nonmotorized travel time used the distance skim multiplied by a factor of average walking speed (3 mph), and transit time used the walk to bus peak time.

Mode choice, the third step, calculates utilities for the three modes. These utilities were exponentiated, added together, and the natural log was taken to get a logsum value for every origin and destination pair. The utility equations for the mode choice model are as follows:

$$\text{drive\_utility} = (\text{coeff\_ivtt} * \text{auto}) + (\text{coeff\_cost} * \text{auto\_cost} * \text{DIST}) \tag{3.1}$$

$$\text{nonmo\_utility} = (\text{k\_nmot} + 20 * (\text{coeff\_walk1} * \text{nonmotor})) \tag{3.2}$$

$$\text{trans\_utility} = \text{k\_trn} + (\text{coeff\_ivtt} * \text{transit}) \tag{3.3}$$

The mode choice parameters (constants and coefficients) were obtained from the USTM Resiliency Model. These values are shown in Table 3.1 and Table 3.2.

Table 3.1: Mode Choice Coefficients

Name	HBW	HBO	NHB
CIVTT	-0.0450	-0.0350	-0.0400
CCOST	-0.0016	-0.0016	-0.0016
CWALK1	-0.0900	-0.0700	-0.0800
AUTOCOST	18.3000	18.3000	18.3000

Table 3.2: Mode Choice Constants

Name	HBW	HBO	NHB
K_TRN	-0.5140	-0.9853	-1.3020
K_NMOT	1.7602	0.5448	-0.5359

Table 3.3: Destination Choice Parameters

VAR	HBW	HBO	NHB
HH	0.000000	1.018700	0.2077000
I(OTH_EMP + OFF_EMP)	0.000000	0.806400	0.5626000
OFF_EMP	0.458600	0.403200	0.2816000
OTH_EMP	1.682700	0.403200	0.2816000
RET_EMP	0.608700	3.813800	5.1189000
DISTCAP	70.000000	40.000000	40.0000000
CLSUM	1.000000	1.000000	1.0000000
CDIST	-0.080100	-0.172800	-0.1157000
CDISTSQ	0.002600	0.003400	0.0035000
CDISTCUB	-0.000009	-0.000011	-0.0000133

The final step, destination choice, uses the mode choice logsum as the primary impedance, and a size term calculated using zonal employment. The destination choice utility is the impedance term added to the log of the size term. The destination choice parameters (constants and coefficients) were also obtained from the USTM Resiliency Model. These values are shown in Table 3.3.

The four-step model gives data in terms of utilities and probabilities. This can be turned into production and attraction trips, and then a highway assignment can be applied using CUBE.

3.3 Parameter Sampling

With this four-step model, MC and LHS methods were used to determine the possible combinations of parameter variance. To identify a standard deviation for each parameter, a coefficient of variation was used. A set coefficient of variation of 0.10 was used for all six mode choice input parameters, and four of the destination choice parameters (HH, OFF_EMP, OTH_EMP, and RET_EMP). Literature had identified a coefficient of variation of 0.30, but for this analysis that caused an unrealistic value of time, and thus it was changed to be 0.10 (Zhao & Kockelman, 2002). The standard deviation was equal to 0.10 multiplied by the mean, where the mean values in this situation are the base scenario parameters (as identified in Table 3.1, Table 3.2, and Table 3.2).

The MC random sampling uses the R function of `rnorm`. LHS uses the `lhs` package in R. Since this package only chooses variables on a zero to one scale, the values given use a function to put the random sampling on the right scale needed for the given parameter. The full code for both methods can be found in a public GitHub repository. 100 and 600 draws of random samples for both methods are generated. With these generated parameters, the mode choice model step was run for every set of input parameters for each purpose. The mean logsum value for each run was determined to compare each continuous draw. This allowed us to see how many iterations of which sampling type would be sufficient to show a full range of possible outcomes.

3.4 Summary

A standard four-step model was created in R to create trips, and evaluate sampling methodologies.

4 Sensitivity Analysis Results

4.1 Overview

4.2 Sampling Findings

The parameters generated were compared for both sampling methods. Figure 4.1 shows the distributions for the HBW parameters when using 100 draws, and Figure 4.2 shows how that changes when using 600 draws. These distributions show that LHS gives normally distributed parameters with fewer draws than MC sampling. At 100 draws LHS shows a nearly perfect normal distribution, where there are some discrepancies for the MC generated parameters. Without looking at the mode choice results, these Figures show that LHS is likely to estimate the full variance of the results with much fewer draws.

To determine if LHS is effective at a reasonable amount of iterations, the standard deviation was calculated for each additional draw. This value shows how much the mean mode choice logsum value for all zones can vary. When the standard deviation for the draws stabilizes, that shows that the amount of generated parameters has captured all of the possible variances of the results. This can be visualized for each purpose. The HBW results for the cumulative standard deviation are shown in 4.3. The results for the other two purposes (HBO and NHB) are in 4.4 and 4.5 respectively.

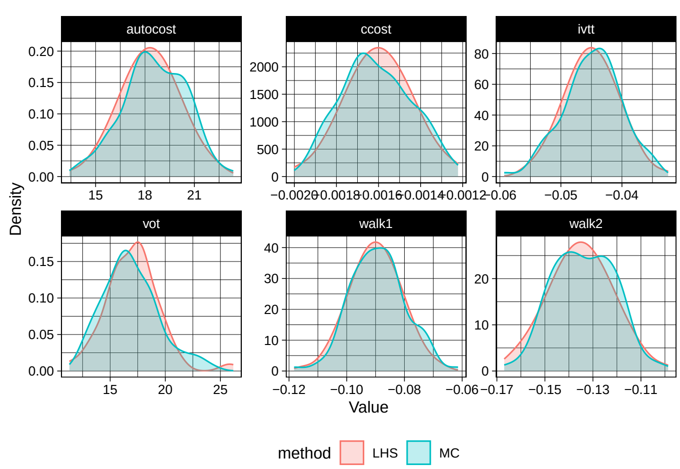

Figure 4.1: HBW Distributions for Input Parameters with 100 Draws

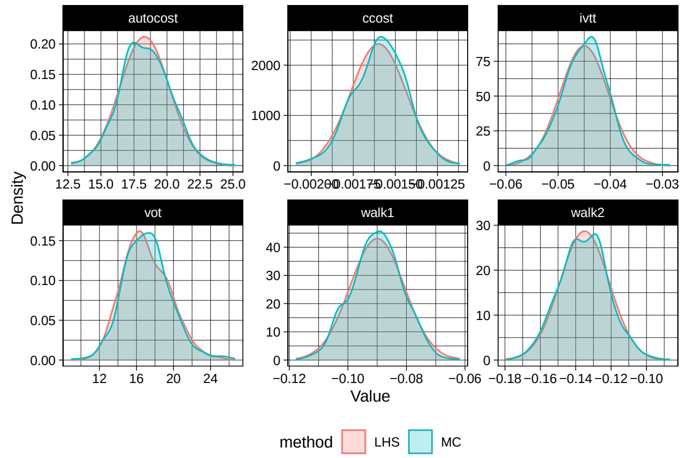

Figure 4.2: HBW Distributions for Input Parameters with 600 Draws

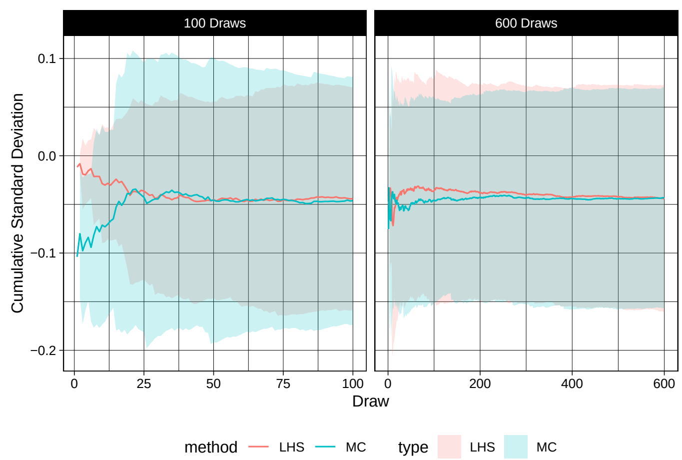

Figure 4.3: HBW Mean Logsum Standard Variation with 100 and 600 Draws

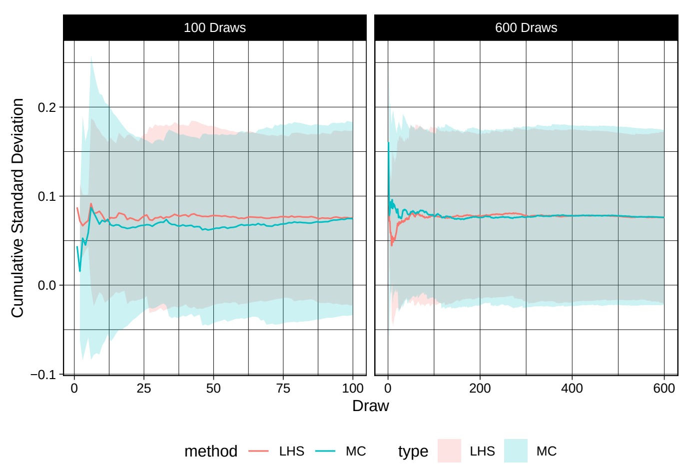

Figure 4.4: HBO Mean Logsum Standard Variation with 100 and 600 Draws

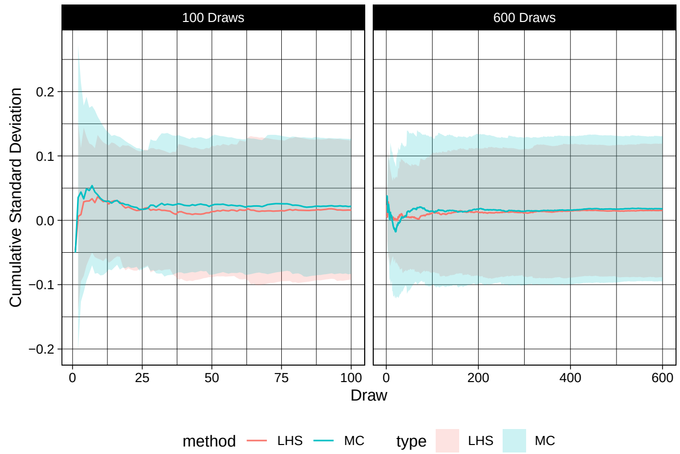

Figure 4.5: NHB Mean Logsum Standard Variation with 100 and 600 Draws

For all three trip purposes, the LHS method had its standard deviation stabilized between 100 and 200 draws. The MC method had still not stabilized to the same extent after 600 draws. This shows us that Latin Hypercube Sampling greatly decreases the iterations needed to approximate random sampling methods. Since LHS captures the possible variance at a small enough amount of iterations, it can be used for large transportation demand models.

4.3 Roanoke Valley Results

A hundred iterations of Latin Hypercube sampling was created with the RVPTO data. Some ways this data can be analyzed to evaluate the range of results is with the highway assignment applied after the four-step model. The coefficient of variation among all iterations for every link in the transportation network can be calculated and compared. Figure 4.6 shows, by facility type, how the total volume varies. It can be seen that the higher volume roads, have less variability, and the roads with low volume have more variability.

Variation among a link can also be visualized with a density plot of the total volume across all iterations. In Figure 4.7 there are three links total volume shown each with different facility types. The green dashed lines are the links measured AAWDT value, and the red dashed lines are the base scenario total volume value. These plots show that across each facility type with an AAWDT value, the variance in total volume is within the range of uncertainty already expected within the model. Since the AAWDT is so different from the Total Volume value, the uncertainty lies more within the model, than within the change in parameter values.

Figure 4.8 displays both standard deviation, and the corresponding average volume amoung iterations on each link. This shows that there is a higher standard deviation on roads with high volumes, but a change in 600 vehicles on a road with 40,000 daily volume is minuscule. That is

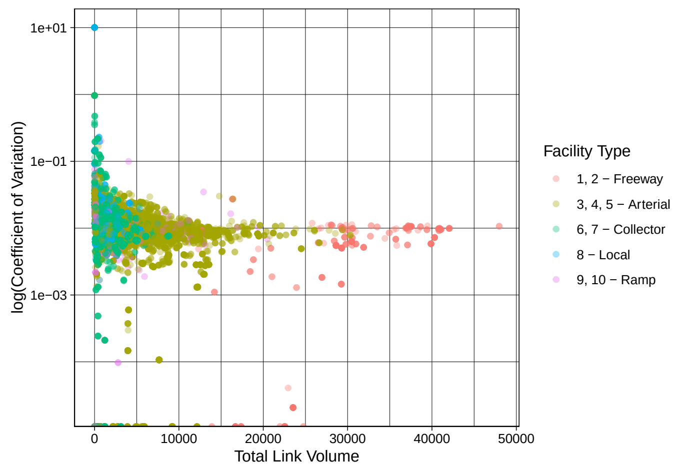

Figure 4.6: Coefficient of Variation of Total Link Volume

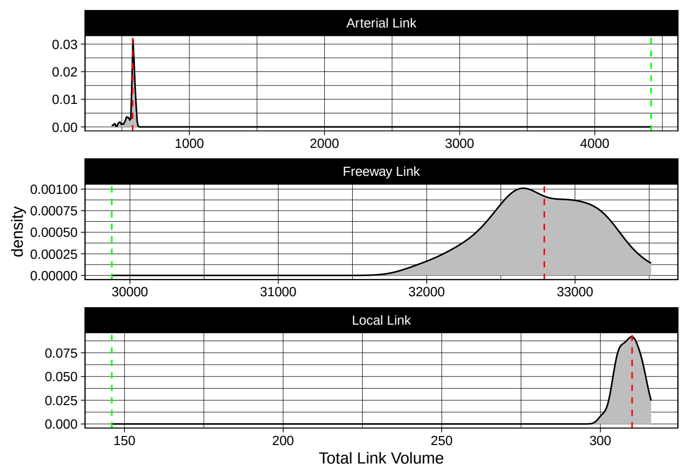

Figure 4.7: Total Volume Density for an Individual Link by Facility Type

only a 1% variation in volume which is insignificant.

Another way to visualize changes among each iteration is to look at how mode choices change. Figure 4.9 shows that for home based work trips by each origin, the variance among each mode is less than 20%, and that transit varies the most.

4.4 Summary

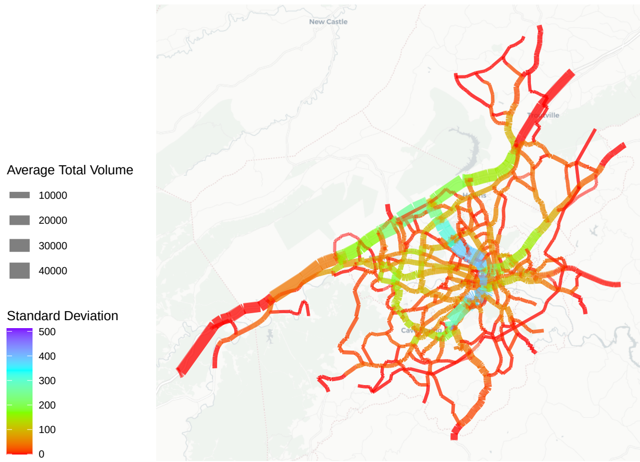

Figure 4.8: Standard Deviation of Total Volume

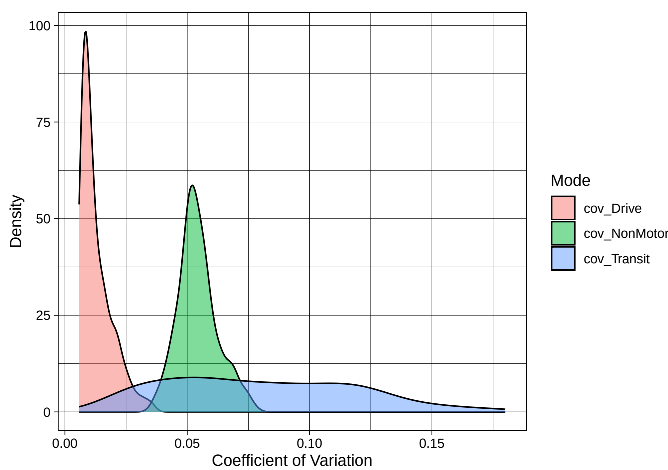

Figure 4.9: Origin Desnity for Coefficient of Variation by Mode for HBW Trips

5 Conclusions

References

Rasouli, S., & Timmermans, H. (2012). Uncertainty in travel demand forecasting models: Literature review and research agenda. *Transportation Letters*, *4*(1), 55–73.

Yang, C., Chen, A., Xu, X., & Wong, S. (2013). Sensitivity-based uncertainty analysis of a combined travel demand model. *Transportation Research Part B: Methodological*, *57*, 225–244.

Zhao, Y., & Kockelman, K. M. (2002). The propagation of uncertainty through travel demand models: An exploratory analysis. *The Annals of Regional Science*, *36*(1), 145–163.