ABSTRACT

A COMPARISON BETWEEN REVEALED PREFERENCE (RP) AND STATED PREFERENCE (SP) BASED ON RESULTS OF SIMULATIONS

by
Qiuzi Chen

This research quantifies the potential biases resulted from two different data generation methods used in transportation modeling: Revealed Preference (RP) and Stated Preference (SP) techniques. Revealed Preference technique is the conventional approach to generate data. It relies on observed or reported data of actual behavior. Stated preference technique is a new data generation method. It creates transportation scenarios using hypothetical data. Conventional studies favor the use of revealed reference. However, full description of advantages and weaknesses of using revealed preference technique is not available in literature, neither point-to-point comparison between stated preference and revealed preference techniques. This research contributes to the literature by demonstrating the relative magnitude of biases inherent to both approaches.

The method to explore approach-specific to generate data is simulation. The simulation work concentrates on biases found in RP and SP in the statistical estimation component, although biases also exist in the forecasting component. Simulation in RP case focuses on errors-in-variables; while, simulation in SP case concentrates on the internal design of the data matrix.

Based on the results from the simulations, the research points out the potential biases in two models used to forecast model shift behavior in New Jersey. The work concludes with a list of future research needs.

A COMPARISON BETWEEN REVEALED PREFERENCE (RP) AND STATED PREFERENCE (SP) BASED ON RESULTS OF SIMULATIONS

by
Qiuzi Chen

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirement for the Degree of
Master of Science in Transportation

Committee for the Interdisciplinary Program in Transportation

January 1995

APPROVAL PAGE

A COMPARISON BETWEEN REVEALED PREFERENCE (RP) AND STATED PREFERENCE (SP) BASED ON RESULTS OF SIMULATIONS

QIUZI CHEN

Dr. W. Patrick Beaton, Thesis Advisor — Date
Professor, Department of Social Science
and Policy Studies, Institute for Transportation, NJIT

Dr. Kyriacos Mouskos, Committee Member — Date
Assistant Professor, Civil and Environmental Engineering, NJIT

Dr. Naomi Rotter, Committee Member — Date
Professor, School of Industrial Management, NJIT

BIOGRAPHICAL SKETCH

Author: Qiuzi Chen

Degree: Master of Science in Transportation

Date: January 1995

Undergraduate and Graduate Education:

- Master of Science in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 1995

- Bachelor of Arts in Tourism
  Nan Kai University, Tianjin, P.R.China, 1992

Major: Transportation

Professional Positions:

- Research Assistant, New Jersey Institute of
  Technology, Newark, NJ 1993-1995

Publications and Presentations:

Qiuzi Chen, Patrick Beaton, and Hamou Meghdir,
"A Profile of Employee Transportation Coordinator."
Transportation Research Board.
Washington, D.C. January 1995

Qiuzi Chen, Patrick Beaton, and Hamou Meghdir,
" The Role of the Employee Transportation Coordinator
Employee Trip Reduction Program."
1994 ACT International Conference Proceedings.
Miami, FL. September, 1994

This thesis is dedicated to
my mother, a great engineer who had a profound
impact on my life and demonstrated a life of independence

ACKNOWLEDGMENT

The author would like to express her sincere gratitude and appreciation to all who provided valuable assistance and advice throughout the course of this research. Special thanks are due to Professor W. Patrick Beaton, Professor Naomi Rotter and Professor Kyriacos Mouskos, my thesis advisors, for their generous advice and guidance and to Dr. Louis J. Pignataro, Director of the Center for Transportation Research and Studies (CTSR) at New Jersey Institute of Technology, for his generous efforts in providing funding for the support of this research. Support for this research was provided from the contract Intelligent Vehicle Highway Systems and the New Jersey Institute of Technology.

The author also would like to thank especially Professor Tony Fowkes and Professor Mark Wardman from Leeds University, London, and Professor John Polak from Oxford University, London, for their invaluable effort in producing research in producing research design and theory fundamental to the basis of my research.

Finally, the author would like to thank Professor Tom Abbott from Rutgers University for his insights in the generation of qualified random numbers and Cathy Chen for her gracious support in programming.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF TABLES
(Continued)

Table Page

LIST OF TABLES
(Continued)

Table **Page**

LIST OF TABLES
(Continued)

CHAPTER 1

INTRODUCTION

Data generation introduces ambiguity and error into research designs. Two data generation techniques common to consumer demand analysis are compared and examined for their unintended consequences toward model estimation.

This research explores the potential biases in both the Revealed Preference (RP) and the Stated Preference (SP) technique based on results of simulation. In a RP study, respondents are asked to report what they did in the past such as yesterday or last week. They are also asked to report the values of the determinants of choice selected by the investigator. When a RP technique is applied in transportation, respondents are usually asked to report their daily travel time, travel cost or commuting modes, etc. On the other hand, in a SP study, respondents are presented a choice set consisting of hypothetical scenarios and asked to rank, rate, or make a choice. The respondents are told that only mode choices specified in the choice set are available to them. Problems arise when each of these generate different conclusions (Bates, 1988). This research demonstrates the problems in the data generation process in the context of new laws encouraging the use of high occupancy vehicles.

1.1 General Background

The federal Clean Air Act Amendments (CAAA 1990) section 108f requires that states with areas of severe non-attainment must show evidence in their state implementation plans that their Average Vehicle Occupancy (AVO) will improve by 25 percent by November 15, 1996. Based on the federal Act, NJ's Employee Trip Reduction Program (ETRP) mandates that every employer with more than 100 employees submit a compliance plan in which specific Employee Commute Option (ECO) strategies are to be implemented at their sites by November 15, 1994. By that date, every affected employer must show that they have implemented the ECO strategies and met the target APO. Both the federal Clean Air Act and NJ ETRP suggest a series of trip reduction strategies: parking management, guaranteed ride home, High Occupancy Vehicle (HOV) subsidy and other soft strategies including shifting work hours and telecommunications. The situation facing employers is clear. Significant changes in work rules and commuting practices must occur. Employers need to know: What specific combinations will permit for a site to improve its APO? And what are the constraints employees have preventing them joining rideshare program or transit? A successful APO forecasting model attempts to answer these questions.

1.2 Motivation

NJ DOT has provided a model called "NJECO" to help employers forecast the effect of the ECO strategies on the site's APO(Comsis, 1994). The model was estimated by using the Revealed Preference (RP) techniques and was transferred from California. On the other hand, Beaton (1992) developed a forecasting model in New Jersey by using Stated Choice (SC) technique. The model was developed at a specific site in New Jersey: the Matsushita Electric Corporation of America (MECA). The purposes and the estimation process of the two models are the same. However, the estimated effects of trip reduction strategies in improving the Average Passenger Occupancy (APO) deviate between these two models. The difference between these two models lies in data generation. RP approach relies on reported commuting choices and commuting habits combined with census data and data kept by the company. SC approach generates hypothetical data by presenting respondents hypothetical choice scenarios and asking them to make a choice. Do the underlying research design differences produce any difference in the estimated coefficients? If yes, how significant is the difference? My thesis explores these questions by using a simulation technique.

1.3 Introduction to Stated Preference and Revealed Preference Techniques

1.3.1 Stated Choice Technique in Transportation

The stated choice approach to rational compensatory decision making analysis simplifies the choice making context, target, and time by making a verbal or pictorial description of a choice problem. Respondents are asked to examine the choice tasks presented to them and pick one alternative they prefer most in the choice set. This approach deals with hypothetical situations involving stated or intended behavior. Usually, each respondent is asked to examine 16 to 27 choice tasks depending on number of attributes and number of value levels for each design variable. Usually, a fractional factorial experimental design is used to produce a set of orthogonal design variables where each design variable is comprised of 2 or 3 value levels.

The setup for an SP study includes the following steps:

1. Focus group meeting,

2. Selection of the Sample,

3. Selection of number of attributes and number of levels within an attribute,

4. Selection of the measurement of choice, and

5. Data analysis

Focus group meetings are held in order to present the researchers with a better understanding of the target population. Therefore, researchers are able to set up appropriate design variables and assign reasonable values to the design variables.

The major consideration in selecting a sample is the representiveness of the target population. The target population must consist of persons who are knowledgeable and capable of responding to the choices offered in an experiment. An SP study does not require that respondents are currently making trade-offs among the choices that are presented to them. However, the choices presented to them must be meaningful in terms of purpose, rationality and efficiency.

The number of attributes and the number of levels of each attribute depends on researchers' interests. If more attributes and more levels of each attribute are included, more information will be obtained. However, the problem of respondent fatigue will occur as more attributes and levels of each attribute are included (Fowkes, 1988).

Respondents can be asked to rank, rate or make a choice of the presented choice set. Stated Choice (SC) is considered as the closest one to the reality and thus is applied most often.

Multinominal logit model is the most often one to be used in revealed preference disaggregate travel demand modeling and is also used in a SP or SC study. The theoretical frame, model formulation and the statistical estimation procedure will be described in Chapter 2.

1.3.2 Revealed Preference Technique in Transportation

In a RP study, respondents are asked to report what they did in the past. In transportation, respondents are usually asked to report their daily travel time, travel cost, mode choice and number of transfers if bus or train is actually used. The RP technique requires the respondents in the sample to be making trade-offs among the choices researchers are interested in. This requirement increases the difficulty of identifying the sampling frame and increases the cost of the study. In addition, unlike in an SP study, where each respondent makes several choices, only one choice can be observed from every respondent in an RP study.

Multinominal logit model is the most common one to be used for data analysis in an RP study. The maximum likelihood estimation procedure is employed to estimate the parameters in the choice model.

1.4 Thesis Structure

The thesis will be prepared in five chapters. Chapter one introduces the problem facing employers, researchers, and planners as well as the need to compare Revealed Preference (RP) and Stated Choice (SC) techniques by simulation. Chapter two discusses individuals' logit decision making process. Chapter three explores the consequences of problems associated with SP and RP by mathematical approach. Chapter four discusses the theoretical difference between Revealed Preference and Stated Choice techniques in terms of data generation and explores these differences by simulation. Chapter five applies the results of the simulation to the RP model and the SC model that have been estimated or used in New Jersey, presents conclusions from study, and suggests some directions for future research.

CHAPTER 2

LOGIT DECISION MAKING PROCESS IN TRANSPORTATION

2.1 Introduction

Section 2.2 introduces the underlying theoretical model for the rational individual's decision making process. The theory of utility maximization combined with the concepts of derived demand and indirect utility functions allows researchers to specify the components of the utility function and explain why choice A rather than choice B is chosen. Section 2.3 introduces the set of assumptions on the distribution of the random utility term in the indirect utility function. Section 2.4 introduces the procedure used in estimating the parameters in the utility function: Maximum Likelihood Estimation Procedure.

2.2 Theory of Discrete Choice

2.2.1 Indirect Utility

Utility in the study of consumer behavior is interpreted as *the "satisfaction" derived from the consumption of alternative consumption bundles (Phlips, 1974)*. The utility function for a consumer is expressed as:

$$u = f(x_1, x_2, \ldots, x_n) \tag{2.2.1.1}$$

Equation 2.2.1.1 is a direct utility function in that it takes commodities as arguments. Applying the direct utility function to the study of a commuter's mode choice behavior in transportation, equation 2.2.1.1 can be expressed as:

$$u = f(x_j, x_H) \tag{2.2.1.2}$$

In equation 2.2.1.2, x_j is commuter transportation services such as SOV or HOV, etc. x_H represents all the other goods and services in an individual's consumption, called Hicks goods (Ferguson, 1972). Symbol u is the satisfaction a commuter gets from consuming transportation service j and Hicks goods. Transportation services satisfy the three properties essential for the identification of a commodity: nonnegativity, divisibility, and unbounded from above (Phlips, 1974). Individuals are assumed to maximize their utility subject to constraints in money and time budgets allocated to commuting. The optimal level of commuting transportation service (x_j^*) generates a demand function ϕ having arguments such as travel cost, travel time, and income budget for commuting, etc. The expression for x_j^* is:

$$x_j^* = \Phi(\alpha_1, \alpha_2, \ldots, \alpha_n, i, t) \tag{2.2.1.3}$$

Equation 2.2.1.3 states that the prices or costs for transportation, tastes, anticipations regarding the future $\alpha_1, \alpha_2, \ldots, \alpha_n$ as well as the consumer's income budget i and time budget t determine the demand for transportation services. Equation 2.2.1.3 is an individual's demand function for commuting, given income budget for commuting i, time budget allocated for commuting t, and various aspects of the mode. However, the demand for commuting is a derived demand. Individuals do not usually obtain utility directly from commuting but from work accomplishment and paycheck (Button, 1993). In order to accomplish work and receive paycheck, individuals must commute to work. The idea that demand for commuting is a derived demand implies that for most of the commuters, the utility of commuting is negative. Here commuters minimize the disutility of commuting at the optimal demand for transportation services.

If we place equation 2.2.1.3 into equation 2.2.1.2, we obtain a function establishing a direct link between the utility derived from a commuter's consumption of transportation services and various price aspects of transportation services and the commuter's income budget, called indirect utility function, which can be written as:

$$u = f^{*}(\alpha_1, \alpha_2, \ldots \alpha_n, i, t) \tag{2.2.1.4}$$

Where star * refers to the utility maximizing solution to the direct.

Recognition of the commuter's utility function for consuming transportation services as an indirect utility function is of vital importance in studying the commuter's decision making process. The policy variates of interest are changes in price of the various aspects of a mode and changes in an individual's income.

The relationship between the direct utility and indirect utility function is a duality relationship. For simplicity, I only consider commuting trips and all the other goods and services (Hicks goods). In addition, I assume that travel cost and travel time are the only two inputs supplied by the commuter. The objective function in the primal shows theutility function for commodities such as transportation services and Hicks goods.

Primal Problem

Maximize:

$$U = c_j x_j + c_H x_H \tag{2.2.1.5}$$

Subject to:

$$\begin{aligned} &p_{1j} x_j \leq i_j \\ &p_{2j} x_j \leq t_j \\ &p_{1H} x_H \leq i_H \\ &p_{2H} x_H \leq t_H \\ &i_j + i_H = i \\ &t_j + t_H = t \\ &x_j > 0, x_H > 0 \end{aligned} \tag{2.2.1.6}$$

In equations 2.2.1.5 and 2.2.1.6, U is the total utility an individual obtains from commuting transportation services and Hicks goods,

x_j refers to the number of commuting services consumed by the commuter,

x_H represents the number of Hicks goods consumed by the commuter,

c_j is the utility per unit of commuting service consumed by the commuter (given the derived nature of commuting services, this value will be negative),

c_H is the utility per unit of Hicks goods consumed by the commuter,

p_{1J} is the \$ spent per commuting service,

p_{2J} is the minutes spent per commuting service,

p_{1H} is the price of one unit of Hicks goods,

p_{2H} is minutes spent for one unit of Hicks goods,

i_J is income budget for commuting services,

t_J is the time budget for commuting services,

i_H is the income budget for all the other Hicks goods,

t_H is the time budget for all the other Hicks goods,

i is the total income budget, and

t is the total time budget.

Given time and income budgets and prices, the commuter must choose an optimal amount of consumption of transportation services in contrast to Hicks goods in order to maximize utility.

Dual Problem:

Minimize:

$$y = \alpha_1 i_J + \alpha_2 t_J + \alpha_3 i_H + \alpha_4 t_H \qquad (2.2.1.7)$$

Subject to:

$$\begin{aligned}
&\alpha_1 p_{1J} + \alpha_2 p_{2J} \geq c_J \\
&\alpha_3 p_{1H} + \alpha_4 p_{2H} \geq c_H \\
&i_J + i_H = i \\
&t_J + t_H = t
\end{aligned} \qquad (2.2.1.8)$$

In equations 2.2.1.7 and 2.2.1.8, y is the consumer's total utility of spending the available resources in commuting and Hicks goods,

α_1 is the utility per dollar spent for commuting,

α_2 is the utility per minute spent for commuting,

α_3 is the utility per dollar spent for Hicks goods, and

α_4 is the utility per miniute spent for Hicks goods.

The objective of the dual problem can be interpreted as minimizing the utility of spending the available resources needed in the process of consuming commuting services and Hicks goods such that utility in the primal will be maximized. The constraint function in the dual problem states that the contribution to the utility of the dollars and time used for commuting and Hicks goods must be enough to produce the optimum amount of commuting services and Hicks goods respectively.

The solution in the dual problem leads to the specification of the indirect utility function used in either the revealed or the stated preference approaches to discrete choice analysis. The constraint functions of the dual are the indirect utility functions for commuting services and Hicks goods. Travel cost: p_{1J}, and travel time: p_{2J} are the attributes for the utility function specifying commuting services; while,the price of Hicks goods: p_{1H}, and time spent for Hicks goods: p_{2H} are the arguments of the indirect utility function for Hicks goods. The typical problem facing revealed and stated preference studies is the estimation of the parameter estimates for $\alpha_1, \alpha_2, \alpha_3$, and α_4.

2.2.2 Utility Maximization and Discrete Choice Theory

Utility Maximization is a fundamental condition in the discrete choice theory. It states that alternative i is prefered to alternative j if $U_i > U_j$, $j \neq i$. The utility is composed of two elements: systematic utility V_i and random utility term ε_i, of which only the systematic utility can be observed. The random utility term reflects individual ideosyncracy (Hensher & Johnson, 1981). Essentially, a population of consumers is created in which the individuals within the population are represented by the values assumed by the random utility term.

The random utility model for individual q is expressed as:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{2.2.2.1}$$

If alternative i is prefered to alternative j,

$$V_{iq} + \varepsilon_{iq} > V_{jq} + \varepsilon_{jq}$$

$$V_{iq} - V_{jq} > \varepsilon_{jq} - \varepsilon_{iq}\,, \quad j \neq i \tag{2.2.2.2}$$

The elements of $\varepsilon_{jq} - \varepsilon_{iq}$ can not be directly measured. Random utility term is essentially a single aggregate residual term. Therefore, the relationship $V_{iq} - V_{jq} > \varepsilon_{jq} - \varepsilon_{iq}$

can not be determined from direct observation. However, given an assumed probability distribution for the residual term, a probability that $\varepsilon_{jq} - \varepsilon_{iq}$ is less than $V_{iq} - V_{jq}$ can be assigned.

$$P_{iq} = P[(\varepsilon_{jq} - \varepsilon_{iq}) < (V_{iq} - V_{jq})], \text{ for all } j \neq i \tag{2.2.2.3}$$

The above probability function means that the probability that alternative i is chosen is equal to the probability that the difference between ε_{jq} and ε_{iq} is less than the difference between V_{iq} and V_{jq}. By simply rearranging the equation 2.2.2.3, equation 2.2.2.3 can be expressed as:

$$P_{iq} = P[\varepsilon_{jq} < V_{iq} - V_{jq} + \varepsilon_{iq}], \text{ for all } j \neq i \tag{2.2.2.4}$$

Since ε_{iq} is random with an assigned distribution, we can define all the possible values of ε_{iq} as b_l (l=1,2,3,....r). Therefore, equation 2.2.2.4 can be expressed as:

$$\begin{aligned} P_{iq} = {} & P[\varepsilon_{iq} = b_1, \text{and } \varepsilon_{jq} < V_{iq} - V_{jq} + b_1, j \neq i] \\ & + P[\varepsilon_{iq} = b_2, \text{ and } \varepsilon_{jq} < V_{iq} - V_{jq} + b_2, j \neq i] \\ & + \cdots \\ & + P[\varepsilon_{iq} = b_r, \text{ and } \varepsilon_{jq} < V_{iq} - V_{jq} + b_r, j \neq i] \end{aligned} \tag{2.2.2.5}$$

Where r is equal to the number of all possible values of ε_{iq}.

Alternatively, equation 2.2.2.5 can be expressed as:

$$P_{iq} = P[\varepsilon_{iq} = b_1][P(\varepsilon_{jq} < V_{iq} - V_{jq} + b_1,\ j \neq i)]$$

$$+\cdots$$

$$+P[\varepsilon_{iq} = b_r][P(\varepsilon_{jq} < V_{iq} - V_{jq} + b_r,\ j \neq i]$$

$$= \sum_{l=1}^{r} P[\varepsilon_{iq} = b_l][P(\varepsilon_{jq} < V_{iq} - V_{jq} + b_l), j \neq i] \qquad (2.2.2.6)$$

If we further assume a continuous distribution of $b_l's$ $(-\infty, \infty)$, equation 2.2.2.6 can be written as:

$$P_{iq} = \int_{-\infty}^{\infty} P[\varepsilon_{iq} = b_l][P(\varepsilon_{jq} < V_{iq} - V_{jq} + b_l), j \neq i]\, db_l \qquad (2.2.2.7)$$

Equation 2.2.2.7 is a general formulation of a choice model expressing the relationship between the probability of selecting an alternative and the attributes of the alternatives in the choice set, under the condition that the distribution of random utility term is defined by researchers.

Conversion of equation 2.2.2.7 to an operation model is described in the following section.

2.3 Model Formulation

A number of assumptions have to be introduced in order to convert model 2.2.2.7 developed in the previous section to an operational model. The main assumption to develop a simple operational model is: Independence-from-Irrelevant Alternative (IIA). The IIA property states that *the ratio of the probabilities of choosing one alternative over another (where both alternatives have a non-zero probability of choice) is unaffected by the presence or absence of any additional alternatives in the choice set (Hensher & Johnson, 1981)*. Another assumption added by McFadden (1975) is positivity. Given the transportation related attributes and socioeconomic characteristics, the probability of a specific alternative being chosen in a choice set must be greater than zero for all possible alternatives in the choice set.

Suppose the utility model for an individual q is expressed as:

$$U_{jq} = V_{jq} + \varepsilon_{jq}, for\ j = 1,2,\ldots.J \qquad q = 1,2,\ldots.Q \tag{2.3.1}$$

Where j is any alternative in the choice set J and q is any individual in the sample Q.

If we assume that the random utility term ε_{jq} is independently and identically distributed with weibull distribution, the definition of weibull distribution in terms of ε_j's is:

$$P(\varepsilon_{jq} \le \varepsilon) = \exp\{-\exp[-\mu_q(\varepsilon - \alpha_{jq})]\} \tag{2.3.2}$$

Where μ_q is the shape parameter (variance of ε_{jq}) across individuals,

α_{jq} is the location parameter (mean value) of ε_{jq} across both individuals and alternatives, and,

ε can be taken as a random variable or any given value on the distribution (ε is taken as any given value on the distribution in Johnson and Hensher's version to derive the logit model, which is also introduced here).

Assuming the location parameter α_{jq} equal to 1 and the shape parameter μ_q equal to 0, which is applied in empirical studies most often, equation 2.3.2 can be simplified as:

$$P(\varepsilon_{jq} \leq \varepsilon) = \exp(-\exp - \varepsilon) = e^{-e^{-\varepsilon}} \tag{2.3.3}$$

Where ε represents any arbitrary value from $-\infty$ to ∞ within the distribution of ε_{jq}.

Assuming that each individual has an identical distribution of ε_j and dropping the subscribe q in equation 2.2.2.4 in the previous section, we get:

$$P_i = P[\varepsilon_j < V_i - V_j + \varepsilon_i], \text{ for all } j \neq i \tag{2.3.4}$$

Since, by assumption, each ε_j is independently distributed, for any given value of ε_i, defined as b, the probality that $\varepsilon_1 < b + V_i - V_1, \varepsilon_2 < b + V_i - V_2, \ldots, \varepsilon_J < b + V_i - V_J, j \neq i$ simultaneously (jointly) can be written as:

$$P_i = P(\varepsilon_j < b + V_i - V_j) \text{ for all } j \neq i)$$

$$= \prod_{\substack{j=1 \\ j \neq i}}^{J} \exp(-\exp-[b + V_i - V_j]) \tag{2.3.5}$$

Recall that equation 2.3.3 specifies a definition of cumulative weibull distribution in terms of ε_j. Similarly, the definition of cumulative weibull distribution in terms of ε_i can be expressed as:

$$P(\varepsilon_i < \varepsilon) = \exp(-\exp-\varepsilon) = e^{-e^{-\varepsilon}} \tag{2.3.6}$$

However, we are interested in the probability that ε_i equals some given value: b. By definition, the derivative of a cumulative probability distribution (equation 2.3.6): the probability density function, gives the probability when $\varepsilon_i = b$.

The probability density function when $\varepsilon_i = b$ can be expressed as:

$$\frac{\partial F}{\partial b} = \exp(-\exp(-b) - b) = e^{-e^{-b} - b} \tag{2.3.7}$$

Equation 2.3.5 is a probability function for all $\varepsilon_j, j \neq i$, when $\varepsilon_i = b$. Equation 2.3.7 gives a probability when $\varepsilon_i = b$. The joint marginal density function of ε_j can be obtained by multiplying equation 2.3.5 by equation 2.3.7:

$$e^{-e^{-b}-b}\prod_{\substack{j=1 \\ j\neq i}}^{J} e^{-e^{-(b+V_i-V_j)}} \tag{2.3.8}$$

Which can be simplied as:

$$= \exp(-\exp(-b)-b)\ \exp[-\sum_{j=1}^{J}\exp(-(b+V_i-V_j))]$$

$$= e^{-e^{-b}-b} e^{-\sum_{\substack{j=1 \\ j\neq i}}^{J} e^{-(b+V_i-V_j)}} \tag{2.3.9}$$

Similar to equation 2.2.2.7, the probability of choosing a particular alternaitve i can be obtained by integrating the joint marginal density function of ε_i (2.3.9) over all possible values of ε_i:

$$P_i = \int_{b=-\infty}^{b=\infty} \exp(-\exp(-b)-b)\exp[-\sum_{j=1}^{J}\exp-(b+V_i-V_j)]db \tag{2.3.10}$$

Rearranging equation 2.3.10, we obtain:

$$P_i = \int_{b=-\infty}^{b=\infty} \exp(-\exp(-b)-b)\exp\{-\exp(-b)*[\sum_{j=1}^{J}\exp(V_j-V_i)]\}db \tag{2.3.11}$$

Let: z = exp(-b), and

$a = \sum_{j=1}^{J} \exp(V_j - V_i)$, we get:

$$z = e^{-b} \tag{2.3.12}$$

b=-lnz

$$db = (-\frac{1}{z})dz \tag{2.3.13}$$

In the equation 2.3.11, notice that: $z = \infty$, when $b = -\infty$, $z = 0$, when $b = \infty$.

Then, equation 2.3.11 can be written as:

$$P_i = \int_{\infty}^{0} z * e^{-z} e^{-za} (-\frac{1}{z}) dz$$

$$= \int_{0}^{\infty} e^{-z} e^{-za} dz$$

$$= \int_{0}^{\infty} e^{-z(1+a)} dz \tag{2.3.14}$$

$$= -e^{-z(1+a)} * \frac{1}{1+a} \Bigg|_{0}^{\infty}$$

We know that: $\exp(-\infty) = 0$, when $z = \infty$, $\exp(0) = 1$, when $z = 0$.

Eaquation 2.3.14 can be calculated:

$$P_i = -\frac{1}{1+a}(0 - 1) = \frac{1}{1+a} \tag{2.3.15}$$

$$P_i = -\frac{1}{1+a}(0-1) = \frac{1}{1+a} \tag{2.3.15}$$

Where a = $\sum_{j=1}^{J} \exp(V_j - V_i)$

Furthermore, equation 2.3.15 can be expressed as:

$$P_i = \frac{1}{1 + \sum_{j=1}^{J} \exp(V_j - V_i)} \tag{2.3.16}$$

Rearranging equation 2.3.16, we obtain:

$$P_i = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)} \tag{2.3.17}$$

Given all the assumptions above, equation 2.3.17 is the basic multinomial logit (MNL) model. Equation 2.3.17 expresses the relationship between selection probability and the attributes of the alternatives.

2.4 Statistical Estimation Process

Maximum Likelihood Estimation (MLE) is used to estimate parameters in the choice model. The basic idea of MLE is that the estimated Maximum Likelihood parameters can

$(z_1, z_2, \ldots, z_n)$ on a random variable Z from a population characterized by an unknown parameter θ. The joint probability density function of the sample is:

$$f(z_1|\theta) f(z_2|\theta), \ldots, f(z_n|\theta) \tag{2.4.1}$$

A joint probability density function as equation 2.4.1 is usually interpreted considering Z as a variable and θ as fixed. However, in the MLE procedure, Zs are considered as fixed and θ as a variable. Equation 2.4.1 is considered as a likelihood function. Maximization of equation 2.4.1 with respect to θ gives the optimal value of θ which is most likely to generate the sample we observed. Extension of one parameter θ to more than one parameters in a MLE estimation gives a likelihood function which is suitable to be used in estimating parameters in either an RP or SP model choice model. Recall equation 2.3.14 that the probability of an individual q's choosing alternative i is:

$$P_{iq} = \frac{\exp(V_{iq})}{\sum_{j=1}^{J} \exp(V_{jq})} \tag{2.4.2}$$

of which the systematic utility V_{jq} is usually assumed to be a linear additive function.

Assume V_{jq} is expressed as:

$$V_{jq} = \sum_{k=1}^{K} \alpha_{jk} x_{jkq} \tag{2.4.3}$$

$$V_{jq} = \sum_{k=1}^{K} \alpha_{jk} x_{jkq} \tag{2.4.3}$$

Where V_{jq} is individual's systematic utility for alternative j,

x_{jkq} is the kth attribute for individual q and alternative j,

K is the total number of attirbutes, and

α_{jk} is the parameter for the kth attribute in the utility function for alternative j.

In either an RP or an SP study, for every respondent in the sample, we observe choices made by the respondent as well as the values of x_{jkq} for all alternatives. Assume that every respondent in the sample makes choices independently, the joint likelihood function of the sample can be expressed as the multiplication of the probability of each individual's choosing alternative j. This can be expressed as:

$$L = \prod_{q=1}^{n_1} P_{1q} \cdot \prod_{q=n_1+1}^{n_1+n_2} P_{2q} \cdots \prod_{q=Q-n_J+1}^{Q} P_{Jq} \tag{2.4.4}$$

The observations in the equation 2.4.4 are ordered so that the first n_1 observations choose alternative 1, the next n_2 observations choose alternative 2,....and the last n_J observations choose alternative J.

To mathematically work with equation 2.4.4, the logarithm of equation 2.4.4 (L^*) is as follows:

$$L^{*}=\sum_{q=1}^{n1}P_{1q}+\sum_{q=n1+1}^{n1+n2}P_{2q}+\ldots+\sum_{q=Q-nJ+1}^{Q}P_{Jq} \qquad (2.4.5)$$

In equation 2.4.5, L^{*} can be maximized with respect to all the αs in the systematic component of the utility. The resulting estimates of all αs are the MLE estimates for the parameters in the utility function.

The computer program used in simulation work: ALOGIT, uses an iterative procedure to estimate the parameters. The system assigns initial values of all α_s first. These values are used to calculate the value of V_{Jq}, which is then used to calculate P_{Jq}. After obtaining each respondent's P_{Jq}, the system will use these values to calculate a starting value of L^{*} in equation 2.4.5. Then in a Newton-Raphson method, the system will continuously search for better values of α_s until the increase in L^{*} does not exceed a preassigned value. The Newton-Raphson method calculates the slope and the curvature of the function as well as the value of L^{*}. *The slope of the function tells what direction the coefficients must be changed in order to improve the function. The curvature of the function tells how far the coefficients must be changed before L^{*} stops increasing (ALOGIT User's Guide, 1992).*

CHAPTER 3

METHODOLOGICAL DIFFICULTIES: A REVIEW OF THE LITERATURE

The literature review begins by examining the pros and cons of revealed preference and stated preference techniques. In section 3.2, problems in estimation process using both RP and SP data are discussed.

3.1 Pros and Cons of RP and SP Approaches

Revealed preference is the conventional approach to estimating transportation choice models and forecasting travel behavior. However, use of the stated preference approach is increasing due to its efficiency in the use of data and reduced cost of conducting a study compared to revealed preference approach (Bradley & Kroes, 1992). The discussions on pros and cons of SP and RP are listed as follows (Jones, 1990):

1. Model Specification: In an RP study, researchers either observe respondents' behaviors or ask respondents report their past behavior, and then model the relationship between the choice and the factors that researchers think affect the respondents' decision making process. Therefore, researchers will not be able to observe all the factors that affect in the respondent's decision making process. Consequently, they can omit some important variables or misspecify the form of the factors in the model. The consequences of omitting important variables or misspecifying some factors when modeling actual behavior produce biased estimates among the remaining variables and poor forecasts.

In an SP study, researchers define the attributes of the alternatives they are interested in and ask respondents make trade-offs among the attributes and choose the most desirable alternative. Therefore, the problems of omitting important variables or misspecifying some factors in the model in an RP study do not apply in the estimation component of the SP study.

2. Statistical Estimation: RP approach is challenged in two dimensions of statistical estimation: correlation among explanatory variables and errors in variables.

(1) Correlation among explanatory variables: In an RP study, researchers have little control over the explanatory variables. As a consequence, correlation among explanatory variables are often very strong. The correlation among explanatory variables prevents researchers from estimating the relative effect of each factor on respondents' decision making by increasing the standard errors (Fowkes & Wardman, 1994). In an SP study, the variables of interest are set up in an orthogonal data matrix. This arrangement essentially eliminates problems of multicolinearity.

(2) Errors in variables: The strongest argument favoring RP over SP reflects the factual nature of choice. This argument is correct if it refers to the actual choice. The choice values in an RP study are of higher quality than the choice data in an SP study. In RP the choice actually occurred; the SP data is a response to a hypothetical situation. However, this argument is inaccurate when it refers to the identification of rejected alternatives as well as the attributes of the chosen as well as the rejected alternatives. The attributes in an RP study are assigned values by respondents to transportation services. It is common for respondents to have errors when reporting their values. The consequences

of errors in variables problems to the statistical estimation process is described and demonstrated in Chapter four . In a SP case, because the design variables are fixed, errors in variables is not an issue in statistical estimation process.

3. Range of applicability: The RP approach can not be used in the situations where the choice alternatives of interest do not exist. However, in a SP study, respondents can be given a detailed description of an alternative condition even though it does not exist in the real world.

4. Types of variables: Given the level of effort needed to construct and administer an RP study, the RP approach is used to estimate the relative effect of major transportation service variables such as travel time and travel cost but not those hard-to-measure transportation variables such as comfort level and safety.

5. Omitted alternatives in RP: One of the strict requirements to conduct an RP study is that every respondent in the sample must make trade-offs among the alternatives in the choice set during the time period of interest. The RP data usually contains information on the mode the respondent chose as well as the attribute values of the mode chosen. Revealed preference studies seldom have information on the rejected modes for that specific respondent.

One solution that has been recommended for this problem is an increase in the size of the sample. Here the hope is to bring a more diverse range of respondents into the sample. However, this strategy does not resolve the fundamental problem. Measurement error still exists and non-compensatory tradeoffs are built into the RP model.

The linear additive utility function utilized in both RP and SP is a source of problem to both approaches. Consumers are assumed to weigh the attributes within and across their mode specific utility functions in the process of making a choice. In RP, the commuter is assumed to have considered the attributes included in the model and chosen the mode used. In an SP study, various choice scenarios are designed for respondents to make trade-offs. The information for an alternative is sufficient for researchers to estimate parameters in the choice model. If a respondent does not make any trade-offs through the whole set of choice scenarios, noncompensatory behavior will be assumed and the response will be taken out of the study.

The arguments mentioned above discuss the strength and shortcomings of both RP and SP approaches. Errors in SP mostly belong to response bias, which leads to a deviation of the SP choice from the actual choice. Bates (1988) described this problems as a scaling factor in forecasting. Discussion of scaling factor is given in the section 3.3.

3.2 Econometric Issues in Both RP and SP

As data generation methods, both RP and SP are subject to certain errors that can not be avoided in the statistical estimation process. Errors associated with RP include:

1. Errors-in-variables,
2. Model specification,
   (1) Omitted variables,
   (2) Misspecify the functional form of the independent variable,
   (3) Interaction among independent variables, and

3. Measurement error in the dependent variable.

The RP data are based on reported data in that respondents are asked to report their travel time, travel cost, or travel modes, etc. Errors in the independent variables are assumed to be random with a given distribution. The effect of errors in variables will force the estimates to attenuate toward zero, depending on the magnitude of the variance of the error term.

The classical Ordinary Least Square (OLS) technique can be used to derive estimate of β which is comparable to the Maximum Likelihood Estimate (Fowkes, Wardman & Holden, 1994). The authors indicate that it makes no material difference whether ε_i are assumed Weibull rather than Normal when it comes to examining the consequences of non-orthogonality on parameter estimates. I will use the same approach in examining the consequences of errors in variables on parameter estimates.

Suppose the true regression model is:

$$y_i = \beta x_i + \varepsilon_i \tag{3.1}$$

Where ε_i is normally distributed with 0 mean value and variance δ_ε^2.

While the actual regression model is:

$$y_i = \beta x_i* + (\varepsilon_i - \beta v_i) = \beta x_i* + \varepsilon* \tag{3.2}$$

Where $x_i* = x_i + v_i$.

Assuming the error term: v_i, in X is normally distributed with 0 mean and has no serial correlation, and also is independent of the error: ε_i, in the true equation. The covariance between x_i* and ε_i* is:

$$Cov(\varepsilon_i*, x_i*) = E[(\varepsilon_i - \beta v_i)(x_i + v_i)] = -\beta\, \delta_v^2 \tag{3.3}$$

Then the ordinary least-square estimator β':

$$\beta' = \frac{\sum x_i^* y_i}{\sum x_i^{*2}}$$

$$= \frac{(\sum x_i + v_i)y_i}{\sum (x_i + v_i)^2}$$

$$= \frac{\sum (x_i + v_i)(\beta x_i + \varepsilon_i)}{\sum (x_i + v_i)^2}$$

$$= \frac{\beta \sum x_i^2 + \sum x_i\varepsilon_i + \beta \sum v_i\varepsilon_i + \sum v_i\varepsilon_i}{\sum x_i^2 + \sum v_i^2 + 2\sum x_i v_i} \tag{3.4}$$

Since ε_i and v_i are all stochastic, the bias of β' is difficult to estimate. However, the consistency of β' can be evaluated by evaluating the expression of β' in the limit as the sample size gets large.

$$\text{plim } \beta' = \text{plim } \frac{\beta \sum x_i^2}{\sum x_i^2 + \sum v_i^2}$$

$$= \frac{\beta \operatorname{var}(x)}{\operatorname{var}(x) + \delta_v^2}$$

$$= \frac{\beta}{1 + \delta_v^2 / \operatorname{var}(x)} \tag{3.5}$$

This suggests that the presence of errors in variables in a two-variable regression model will lead to an underestimation of the true regression parameter.

A comparable analysis of the consequences of errors in variables upon the multinomial logit model is not available in the literature. However, an analysis related to the multinomial probit is available (Yatchew & Griliches, 1985). Probit as well as logit are members of a general class of models identified as generalized extreme value (GEV) models. The members of GEV are used to examine multinomial choice problems. The findings in the case of probit analysis can indicate the consequences of errors in variables for logit. The following shows the consequences of errors in variables as found in the probit model.

Suppose the basic probit model we are working with is expressed as:

$$y_i = \beta\, x_i + \varepsilon_i \tag{3.6}$$

$$\begin{aligned} y_i^* &= 1 \quad \text{if } y_i \geq 0 \quad i = 1, \ldots, K \\ &= 0 \quad \text{if } y_i < 0 \end{aligned} \tag{3.7}$$

Where y_i is not observed, y_i^* is observed, and

x_i is normal with mean zero and variance δ_x^2,

ε_i is normal with mean zero and variance δ_ε^2.

Suppose that x_i is measured with error, and

$$x_i^* = x_i + v_i \tag{3.8}$$

Where x_i^* is observed data of x_i with error built in, and

v_i is normal with mean and variance δ_v^2 and uncorrelated with ε_i.

Then, equation 3.6 can be written as:

$$y_i = \beta x_i^* - \beta v_i + \varepsilon_i \tag{3.9}$$

The resulting MLE estimator of β in equation 3.9 is:

$$\beta' = \frac{\beta\delta_x^2 / (\delta_x^2 + \delta_v^2)}{\sqrt{\delta_\varepsilon^2 + \beta^2 (\frac{\delta_v^2 \delta_x^2}{\delta_v^2 + \delta_x^2})}} \tag{3.10}$$

Comparing equation 3.10 with the OLS model, equation 3.5, probit estimates based upon errors in variables data show a compounded bias toward zero.

Model specification errors include omitted variables, misinterpretation of the way the independent variable is entered and interaction factor. In an RP study, researchers can

Model specification errors include omitted variables, misinterpretation of the way the independent variable is entered and interaction factor. In an RP study, researchers can not always observe every factor in the respondents' decision making process. Therefore, important determinants of choice can be easily omitted in the RP studies. Yatchew and Griliches (1985) indicated there are various asymptotic biases in MLE estimates in a probit model if variables are omitted in the model. The type of bias depends on the assumption that the sample is drawn. If OLS techniques are used, the estimators will be biased and inconsistent (Pindyck & Rubinfeld, 1991). Form of the estimation equation is another source of error. The misspecification of a linear model when the true model is nonlinear can also lead to biased and inconsistent parameter estimators (Pindyck & Rubinfeld, 1991).

Interaction is another source of error that can influence the estimation process. Interaction can occur at both the level of association between fixed independent variables and at the causation within a structural model. Little has been done to explore the impact of interaction errors on the logit model.

Measurement error focuses attention upon the elements of experimental design. In RP studies, efforts to reduce its impact on estimation have lead researchers to make assumptions about *spatial aggregation in zoning systems, ticket type, fuel type, etc. (Bradley & Kroes, 1990)*. Measurement errors are likely to be random, with an effect similar to those of errors in variables discussed above.

On the other hand, errors associated with SP include:

1. Response error: Response error includes respondent fatigue and respondent learning.

(1) Respondent fatigue: Respondent fatigue is likely caused by complexity and longevity of the survey. Thus the dependent variable involves another type of error η_i other than the regression error ε_i. The variance of η_i increases as a respondent goes through the survey. Discussion on how η_i affects the forecasting component is described later in this chapter.

(2) Respondent learning: Respondent learning refers to respondents learn as they go through the survey. Therefore, the variance of η_i decreases as the respondents learn how to perform the task.

Both respondent fatigue and learning represent another type of error η_i in dependent variable, so there will be no problems in the estimating the parameters in the model. However, problems will appear in forecasting.

2. Justification bias: Respondents may rationalize their actual behavior in a way they think they are supposed to in an experiment. For example, they may choose High Occupancy Vehicle (HOV) modes such as carpool and vanpool when they think they are expected to. The effect of the justification bias is a shift to the alternative-specific constant to favor those modes that are expected to be chosen.

3. Policy response bias: Respondents may consciously bias their answers in the experiment hoping that their answers would affect the policy-making. The consequences of policy response bias are expected to have similar effects as that of justification bias.

4. Model specification errors: Model specification error describes the situation within SP where constraints beyond those specified in the design variables, such as weather, car availability, etc., affect choice decisions (Bradley & Kroes, 1990). The effect of this error is similar to that of omitted variables in the RP case.

The potential biases resulted from the errors associated with RP and SP can be quantified by simulation. However, this research focuses on several potential error effects that are common to the estimation phase of RP and SP. In the RP case, simulation work focuses on errors-in-variables; while, in the SP case, simulation work focuses on varying the internal design of data matrix.

3.3 A Note on the Forecasting Problem in SP

Most of the biases linked to SP occur in the experimental design phase (Bonsall, 1983). Response bias can be controlled during the survey design and administration process; however, problems remain in the forecasting phase. Bates (1988) discussed the effect of the response bias in estimation and forecasting components, which is described as follows.

Suppose that the dependent variable involves a measurement error η_i, the researchers observe a pseudo-utility $\hat{U}$ instead of the true utility U. The relationship between pseudo-utility and the true utility can be defined as:

$$U_i = V_i + \varepsilon_i = \hat{U} + \eta_i \tag{3.11}$$

Where η_i is independently distributed with the same type of distribution as ε_i with constant variance δ_η^2. By simply rearranging equation 3.11, the observed utility can expressed as:

$$\hat{U}_i = V_i + (\varepsilon - \eta_i) = V_i + \varepsilon_i^* \qquad (3.12)$$

Where ε_i^* is the new error term combining both ε_i and η_i.

From equation 3.12, the regression error term ε_i is combined with η_i. Problems occurred in the estimation can be controlled. However, problem comes into forecasting. In forecasting, utility is calculated based on the utility model and the available information on attributes. If the following is observed by researchers in the estimation process:

$$\hat{U}_i > \hat{U}_j, i \neq j, j \in A,, \text{ or}$$

$$U_i - \eta_i > U_j - \eta_j, j \neq i, j \in A \qquad (3.13)$$

Equation 3.13 does not necessarily imply that: $U_i > U_j$.

In estimation, $\varepsilon_i^* = \varepsilon_i - \eta_i$, the new variance for ε_i^* is $(\delta_\varepsilon^2 + \delta_\eta^2)$, which is greater than that would use in forecasting, generally. In forecasting, the scale of the coefficients in V_i to the random term has changed compared to those in estimation component. Thus the estimated demand for a specific alternative are affected. Therefore, the error term used in estimation must be apportioned, theoretically. Or, alternatively, the scale of the coefficients in V_j to the random term must be corrected. Unfortunately, Bates (1988) indicated that *knowledge on how to do this in practice is lacking.* This problem with scaling factor can be demonstrated by simulation. However, this research only deals with problems in the estimation process.

3.4 Conclusions

Multinomial or discrete choice theory has advanced planners ability to model the demand for transportation infrastructure. The two primary approaches to operationalizing discrete choice theory are Revealed Preference and Stated Preference studies. Both approaches have their strengths and weaknesses. Neither approach can be used as a benchmark for truth. Rather, both approaches must be used, where possible, in a complementary fashion.

CHAPTER 4

SIMULATION

4.1 Introduction

This chapter starts with the description of the simulator and the methodology used to design and construct control models and various test variations. A random number generator written in UNIX C is used to simulate the random utility term in the utility model. Concern that the results from the simulation might not be due to a change within the model but due to a poor random number generator motivated the use of statistical tests examining the quality of random number generator. The rest of the chapter describes the construction of the control models and various test variations. The test variations include: (1) Test of Varying Number of Value Levels for Design Variables, (2) Test of Varying Middle Placement within the Design Variable, and (3) Test of Errors in Variables. The chapter concludes with the comparisons between control model and various test variations.

4.2 Simulator

4.2.1 Components of Simulator

This research uses the simulator originally developed by Tony Fowkes and Mark Wardman of Leeds University (1988). The simulator represents an individual decision making process based on the principle of utility maximization (see chapter 2). The simulator can be imagined as a data matrix consisting of:

1. Logit structural model,
2. Attributes for each mode or alternative represented in the model,
3. Structural parameters for each alternative,
4. The number of value levels given to the attributes,
5. Values of the attributes, and
6. A random utility generator.

It is assumed that an individual's probability of choosing the mode i out of a choice set is governed by the decision making process implied by the logit model. The structural model is expressed as:

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \tag{4.2.1.1}$$

In the simulator, each row represents an individual's decision. The first column consists of values representing the mode choices the simulated individual makes based on utility maximization. Equation 4.2.1.1 states that the probability of an individual's choosing an alternative i is the ratio of the exponential of the utility of choosing alternative i to the sum of the exponential of the utilities of choosing the other alternatives. The individual will choose the alternative with the largest utility.

The next several columns are headed by the attributes of the alternatives in the choice set. The subdata matrix under the headings of the attributes are designed to be orthogonal in relation to the other design variables in the choice set. The orthogonal design of the combinations of values of attributes are taken from GE Research and Development Center's Catalog (1964). In an actual SP experiment, these combinations are actually presented to the respondents and they are requested to make a choice out of the alternatives. The attributes of an alternative are the arguments or independent variables affecting the individual's decision making process; examples of the independent variables include travel time and travel cost, etc. The design variables are usually assigned 2 or 3 values. Once the number of levels of the attributes is determined, the number of the choice scenarios that must be completed by each respondent can be found from the fractional factorial design plan. In addition, the orthogonal design for the values of the attributes by choice task is obtained in this catalog. Replacing the initial values of the attributes in the design plan with the values designed for the attributes of interest, a particular set of choice scenarios can be obtained.

In chapter 2, I show that an individual's utility in choosing alternative i consists of systematic utility V_i and the random utility term ε_i. The set of random utility terms: $\{\varepsilon_i\}$ represents individuals within the population and is assumed to have a weibull distribution. In the simulator, the columns following the last attribute are used to generate various uniformly distributed random numbers. A unique distribution is assigned to each alternative. Conversion of the uniform distribution into a weibull distribution is don by the following formula (Fowkes,1988):

$$\varepsilon_w = -\ln(-\ln(\varepsilon_u)) / (1.28 / stdev) \qquad (4.2.1.2)$$

Where ln is the function of natural logarithm,

ε_w is the random utility term with weibull distribution ranging from 0 to 1,

ε_v is the uniformly distribution random number ranging from 0 to 1 generated by the random number generator, and

stdev refers to the specified standard deviation of random utility term in the utility function.

In the simulation, the parameters of the attributes are assigned by the researchers. Therefore, the simulator calculates the systematic utility for each alternative. Total utility is derived by adding a random utility ε_w with weibull distribution to the systematic utility. Based on utility maximization, the simulator assigns a corresponding value to the choice variable in the first column.

After the values of the choice variables have been assigned by the simulator, the columns consisting of the choice variable and the attributes of the alternatives in the choice set become the input to a computer program that estimates the coefficients of the attributes of each alternative in the choice set.

4.2.2 Random Numbers Generator

The random numbers generator in UNIX C is used in the simulation to generate the random numbers with uniform distribution. Then the uniform distribution is transformed to weibull distribution. The random numbers generator must produce values that are identically and independently distributed with uniform distribution. Two tests are performed to assure the quality of the generator.

The random numbers generator taken for this study uses a numeric algorithm. Once a seed value is given, a sequence of random numbers is determined in that each random number is determined by one or several algorithms sequentially. There are numerous discussions on the randomness of random numbers generated by a fixed mathematical formula. However, most agree that *arithmetic generators, if designed carefully, can produce numbers that appear to be independent with uniform distribution and can pass a series of statistical tests (Law & Kelton, 1991).* That is, acceptable random numbers must be independent and exhibit no correlation with each other. Assumptions of independence and correlation are examined by using two statistical tests.

Chi-square test is used to test uniformity on random numbers. A Runs-up test is used to test independence of random numbers. Descriptions of two tests are as follows. Following the description is the an example in which 800 random numbers generated by the program are examined.

1. Chi-square test: The test examines whether the generated random numbers distribute uniformly between 0 and 1.

Suppose there are n random numbers to be tested. Divide n random numbers into k subintervals with equal length. As a general principle, k should at least be 100 and n/k should at least be 5. Let f_j be the number of the random numbers in the jth subinterval. The Chi-square with k-1 degree of freedom can be expressed as (Law & Kelton, 1991):

$$x^2 = \frac{k}{n}\sum_{j=1}^{k}(f_j - \frac{n}{k})^2 \quad (4.2.2.1)$$

The null hypothesis is: The selected n random numbers are independently and identically distributed with uniform distribution ranging from 0 to .

The alternative hypothesis: The selected n random numbers are not independently and identically distributed with uniform distribution on ranging from 0 to 1 at 95% confidence level.

For large n and k, the estimated formula for the critical value of the chi-square distribution is expressed as:

$$x_{k-1,1-\alpha}^2 \approx (k-1)\{1 - \frac{2}{9(k-1)} + z_{1-\alpha}\sqrt{2/[9(k-1)]}\}^3 \quad (4.2.2.2)$$

Where $z_{1-\alpha}$ is the upper $1-\alpha$ critical point of the normal distribution with mean of 0 and variance of 1,

α is the probability of type I error, and

k-1 is the degree of freedom.

The null hypothesis will be rejected at level α when $x^2 > x_{k-1,1-\alpha}^2$, which means that the selected n random numbers are not independently and identically distributed with uniform distribution.

The simulation uses the utility "Random()" in UNIX C as its random number generator. Just for a simple example, number 3 was chosen as the seed value for Random() to generate 800 random numbers and divide those random numbers into 100 subintervals, we get:

$$n = 800, k = 100$$
$$x^2 = 80.875 \qquad (4.2.2.3)$$
$$x_{99,0.95}{}^2 = 123.23$$

The null hypothesis is not rejected at $\alpha = 0.05$. Therefore, the selected 800 random numbers are independently and identically distributed.

2. Runs-up test: This test tests the assumption of independence. The test examines the sequence for unbroken subsequences of maximal length within which the random numbers increase monotonically; such a sequence is called a run up *(Law & Kelton, 1991)*. The test formula is given as:

$$R = \frac{1}{n}\sum_{i=1}^{6}\sum_{j=1}^{6} a_{ij}(r_i - nb_i)(r_j - nb_j) \qquad (4.2.2.4)$$

Where r_i and r_j are the number that the subsequence is equal to i or j,

a_{ij} is the (i,j)th element of the given matrix

$$\begin{bmatrix} 4529.4 & 9044.9 & 13568 & 18091 & 22615 & 27892 \\ 9044.9 & 18097 & 27139 & 36187 & 45234 & 55789 \\ 13568 & 27139 & 40721 & 54281 & 67852 & 83685 \\ 18091 & 36187 & 54281 & 72414 & 90470 & 111580 \\ 22615 & 45234 & 67852 & 90470 & 113362 & 139476 \\ 27892 & 55789 & 83685 & 111580 & 139476 & 172860 \end{bmatrix}$$

b_i and b_j are given by:

$$(b_1, b_2, \ldots, b_6) = (\frac{1}{6}, \frac{5}{24}, \frac{11}{120}, \frac{19}{720}, \frac{29}{5040}, \frac{1}{840}).$$

The same set of 800 random numbers as in the chi-square test was used in the runs-up test. It is recommended (Law & Kelton, 1991) that for large n ($n \geq 4000$), R will have an approximate chi-square distribution with 6df, under the null hypothesis that the n random numbers are independently and identically distributed. In our test, 800 observations are far less than 4000. However, the calculation of critical value of R when n is less than 4000, was not mentioned by Law and Kelton. Therefore, the critical value of $\chi_{6,0.90}{}^2$ is used in this test to examine the assumption of independence for the 800 random numbers.

The calculation showed that for the selected 800 random numbers, $R = 6.26$, $\chi_{6,0.90}{}^2 = 10.6$. Therefore, the null hypothesis of independence is not rejected at level $\alpha = 0.10$.

4.2.3 Standard Deviation of the Random Utility Term

The variance of the random utility term determines the scale of the parameter estimates to be recovered in the utility model. The variance also determines the ability to solve the maximum likelihood equation. Too little variance results in no convergence within the maximum likelihood equation and therefore no recoverable estimators. Where the variance is too large, then the deterministic utility is minimal with respect to the random utility and statistically significant estimators are not recoverable. Balance between these two extremes must be achieved.

Selection of an appropriate standard deviation leads to good estimates of the parameters in the choice model. If the standard deviation is specified too large, the estimated model will contain too much error and have low t-ratios; if the standard deviation is specified too small, the model will fail to converge (Fowkes, 1988). The method used to determine an appropriate standard deviation of the random utility term is error and trial method. The criteria used to select a proper standard deviation include:

1. Minimizing the average difference of the estimated coefficients from true parameters,

2. Maximizing the standard error of the estimated coefficients,

3. Maximizing the "t" ratios for the estimated coefficients, and

4. Maximizing the Rho-squared with respect to constant.

The estimates in the control model are expected to recover the parameters best. Therefore, the average difference of the estimated coefficients from true parameters was the most important criteria. A set of values ranging from 0.07 to 7.0 for standard deviation is tested. A standard deviation of 0.7 was determined to produce the best estimators.

4.3 Methodology

A set of control models and test variations was prepared representing realistic commuting situations. There is a control model designed for each test design. The coefficients derived by Beaton (1992) in New Jersey are taken as the structured parameters of the logit model used throughout the simulation. The values assigned to the independent variables were derived from focus group meetings with employees of the employment sites where the original SP experiment was held. The control model is specified in a design matrix consisting of the attributes of the choice alternatives and their random utilities. The chosen alternative is determined by utility maximization. The test variations are prepared by changing one or more properties of the control model.

In my simulation, the choice set faced by each simulated decision-maker consists of 16 choice tasks in which seven orthogonal-design variables are involved. By repeating these 16 choice tasks 50 times, 800 observations are constructed in the EXCEL workbook and saved as a text file. The saved text file as an input file is read by a program written in C (for detail, see Appendix-A). After it reads the input file, the program returns a random number with uniform distribution ranging from 0 to 1 and then transforms the uniform distribution to weibull distribution for each of the 800 observations. The program combines a set of parameters, and the values of the attributes and computes a systematic utility value for each alternative for each individual. The overall utility of each alternative for each individual is derived from the sum of the systematic utility and the random utility term with weibull distribution. The program assigns the alternative with the largest total utility as the chosen alternative. The output from the program is now the input to a computer program used to test the ability of the logit model to return estimates identical to their parameters.

ALOGIT program is chosen as a specific computer program used in my simulation to estimate the parameters of the attributes. Chapter 2 has discussed the detailed information about the estimation process in ALOGIT.

In order to test the ability of the logit program to recover the parameters of the attributes, the control model in each test design is estimated 50 or 100 times by varying random utility term each time. The resulted estimates for each parameter in the control model are plotted as a frequency distribution. An asymptotic 2-tailed t-test is then used to compare the estimates in the control model with the parameters.

Each test variation is simulated 50 or 100 times in the same way as the control model in order to obtain the central tendency and variability of the estimates of parameters in the test condition. A set of asymptotic 2-tailed t-tests concerning the means is performed. The tests are used to determine whether the mean value of the distribution of sample means for each design variable in the test model is significantly different from the corresponding parameter.

The estimates of parameters by ALOGIT program are the scaled values. The real estimates of parameters are obtained from the equation (Ben-Akiva, 1989):

$$\text{Real estimate} = [\,\text{Scaled estimate} / (\pi / (\sqrt{6} * stdev))\,] \tag{4.3.1}$$

After the estimates are released from the scaling factor, a set of statistical tests is performed to compare the difference between the control models and the test variations.

Tests of significant difference between the mean value of the distribution of sample means for each attribute use an asymptotic two-tailed t-test at 95% confidence level.

Suppose that the null hypothesis is: $H_0\!: x_i = \mu_i$, and

the alternative hypothesis is: $H_a\!: x_i \neq \mu_i$.

The test formula is expressed as:

$$t = (x_i - \mu_0) / (s / \sqrt{n}) \qquad (4.3.2)$$

where s is the standard deviation of the sampling distribution,

x_i is the mean of the sample distribution for design variable i ,

μ is the true parameter in the model, and

n is the degree of freedom.

At 95% confidence level, the critical value of t is 1.96. If the t-ratio is either greater than 1.96 or less than -1.96, the null hypothesis is rejected.

4.4 Test of Varying Number of Value Levels for Design Variables

4.4.1 Control Model Design

Varying the number of value levels tests the ability of the logit program to recover the parameters of the design variables while the number of value levels of the design variables is varied. The control model used in this test contains 6 design variables and one alternative-specific constant for vanpool. Of the six design variables, three are 2-level and the other three are 3-level variables.

The parameterized systematic utility models for the three alternatives are specified as:

$$V_{sov} = -0.16*PK\$\$ + 0.54*PKspace$$

$$V_{carpool} = -0.037*CPWT + 1.13*GRHcp$$

$$V_{vanpool} = 0.29 - 0.048*VPWT + 1.13*GRHvp \qquad (4.4.1.1)$$

Where PK$$ is the parking charge,

PKspace is the availability of parking space,

CPWT is the extra time over SOV when carpool is used,

GRHcp is the availability of Guaranteed Ride Home which requires 25 minutes waiting time for carpoolers,

VPWT is the extra time over SOV when vanpool is used, and

GRHvp is the availability of Guaranteed Ride Home which requires 25 minutes waiting time for vanpoolers.

The values and the levels of the design variables are shown in the following table:

Table 1: Value Levels of the Design Variables for the Control Model in the Test of Varying the Number of Value Levels within the Design Variables

Variables	Level one	Level two	Level three
PK$$	$0	$3	$7
PKspace	0	1	
CPWT	0min.	10min.	20min.
GRHcp	0	1	
VPWT	5min.	25min.	35min.
GRHvp	0	1	

4.4.2 Findings: The Control Model

One hundred simulations were performed to obtain a central tendency and a variability of a distribution of sample means for each estimator in the control model.

Table 2: Descriptive Statistics for the Distribution of Sample Means of the Estimators in the Control Model in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
Medians	-0.16	0.55	-0.037	1.11	-0.049	1.16	0.32
Std.Error	0.002	0.009	0.003	0.01	0.001	0.01	0.01
Ave. % chg.*	0.01%	0.25%	8.6%	0.34%	0.8%	1.1%	5.6%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

Table 2 above shows that:

1. For the control model, there is no significant difference between the mean values of the distributions of sample means for all estimators and the parameters.

2. The median values are found within two standard errors of the parameters. Therefore, the central tendency of the distribution of the sample mean for every estimator is not affected by outliers.

3. The skewness values (see Table B-1 in Appendix-B) show that the distribution of sample means for GRHcp has a slight positive skewness. The distributions for PK$$, PKspace, GRHvp, and the constant for vanpool have slight negative skewnesses. The distribution of VPWT is more skewed to the right than other variables. The distribution of CPWT is more skewed to the left than other variables.

4. The kurtosis values (see Table B-1 in Appendix-B) indicate that the distributions of sample means for VPWT and CPWT are peaked while the distributions of the others are relatively normal.

4.4.3 Statistical Test: The Control Model

The following table compares the mean value of the distribution of the sample mean for each estimator in the control model to the respective parameters.

Table 3: Comparison of the Distribution of Sample Means for Estimators in the Control Model to the Parameters in the Test of Varying the Number of Value Levels within the Design Variable

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
t-ratios	0.01	0.15	0.917	0.38	0.340	1.03	1.58

None of the t-ratios is either greater than 1.96 or less than -1.96, the null hypothesis is not rejected. Therefore, the mean value of the distributions of the sample means for each design variable in the control model is not significantly different from the true parameter at 95% confidence level.

4.4.4 Test Model Design: Varying the Number of Value Levels for Design Variables

This test examines the effect on the estimated coefficients of the design variables due to various combinations of value levels among the six design variables in the test variations. The combinations are shown in the following table.

Table 4: Combinations of Value Levels for the Design Variables by Test Variations in the Test of Varying Number of Value Levels for Design Variables

Test Variations #	Number of 2-Level Variables	Number of 3-Level Variables	Number of 4-Level Variables
Test Variations			
1	3	2	1
2	3	1	2
3	3	0	3
Control Model	3	3	

4.4.5 Test Variation 1

4.4.5.1 Test Design

The first test variation design includes two 3-level variables, three 2-level variables and one 4-level variable. The values for each level of the six design variables are shown in the following table.

Table 5: Value Levels of the Design Variables for the Test Variation 1 in the Test of Varying the Number of Value Levels within the Design Variables

Variables	Level one	Level two	Level three	Level four
PK$$	$0	$3.00	$7.00	$12.00
PKspace	0	1		
CPWT	0min.	10min.	20min.	
GRHcp	0	1		
VPWT	5min.	25min.	35min.	
GRHvp	0	1		

4.4.5.2 Findings: Test Variation 1

Fifty runs were performed to obtain a central tendency and a variability of the distribution of the sample means for each design variable.

Table 6: Descriptive Statistics of the Distribution of Sample Means for the Estimators in Test Variation 1 in the Test of Varying the Number of Value Levels of the Design Variables Compared to the Control Model and Parameters

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
Ave. % chg.*	0.01%	0.25%	8.6%	0.34%	0.8%	1.1%	5.6%
Test #1							
Means	-0.16	0.54	-0.037	1.14	-0.065	1.14	0.28
Medians	-0.15	0.53	-0.036	1.15	-0.049	1.16	0.27
Std.Error	0.003	0.01	0.0008	0.01	0.012	0.02	0.02
Ave. % chg.*	2.03%	0.6%	0.73%	1.2%	35.5%	1.2%	2.1%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

Table 6 above shows that:

1. The difference between the mean value of the distribution of the sample means for VPWT changes from 0.8% under the control condition to 35.4% under the condition of test variation 1.

2. The mean values of the distributions of the sample means for all estimators are within two standard error around the parameters.

3. The median values of the distributions of sample means for all estimators are found within two standard errors of the parameters. Therefore, the central tendencies of the distributions of the sample means for all estimators are not affected by outliers;

However, the sample mean of VPWT is approaching the critical value while the median is essentially the same as the parameters.

4. Skewness values (see Table B-2 in Appendix-B) show that the distributions of sample means for PKspace, GRHcp, GRHvp and the constant in the utility model of vanpool have slight positive skewnesses. The distributions for PK$$ and CPWT have slight negative skewnesses. The distribution of the sample means for VPWT is more skewed to the left than other variables. This is also reflected by the average absolute change of 35.4% of the mean value from the true parameter. The median value also takes the position to the left of the parameter.

5. The kurtosis values (see Table B-2 in Appendix-B) indicate that the distribution of VPWT is peaked while the distributions of the others are relatively normal.

4.4.5.3 Statistical Test: Test Variation 1

Table 7 compares the mean values of the distributions of sample means for the estimators in the test variation 1 to the control model as well as parameters.

Table 7: Comparison of the Mean Values of the Distributions of Sample Means for all Estimators between Test Variation 1 and the Control Model as well as the Parameters in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const. for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
t-ratios	0.01	0.15	0.917	0.38	0.340	1.03	1.58
Test #1							
Means	-0.16	0.54	-0.037	1.14	-0.065	1.14	0.28
t-ratios	1.25	0.21	0.329	0.99	1.44	0.82	0.39

None of the t-ratios is either greater than 1.96 or less than -1.96. The null hypothesis is not rejected. Therefore, the mean value of the sample distribution for each design variable in the control model is not significantly different from the true parameter at 95% confidence level.

4.4.6 Test Variation 2

4.4.6.1 Test Design: Test Variation 2

Test variation 2 tests how well the parameters are recovered when two 4-level design variables are involved in the model. The value levels of the six design variables are shown in the following table.

Table 8: Value Levels of the Design Variables in the Test Variation 2 in the Test of Varying the Number of Value Levels within the Design Variables

Variables	Level one	Level two	Level three	Level four
PK$$	$0	$3.00	$7.00	$12.00
PKspace	0	1		
CPWT	0min.	10min.	20min.	30min.
GRHcp	0	1		
VPWT	5min.	25min.	35min.	
GRHvp	0	1		

4.4.6.2 Findings: Test Variation 2

Fifty runs were performed to obtain the central tendency of the distribution of sample means for each estimator.

Table 9: Descriptive Statistics of the Distribution of Sample Means for the Estimators in Test Variation 2 as well as the Control Model and Parameters in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
Ave. % chg.*	0.01%	0.25%	8.6%	0.34%	0.8%	1.1%	5.6%
Test #2							
Means	-0.16	0.53	-0.037	1.14	-0.057	1.14	0.29
Medians	-0.16	0.53	-0.038	1.14	-0.048	1.13	0.26
Std.Error	0.002	0.02	0.0006	0.02	0.008	0.02	0.02
Ave. % chg.*	1.3%	1.4%	1.2%	1.0%	17.9%	1.1%	1.6%

* Ave. % chg. is calculated from: |(Mean Value - Parameter) / Parameter|.

Table 9 shows that:

1. When the number of value levels of CPWT is increased to 4, the difference between the mean value of the distribution of sample means for VPWT and the parameter increases to 17.9% compared to the control model. However, compared to the control model, the differences between the mean value of the distributions of sample means for both CPWT and the constant in vanpool model decrease under the test variation 2.

2. The mean values of the distributions of sample means distributions for all estimators are within two standard error around true parameters.

3. The median values of the distributions of sample means for all estimators are found within two standard errors of the parameters. Therefore, the central tendency of the distribution of the sample means for each estimator is not affected by outliers.

4. Skewness values (see Table B-3 in Appendix-B) show that the distributions of sample means for PK$$ and GRHvp have slight positive skewnesses. The distributions for

PKspace, CPWT, GRHcp and the constant for vanpool have slight negative skewnesses. The distribution of VPWT is more skewed to the left than other variables. This is can be explained by the average percent change of 17.9% between the mean value of the distribution of sample means and the parameter.

5. Kurtosis values (see Table B-3 in Appendix-B) indicate that the distribution of sample means for VPWT is peaked while the distributions of the others are relatively normal.

4.4.6.3 Statistical Test: Test Variation 2

The following table compares the mean values of distributions of sample means for estimators in the test to the control model as well as the parameters.

Table 10: Comparison of the Mean Values of the Distributions of Sample Means for all Estimators between Test Variation 2 and the Control Model as well as the Parameters in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const. for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
t-ratios	0.01	0.15	0.917	0.38	0.340	1.03	1.58
Test #2							
Means	-0.16	0.53	-0.037	1.14	-0.057	1.14	0.29
t-ratios	0.86	0.50	0.700	0.76	1.003	0.80	0.28

None of the t-ratios is greater than 1.96 or less than -1.96, the null hypothesis is not rejected. Therefore, the mean values of distributions of the sample means for all estimators in the test variation 2 is not significantly different from the true parameter at 95% confidence level.

4.4.7 Test Variation 3

4.4.7.1 Test Design

Test variation 3 contains three 2-level design variables, three 4-level variables and one alternative-specific constant for vanpool. The value levels of the six design variables are shown in the following table.

Table 11: Value Levels of the Design Variables for the Test Variation 3 in the Test of Varying the Number of Value Levels within the Design Variables

Variables	Level one	Level two	Level three	Level four
PK$$	$0	$3.00	$7.00	$12.00
PKspace	0	1		
CPWT	0min.	10min.	20min.	30min.
GRHcp	0	1		
VPWT	5min.	25min.	35min.	45min.
GRHvp	0	1		

4.4.7.2 Findings: Test Variation 3

Fifty runs were performed to obtain a central tendency and the variability of the distribution of sample means for each estimator.

Table 12 below shows that:

1. When three variables of PK$$, CPWT, and VPWT are designed to have 4 levels, the difference between the mean value of the distribution of sample means for VPWT increases from 0.8% to 18.9% compared to the control model.

2. The mean values of distributions of sample means for all estimators except PK$$ in the model are within two standard errors around the parameters. The mean value of distribution of sample means for PK$$ in the model is within three standard errors of the parameters.

3. The median of the distributions of sample means for all estimators are found within two standard errors of the parameters. Therefore, the central tendency of the distribution of sample means for each estimator is not affected by outliers.

Table 12 also shows that:

Table 12: Descriptive Statistics of the Distribution of Sample Means for the Estimators in Test Variation 3 as well as the Control Model and Parameters in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
Ave. % chg.*	0.01%	0.25%	8.6%	0.34%	0.8%	1.1%	5.6%
Test #3							
Means	-0.17	0.54	-0.037	1.11	-0.057	1.11	0.31
Medians	-0.17	0.55	-0.037	1.12	-0.049	1.13	0.29
Std.Error	0.003	0.02	0.0007	0.012	0.008	0.02	0.02
Ave. % chg.*	4.3%	0.5%	0.3%	2.1%	18.9%	1.7%	6.2%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

4. Skewness values (see Table B-4 in Appendix-B) show that the distribution of sample means for the constant in vanpool model has slight positive skewness. The distributions for PKspace, PK$$, CPWT, GRHcp and the constant for vanpool have slight negative skewnesses. The distribution of VPWT is heavily skewed to the left. This can be

explained by the average percent change of 18.9% between the mean value of the distribution of sample means and the parameter for VPWT.

5. Kurtosis values (see Table B-4 in Appendix-B) indicate that the distribution of VPWT is peaked while the distributions of the others are relatively normal.

4.4.7.3 Statistical Test: Test Variation 3

The following table compares the mean values of the distributions of sample means for estimators in the test to the control model as well as the parameters.

Table 13: Comparison of the Mean Values of the Distributions of Sample Means for all Estimators between in the Test Variation 3 and the Control Model as well as the Parameters in the Test of Varying the Number of Value Levels of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const. for vanpool
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.16	0.54	-0.040	1.13	-0.048	1.14	0.31
t-ratios	0.01	0.15	0.917	0.38	0.340	1.03	1.58
Test #3							
Means	-0.17	0.54	-0.037	1.11	-0.057	1.11	0.31
t-ratios	2.42	0.17	0.144	2.03	1.07	0.97	1.14

Table 13 shows that the mean values of the distributions of sample means for PKspace, CPWT, VPWT, GRHvp, and the constant for vanpool are not significantly different from the parameters at 95% confidence level. However, for PK$$ and GRHcp, the null hypothesis is rejected. The mean values of the distributions of sample means for

PK$$ and GRHcp are significantly different from the parameters at 95% confidence level. However, the bias in both cases is under 5% as is shown in table 12.

4.4.8 Conclusions

When varying the number of the value levels built with SP experiments, the simulation exercise shows:

1. In two test variations where there is no significant difference between the mean value of the distribution of the sample means for estimators and their parameters. In all three test variations, the mean values for VPWT deviate from its parameter at 35.3%, 17.9%, and 18.9% respectively. However, no significant differences were found between them and the parameter.

2. When three design variables are assigned four value levels, the estimated coefficients of the estimators are slightly biased.

3. The bias that was recovered in the model was less than 5 percent of the parameter.

4.5 Test of Varying the Middle Placement within Design Variable

4.5.1 Control Model Design

The test designs to examine how well the parameters of the design variables can be recovered when the middle value of a design variable: PK$$, is varied. Preliminary tests showed that varying the middle value of a variable entered in a utility model in a linear form does not affect the recoverability of the parameter. The more useful test for placement of the middle value accrues where a curvilinear relationship is hypothesized. Therefore, PK$$ is entered in the utility function as a quadratic term instead of a linear term. In order to increase the slope of the curve, the parameter of the parking charge was changed from -0.16 to -0.21.

The parameterized model in the test of varying the middle placement within the design variable is constructed as:

$$V_{sov} = -0.21 * PK\$\$^2 + 0.54 * PKspace$$

$$V_{carpool} = -0.037 * CPWT + 1.13 * GRHcp$$

$$V_{vanpool} = -0.048 * VPWT + 1.13 * GRHvp + 0.29 * VPsub \qquad (4.5.1.1)$$

In equation 4.5.1.1, *PK$$* is the parking charge,

PKspace is the availability of parking space,

CPWT is the extra time over SOV when carpool is used,

GRHcp is the availability of Guaranteed Ride Home which requires 25 minutes waiting time for carpoolers,

VPWT is the extra time over SOV when vanpool is used,

GRHvp is the availability of Guaranteed Ride Home which requires 25 minutes waiting time for vanpoolers, and

VPsub is the subsidies for vanpoolers.

The values and the levels of the design variables are shown in the following table:

Table 14: Value Levels of the Design Variables for the Control Model in the Test of Varying the Middle Placement within the Design Variables

Variables	Level one	Level two	Level three
PK$$	$0	$3	$7
PKspace	0	1	
CPWT	0min.	10min.	20min.
GRHcp	0	1	
VPWT	5min.	25min.	35min.
GRHvp	0	1	
VPsub	0	$1	$3

4.5.2 Findings: the Control Model

Fifty runs were performed to obtain a central tendency and a variability of the distribution of the sample means for each design variable.

Table 15: Descriptive Statistics of the Distribution of Sample Means for the Design Variables in the Control Model in the Test of Varying the Middle Placement of the Design Variables

Variables	Parking charge (squared)	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Const.for vanpool
Parameters	-0.21	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.21	0.56	-0.037	1.14	-0.048	1.13	0.29
Medians	-0.22	0.56	-0.037	1.13	-0.048	1.11	0.29
Std.Error	0.003	0.02	0.0008	0.02	0.0007	0.01	0.008
Ave. % chg.*	2.2%	3.2%	1.1%	0.9%	0.7%	1.8E-05	0.5%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

Table 15 above shows that:

1. The differences between the mean values of the distributions of the sample means for all seven design variables range from lowest 1.8E-05 for VPWT to highest 3.2% for PKspace.

2. The mean values of the distributions of sample means for all seven design variables are within two standard errors around true parameters. The control model is unbiased.

3. The median values of the distributions of sample means for estimators except PK$$ are found within two standard errors of respective parameters. The median value for PK$$ (squared) is found within 3 standard errors of its parameter.

4. Skewness values (see Table B-5 in Appendix-B) show that the distributions of sample means of PKspace, GRHcp, GRHvp, and VPsub have slight positive skewnesses. The distributions for PK$$, CPWT, and VPWT have slight negative skewnesses.

4. Kurtosis values (see Table B-5 in Appendix-B) indicate that the distributions of the sample means for all seven design variables are normal.

4.5.3 Statistical Test: Control Model

The following table compares the mean values of the distributions of sample means for all seven design variables in the control model to the respective parameters.

Table 16: Comparison of the Mean Values of the Distributions of Sample Means for all Design Variables between the Control Model and the Parameters in the Test of Varying the Middle Placement of the Design Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.21	0.54	-0.037	1.13	-0.048	1.13	0.29
Control							
Means	-0.21	0.56	-0.037	1.14	-0.048	1.13	0.29
t-ratios	1.62	1.03	0.482	0.63	0.505	0.001	0.17

The null hypothesis is not rejected. The mean values of the distributions of sample means for all seven design variables are not significantly different from the parameters at 95% confidence level.

4.5.4 Test Model Design

The test examines how well the parameters of the seven design variables are recovered when the middle value of the quadratic term: PK$$, is changed from $3.00 to $1.00.

4.5.5 Findings: Test Model

Fifty runs were performed to obtain a central tendency and a variability of the distribution of sample means for each design variables.

Table 17: Descriptive Statistics of the Distributions of Sample Means for the Design Variables in the Test Model in the Test of Varying the Middle Placement of the Design Variables

Variables	Parking charge (squared)	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.21	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.21	0.56	-0.037	1.14	-0.048	1.13	0.29
Medians	-0.22	0.56	-0.037	1.13	-0.048	1.11	0.29
Std.Error	0.003	0.02	0.0008	0.02	0.0007	0.01	0.008
Ave. % chg.*	2.2%	3.2%	1.1%	0.9%	0.7%	1.8E-05	0.5%
Test Model							
Means	-0.21	0.55	-0.037	1.14	-0.048	1.12	0.30
Medians	-0.21	0.54	-0.037	1.15	-0.048	1.12	0.29
Std.Error	0.01	0.01	0.0007	0.01	0.0006	0.02	0.008
Ave. % chg.*	2.1%	1.1%	0.24%	1.2%	0.67%	0.8%	2.4%

* Ave. % chg. is calculated from: $\left|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}\right|$.

Table 17 above shows that:

1. The differences between the mean values of the distributions of sample means for all seven design variables and the parameters under the test condition do not vary significantly compared to the control model.

2. The mean values of the distributions of sample means for all seven design variables are within two standard errors around the parameters.

3. The median values of the distributions of sample means for all seven design variables are found within two standard error of the parameters. Therefore, the central tendency of the distribution of sample means for each design variable is not affected by outliers.

4. Skewness values (see Table B-6 in Appendix B-6) show that the distributions of sample means for PKspace, GRHcp, and VPsub have slight positive skewnesses. The distributions for PK$$, CPWT, GRHvp and VPWT have slight negative skewnesses.

4. Kurtosis values (see Table B-6 in Appendix B-6) indicate that the distributions of the sample means for all seven design variables are normal.

4.5.6 Statistical Test: Test Model

The following table compares the mean value of the distribution of sample means for each design variable in the test model to the control model as well as the parameters.

Table 18 below shows that the null hypothesis is not rejected. The mean values of the distributions of sample means for all seven design variables are not significantly different from the true parameters at 95% confidence level.

The null hypothesis is not rejected. The mean values of the distributions of sample means for all seven design variables are not significantly different from the parameters at 95% confidence level.

4.5.4 Test Model Design

The test examines how well the parameters of the seven design variables are recovered when the middle value of the quadratic term: PK$$, is changed from $3.00 to $1.00.

4.5.5 Findings: Test Model

Fifty runs were performed to obtain a central tendency and a variability of the distribution of sample means for each design variables.

Table 17: Descriptive Statistics of the Distributions of Sample Means for the Design Variables in the Test Model in the Test of Varying the Middle Placement of the Design Variables

Variables	Parking charge (squared)	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.21	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.21	0.56	-0.037	1.14	-0.048	1.13	0.29
Medians	-0.22	0.56	-0.037	1.13	-0.048	1.11	0.29
Std.Error	0.003	0.02	0.0008	0.02	0.0007	0.01	0.008
Ave. % chg.*	2.2%	3.2%	1.1%	0.9%	0.7%	1.8E-05	0.5%
Test Model							
Means	-0.21	0.55	-0.037	1.14	-0.048	1.12	0.30
Medians	-0.21	0.54	-0.037	1.15	-0.048	1.12	0.29
Std.Error	0.01	0.01	0.0007	0.01	0.0006	0.02	0.008
Ave. % chg.*	2.1%	1.1%	0.24%	1.2%	0.67%	0.8%	2.4%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

due to different perception and recording abilities among individuals. However, the mode choices respondents choose to use are based on the actual data instead of reported data. Therefore, the choice values found in an RP study are true values while the values for the independent variables are confounded with error. Do errors in the independent variables bias the estimates of the parameters for the design variables? In theory, errors in independent variables will bias the estimates of the variables downwards (see Chapter 3). Simulations in errors in variables seeks to demonstrate and quantify this hypothesis.

A model consisting of seven design variables with no errors is used as the control model for the test. In the test variations, one or two independent variables are measured with error in and these error-involved values will be entered into the ALOGIT program where all the other factors are the same as in the control model. The errors in the independent variables are assumed to have a normal distribution. Fifty runs are performed for each test variation in order to obtain a central tendency of the distribution of the sample means for each design variable. Sensitivity analysis is performed to examine the magnitude of deviation of the mean values of the distributions of the sample means for the design variables from their respective parameters when the percentage error in a design variable is increased.

4.6.2 Algorithm Used to Transfer the Error Term with Uniform Distribution to Weibull Distribution

Errors constructed into the variable VPWT are assumed to have a normal distribution with mean value being zero. The errors are constructed as follows:

$$Error = stdev * (\sum_{j=1}^{12} x_j - 6.0) \qquad (4.6.2.1)$$

In equation 4.6.2.1, x_j are uniformly distributed random numbers ranging from 0 to 1 and *stdev* is the standard deviation set equal to one.

4.6.3 Control Model

4.6.3.1 Control Model Design

The control model in this test is a 3 alternative logit model. It was chosen due to its similarity with stated choice and revealed preference studies currently used to forecast the effect of transportation demand management policies on APO at an employment site in New Jersey. The parameterized deterministic utility function for each alternative is the following:

$$V_{sov} = -0.16*PK\$\$+0.54*PKspace$$
$$V_{carpool} = -0.037*CPWT+1.13*GRHcp$$
$$V_{vanpool} = -0.048*VPWT+1.13*GRHvp+0.29*VPsub \qquad (4.6.3.1.1)$$

In equation 4.6.3.1.1, PK$$ is parking charge in dollars,

PKspace is availability of parking space,

CPWT is the extra time of using carpool over SOV,

GRHcp is availability of GRH which requires 25 minutes waiting time, VPWT is extra time of using vanpool over SOV,

GRHvp is availability of GRH which requires 25 minutes waiting time, and

VPsub is vanpool subsidy.

Variables PK$$, CPWT, VPWT, and VPsub are assigned three levels, while PKspace, GRHcp and GRHvp are dummy variables taking on two levels. The values assigned to each level of the design variables are listed as follows:

Table 19: Levels and Values of the Design Variables in the Control Model in the Test of Errors in Variables

Variables	Level one	Level two	Level three
PK$$	$0	$3	$7
PKspace	0	1	
CPWT	0min.	10min.	20min.
GRHcp	0	1	
VPWT	5min.	25min.	35min.
GRHvp	0	1	
VPsub	$0	$1	$3

4.6.3.2 Findings: The Control Model

One hundred runs were performed to obtain a central tendency of the distribution of the sample means for each design variable.

Table 20 below shows that:

1. In the control model, the mean values of the distributions of sample means for CPWT and VPWT increase about 20% compared to their parameters.

2. The mean values of the distributions of the sample means for all design variables are found within two standard errors of the parameters.

3. The median values of the distributions of the sample means are found within two standard errors of the parameters. Therefore, the central tendency of the distribution of the sample means for every estimator is not affected by outliers. However, in both variables of CPWT and VPWT where 20% deviations occur from their parameters, outliers seem as the cause.

In addition, table 20 also shows that:

Table 20: Descriptive Statistics of the Distributions of Sample Means for the Design Variables in the Control Model in the Test of Errors in Variables as well as Parameters

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%

* Ave. % chg. is calculated from: $\left|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}\right|$.

4. The skewness values (see Table B-7 in Appendix-b) show that the distributions of sample means for PKspace, GRHcp, and GRHvp have slightly positive skewnesses. The distribution PK$$ has a slight negative skewness. The distributions of CPWT and VPWT are skewed to the left, which can be seen from their mean values of the distributions of sample means. The distribution of VPsub is skewed to the right, which can partly be explained from its mean value.

5. The kurtosis values (see Table B-7 in Appendix-B) indicate that the distributions of CPWT, VPWT and VPsub are peaked while the distributions of the others are normal.

4.6.3.3 Statistical Test: The Control Model

The following table compares the mean values of the distributions of the sample means for all design variable in the control model to their parameters.

Table 21: Comparison of the Mean Values of the Distributions of Sample Means for all Design Variables in the Control Model and the Parameters in the Test of Errors in Variables

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65

None of the t-ratios is either greater than 1.96 or less than -1.96, the null hypothesis is not rejected. The mean value of the distribution of the sample means for each design variable in the control model is not significantly different from the parameter.

4.6.4 Test Variations Design

The combination of instrument design flaws and human perceptual characteristics lead to three broad categories of error among independent variables. Where uncertainty of an answer exists and pernicious behavior is not expected, reported values are distributed on either side of the true value. Where the human perceptual apparatus consistently overestimates or underestimates phenomenon then errors can be bunched on one side or the other of the true value but probably not on both sides. When applied to transport problems, Bruzelius (1979) comments: *"..it is widely believed that the motorist underestimates his true costs for a journey. This may be regarded as an instance of* *perceptual error since the underestimation my be due to the fact the driver perceives his costs, which are actually non-fixed, as fixed when deciding on a journey.*" Menon (1993) in a frequency behavior study found that respondents had a tendency to overreport the frequency of occurrence for irregular behavior. In the field of transportation, respondents are often asked " How many times have you been delayed to work on time during last week?". Occurrence of being delayed is an irregular behavior. Therefore, it is likely for respondents to overestimate the occurrence of being delayed.

The issues explores within errors in variables are limited to the three generalized classes of errors: errors randomly distributed around a true value, errors distributed on the positive side of a true value and errors distributed solely on the negative side of a true value.

In order to extend the errors in variables study to the broad issues surrounding logit models, three variations are explored. Initially, one design variable is used to explore the consequences of errors in variables across the three generalized classes of error. Next, two design variables in the same mode specific utility equation are injected with random error; finally, two design variables in two separate utility equations are injected with random error and the consequences for the parameter estimation process is examined.

The designed test variations in the test of errors in variables include:

1. Normally distributed errors in the variable of VPWT,
2. Errors which are skewed to the right of the true values of VPWT,
3. Errors which are skewed to the left of the true values of VPWT,
4. Normally distributed errors in both variables of VPWT and VPsub, and

5. Normally distributed errors in both variables of VPWT and CPWT.

The first test variation simulates a population where the model is given data on one variable, VPWT, that is measured with positive and negative random error. The second and the third variation simulate respondents in the sample who either over or under estimate the value of VPWT. The fourth test variation simulates error in two variables in a single equation; the respondents both over and under estimate the values of VPWT and VPsub. The last test variation examines the impact of reporting error in two variables placed in two separate equations; the test simulates respondents who over and under estimate the values of both VPWT and CPWT.

4.6.5 Hypotheses

Five tests are developed to examine the various aspects of error in variables of the logit model. The comparison is conducted among test variations 1, 2 and 3. In addition, the comparison among test variations 1, 4, and 5 is also performed. The results of the comparisons are hypothesized as follows:

1. The presence of errors in variables will bias and attenuate the estimated coefficients of the design variables in the overall model toward zero especially the estimates of design variables measured with error.

2. Based on the studies of Rosner et.al (1990) , the standard errors of the distributions of sample means for the variables measured with error are expected to attenuate toward zero.

3. The overall average percentage deviation of the mean values of the distributions of the sample means from the parameters over all seven design variables will increase as the average percentage error in the design variables increase.

4. Since the average percentage errors in the design variables among test variations 1, 2, and 3 are similar with each other, the effect of errors in variables under test conditions 1, 2, and 3 on the estimated coefficients of design variables will be similar.

5. Given that one design variable is constructed to be measured with errors, addition of one more independent variable measured with error in a same equation will aggravate the average percentage change of mean values from the true parameters for all design variables in the overall model.

6. Given one design variable measured with error, the addition of one more independent variable measured with error in a different equation will aggravate the average percentage change of mean values of the distributions of the sample means from the parameters for all design variables in the overall model.

4.6.6 Test Variation 1

4.6.6.1 Test Design

Test variation 1 tests the effect on the estimated coefficients of the design variables if respondents in the sample randomly over and underestimate the value of VPWT. The term VPWT has 3 given values in the fractional factorial design to compute the utility of using vanpool. In the test, two error levels of VPWT are designed as follows:

$$\text{Error level 1:}\quad VPWT_{new} = VPWT_{old} + 4.0 * (\sum_{j=1}^{12} x_j - 6.0)$$

$$\text{Error level 2:}\quad VPWT_{new} = VPWT_{old} + 12.0 * (\sum_{j=1}^{12} x_j - 6.0) \tag{4.6.6.1.1}$$

In equation 4.6.6.1.1, $VPWT_{new}$ represents the variable of VPWT whose values are measured with error,

$VPWT_{old}$ represents the variable of VPWT whose values are fixed, and

x_j is the uniform random variate between 0 and 1.

Comparison between true values and error-involved values of VPWT as well as the statistical descriptive report for error-involved values of VPWT are described in Table C-1 in Appendix-C.

4.6.6.2 Findings: Test Variation 1

Fifty runs were performed to obtain a central tendency and variability of the distribution of sample means for each design variable.

Table 22 below shows that:

1. At error level one, the mean values of the distributions of sample means for all design variables except PKspace and CPWT attenuate toward zero. At error level two, the mean values of the distributions of sample means for all design variables except PKspace attenuate toward zero.

Table 22: Descriptive Statistics of the distributions of sample means for the Design Variables and in Test Variation 1 compared to the Control Model as well as the Parameters in the Test of Errors in Variables+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%
Test Variation 1							
Error Level 1							
Means	-0.15	0.54	-0.040	1.12	*-0.044*	1.09	0.27
Medians	-0.15	0.53	-0.035	1.12	*-0.044*	1.08	0.27
Std.Error	0.003	0.02	0.005	0.01	*0.0005*	0.02	0.006
Ave. % chg.*	4.7%	0.5%	6.9%	0.9%	*8.4%*	3.8%	7.1%
Error Level 2							
Means	-0.12	0.55	-0.026	1.11	*-0.027*	0.89	0.17
Medians	-0.12	0.53	-0.028	1.11	*-0.028*	0.88	0.18
Std.Error	0.002	0.01	0.0007	0.01	*0.0004*	0.01	0.006
Ave. % chg.*	24.7%	2.4%	27.8%	1.7%	*42.8%*	21.5%	40.3%

* Ave. % chg. is calculated from: $\left|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}\right|$.

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

Table 22 above also shows that:

2. The deviation of the mean values of the distribution of the sample means for VPWT from its parameter increases when the percentage error in VPWT increases.

3. For both levels of VPWT, the median values of the distributions of sample means for all design variables are found within two standard errors of parameters. Therefore, the central tendency of the distribution of sample means for every estimator is not affected by outliers.

4. At error level 1 (see Table B-8-1 in Appendix-B), the distributions of PK$$, PKspace, and GRHvp have slight positive skewnesses; The distributions of CPWT, GRHcp, VPWT, and VPsub have slight negative skewnesses. At error level 2 (see Table B-8-2 in Appendix-B), the distributions of PKspace, CPWT, and VPWT have slight positive skewnesses; the distributions for PK$$, GRHcp, GRHvp, and VPsub have slight negative skewnesses.

5. At error level 1 (see Table B-8-1 in Appendix-B), the distribution of CPWT is peaked. At level 2 (see Table B-8-2 in Appendix-B), the distributions of all seven design variables are normal.

4.6.6.3 Statistical Test: Test Variation 1

The following table compares the mean values of the distributions of sample means for design variables in test variation 1 and the control model as well as the parameters.

Table 23 shows that at error level 1, the mean values of the distributions of sample means for PKspace, CPWT, and GRHcp are not significantly different from the parameters at 95% confidence level. At error level 2, only the mean values of the distributions of sample means for PKspace and GRHcp are not significantly different from the parameters at 95% confidence level. This suggests that with error built in VPWT, the mean values of the distributions of sample means for 4 more variables are biased compared to the control model. No variable in the control model is biased.

Table 23: Comparisons between the Mean Values of the Distributions of Sample Means for the Design Variables in Test Variation 1 and the Control Model as well as the Parameters+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65
Test Variation 1							
Error Level 1							
Means	-0.15	0.54	-0.040	1.12	*-0.044*	1.09	0.27
t-ratios	2.83	0.16	0.468	0.70	*8.426*	2.63	3.71
Error Level 2							
Means	-0.12	0.55	-0.027	1.11	*-0.027*	0.89	0.17
t-ratios	17.00	0.87	14.939	1.56	*52.651*	17.25	19.32

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

4.6.7 Test Variation 2

4.6.7.1 Test Design

Test variation 2 tests the effect on estimated coefficients of the design variables when all respondents overestimate the value of VPWT. Two error levels of VPWT are designed as follows:

$$\text{Error level 1:}\quad VPWT_{new} = VPWT_{old} + 4.0 * \left| \sum_{j=1}^{12} x_j - 6.0 \right|$$

$$\text{Error level 2:}\quad VPWT_{new} = VPWT_{old} + 12.0 * \left| \sum_{j=1}^{12} x_j - 6.0 \right| \qquad (4.6.7.1.1)$$

Comparison between true values and error-involved values of VPWT as well as the statistical descriptive report of error-involved values of VPWT are described in Table C-2 in Appendix-C.

4.6.7.2 Findings: Test Variation 2

Fifty runs were performed to obtain a central tendency and variability of the distribution of sample means for each design variable.

Table 24: Descriptive Statistics of the distributions of sample means for the Design Variables for under Control Model and Test Variation 2 as well as the Control Model and the Parameters in the Test of Errors in Variables+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%
Test Variation 2							
Error Level 1							
Means	-0.16	0.52	-0.037	1.09	*-0.043*	1.12	0.29
Medians	-0.16	0.51	-0.036	1.09	*-0.043*	1.11	0.29
Std.Error	0.003	0.01	0.0007	0.01	*0.0005*	0.02	0.006
Ave. % chg.*	3.0%	3.3%	0.9%	3.9%	*10.44%*	0.9%	1.5%
Error Level 2							
Means	-0.15	0.47	-0.035	1.04	*-0.033*	1.08	0.28
Medians	-0.15	0.47	-0.036	1.04	*-0.034*	1.05	0.28
Std.Error	0.003	0.02	0.0008	0.01	*0.0004*	0.02	0.007
Ave. % chg.*	5.1%	12.3%	4.20%	7.8%	*30.60%*	4.4%	2.3%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

Table 24 shows that:

1. At error level 1, the mean value of the distribution of sample means for VPsub increases by 1.5%; The mean values of the other six design variables attenuate toward zero. At error level 2, the mean values of the distributions of sample means for seven design variables attenuate toward zero.

2. Similar to the finding in test variation 1, the standard error of the distribution of sample means for VPWT attenuate toward zero as the percentage error in VPWT increases.

3. For both error levels of VPWT, the median values are found within two standard errors of the parameters for all design variables. Therefore, the central tendency of the distribution of sample means for every estimator is not affected by outliers.

4. At error level 1 (see Table B-9-1 in Appendix-B), the distributions of sample means for PK$$, PKspace, GRHcp, GRHvp, and VPsub have slight positive skewnesses; the distributions of CPWT and VPWT have slight negative skewnesses. At error level 2 (see Table B-9-2 in Appendix-B), the distributions of PK$$, GRHcp, VPWT, GRHvp, and VPsub have slight positive skewnesses; the distributions for PKspace and CPWT have slight negative skewnesses.

5. At error level 1 (see Table B-9-1 in Appendix-B), the distribution of sample means for PK$$ is peaked. At level 2 (see Table B-9-2 in Appendix-B), the distributions of sample means for all design variables except the variable of PKspace are normal.

4.6.7.3 Statistical Test: Test Variation 2

The following table compares the mean values of the distributions of sample means for all design variables in test variation 2 to the control model as well as the parameters.

Table 25: Comparisons between the Mean Values of the Distributions of Sample Means for the Design Variables in Test Variation 2 and the Control Model as well as the Parameters[+]

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65
Test Variation 2							
Error Level 1							
Means	-0.16	0.52	-0.037	1.09	*-0.043*	1.12	0.29
t-ratios	1.35	1.32	0.514	3.60	*9.118*	0.57	0.71
Error Level 2							
Means	-0.15	0.47	-0.035	1.04	*-0.033*	1.08	0.28
t-ratios	2.88	3.03	1.873	6.73	*29.45*	2.65	0.98

[+] Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

The above table shows that at error level 1, the mean values of the distributions of sample means for GRHcp and VPWT are significantly different from the parameters at 95% confidence level; at error level 2, only the mean values of the distributions of sample means for CPWT and VPsub are not significantly different from the parameters at 95% confidence level.

4.6.8 Test Variation 3

4.6.8.1 Test Design

Test variation 3 tests the effect on estimated coefficients of the design variables when all respondents underestimate the value of VPWT. Two levels of VPWT are designed as follows:

$$\text{Error level 1: } VPWT_{new} = VPWT_{old} + (-4.0) * \left| \sum_{j=1}^{12} x_j - 6.0 \right|$$

$$\text{Error level 2: } VPWT_{new} = VPWT_{old} + (-12.0) * \left| \sum_{j=1}^{12} x_j - 6.0 \right| \quad (4.6.8.1.1)$$

Comparison between the true values and error-involved values of VPWT as well as the statistical descriptive report for error-involved values of VPWT are described in Table C-3 in Appendix-C.

4.6.8.2 Findings: Test Variation 3

Fifty runs were performed in order to obtain a central tendency and variability of the distributions of sample means for the design variables.

Table 26 below shows that:

1. The test variation 3 is a unique case in terms that the average percentage deviations of the mean values of distribution of sample means for VPWT at both error levels are not the most severely affected by the errors in VPWT. The standard errors and

medians of the distributions of sample means in test variation 3 show that the mean values are not affected by the outliers. This suggests that the uniqueness of test variation was not due to some outliers in the distribution.

Table 26: Descriptive Statistics of the distributions of sample means for the Design Variables in the Test Variation 3 as well as Control Model and the Parameters in the Test of Errors in Variables+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%
Test Variation 3							
Error Level 1							
Means	-0.12	0.68	-0.018	0.97	*-0.043*	1.12	0.29
Medians	-0.12	0.68	-0.014	0.97	*-0.044*	1.12	0.29
Std.Error	0.003	0.01	0.002	0.01	*0.0006*	0.01	0.006
Ave. % chg.*	27.0%	26.4%	52.0%	13.8%	*9.9%*	0.5%	1.4%
Error Level 2							
Means	-0.12	0.60	-0.024	1.15	*-0.037*	0.85	0.16
Medians	-0.11	0.61	-0.024	1.13	*-0.037*	0.85	0.16
Std.Error	0.002	0.02	0.0007	0.01	*0.0004*	0.02	0.009
Ave. % chg.*	27.8%	10.5%	34.2%	1.9%	*22.7%*	25.2%	45.7%

* Ave. % chg. is calculated from: |(Mean Value - Parameter) / Parameter|.

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

In addition, table 26 above shows that:

2. At error level 1, the mean values of the distributions of sample means for PK$$, CPWT, GRHcp, VPWT, and GRHvp attenuate toward zero while the mean values of PKspace and VPsub increase about 26.4% and 1.4% respectively compared to the

parameter. At error level 2, the mean values of the distributions of sample means for all the design variables except PKspace and GRHcp attenuate toward zero. The mean values of the distributions of sample means for PKspace and GRHcp increase about 10.5% and 1.9% respectively compared to their parameters.

3. The standard error of the distribution of sample means for VPWT attenuate toward zero as the percentage error in VPWT increase.

4. For both error levels of VPWT, the median values are found within two standard errors of the parameters for all design variables. Therefore, the central tendency of the distribution of sample means for every estimator is not affected by outliers.

5. At error level 1 (see Table B-10-1 in Appendix-B), the distributions of sample means for PKspace, VPWT, GRHvp, and VPsub have slight positive skewnesses; the distributions of PK$$, CPWT and GRHcp have slight negative skewnesses. At error level 2 (see Table B-10-2 in Appendix-B), the distributions of PKspace, GRHcp, GRHvp, and VPsub have slight positive skewnesses; the distributions for PK$$, CPWT, and VPWT have slight negative skewnesses.

6. At error level 1 (see Table B-10-1 in Appendix-B), the distribution of sample means for CPWT is peaked. At error level 2 (see Table B-10-2 in Appendix-B), the distributions of sample means for all design variables except the variable of VPsub are normal.

4.6.8.3 Statistical Test: Test Variation 3

The following table compares the mean values of the distributions of the sample means for the design variables in test variation 3 to the control model as well as the parameters.

Table 27: Comparisons between the Mean Values of the Distributions of Sample Means for the Design Variables in Test Variation 3 and the Control Model as well as the Parameters+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65
Test Variation 3							
Error Level 1							
Means	-0.12	0.68	-0.018	0.97	*-0.043*	1.12	0.29
t-ratios	16.85	9.82	8.082	11.48	*7.286*	0.37	0.63
Error Level 2							
Means	-0.12	0.60	-0.024	1.15	*-0.037*	0.85	0.16
t-ratios	21.05	3.27	18.183	1.64	*24.342*	18.76	15.05

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

Table 27 shows that at error level 1, the mean values of the distributions of sample means for PK$$, PKspace, CPWT, GRHcp, VPWT are significantly different from the parameters at 95% confidence level. At error level 2, the mean values of the distributions of sample means for all design variables except GRHcp are significantly different from the parameter at 95% confidence level.

4.6.9 Test Variation 4

4.6.9.1 Test Design

The test tests the effect on estimated coefficients of design variables when the respondents both over and under estimate the values of VPWT and VPsub. Two error levels of VPWT and VPsub are designed as follows:

Error level 1: $VPWT_{new} = VPWT_{old} + 4.0 * (\sum_{j=1}^{12} x_j - 6.0)$

$$VPsub_{new} = VPsub_{old} + 0.5 * (\sum_{j=1}^{12} x_j - 6.0) \qquad (4.6.9.1.1)$$

Error level 2: $VPWT_{new} = VPWT_{old} + 12.0 * (\sum_{j=1}^{12} x_j - 6.0)$

$$VPsub_{new} = VPsub_{old} + 2.0 * (\sum_{j=1}^{12} x_j - 6.0) \qquad (4.6.9.1.2)$$

In equations 4.6.9.1.1 and 4.6.9.1.2, $VPsub_{new}$ represents the variable of VPsub whose value is measured with error, and

$VPsub_{old}$ represents the variables of VPsub whose value is measured with error.

Comparison between the true values and error-involved values of VPsub as well as the statistical descriptive report for error-involved values of VPsub are described in Table C-4 in Appendix-C.

4.6.9.2 Findings: Test Variation 4

Fifty runs were performed to obtain a central tendency and variability of the distributions of sample means for design variables.

Table 28: Descriptive Statistics of the distributions of sample means for the Design Variables in the Test Variation 4 as well as the Control Model the Parameters in the Test of Errors in Variables+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%
Test Variation 4							
Error Level 1							
Means	-0.16	0.58	-0.035	1.13	*-0.043*	1.09	*0.23*
Medians	-0.16	0.56	-0.035	1.12	*-0.043*	1.09	*0.23*
Std.Error	0.002	0.02	0.0007	0.01	*0.0004*	0.01	*0.006*
Ave. % chg.*	0.7%	6.5%	4.5%	0.03%	*9.9%*	3.5%	*2.2%*
Error Level 2							
Means	-0.13	0.54	-0.030	1.08	*-0.024*	0.96	*0.02*
Medians	-0.13	0.55	-0.030	1.07	*-0.024*	0.95	*0.01*
Std.Error	0.002	0.01	0.0008	0.01	*0.0004*	0.01	*0.003*
Ave. % chg.*	20.8%	0.7%	18.1%	4.2%	*49.3%*	14.9%	*93.6%*

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

Table 28 shows that:

1. At error level 1, the mean values of the distributions of sample means for all variables except PKspace attenuate toward zero; The mean values of PKspace increases 6.5% compared to the parameter. At error level 2, all the mean values of sample means

for all design variables except PKspace attenuate toward zero; The mean value of PKspace increases by 0.7% compared to the parameter.

2. The standard errors of the distributions of sample means for VPWT and VPsub attenuate toward zero as the percentage error in these two design variables increase.

3. For both error levels of VPWT and VPsub, the median values are found within two standard errors of the parameters for all design variables. Therefore, the central tendency of the distribution of sample means for each design variable is not affected by outliers.

4. At error level 1 (see Table B-11-1 in Appendix-B), the distributions of the sample means for PK$$, PKspace, GRHcp, VPWT, and GRHvp have slight positive skewnesses; the distributions of CPWT and VPsub have slight negative skewnesses. At error level 2 (see Table B-11-2 in Appendix-B), the distributions of PK$$, GRHcp, GRHvp, and VPsub have slight positive skewnesses; the distributions for PKspace, CPWT and VPWT have slight negative skewnesses.

5. At error level 1 (see Table B-11-1 in Appendix-B), the distribution of sample means for PKspace is peaked. At error level 2 (see Table B-11-2 in Appendix-B), the distributions of sample means for all design variables except CPWT are normal.

4.6.9.3 Statistical Test: Test Variation 4

The following table compares the mean values of the distributions of sample means for all design variables in test variation 4 to the control model as well as the parameters.

Table 29: Comparisons between the Mean Values of the Distributions of Sample Means for the Design Variables in Test Variation 4 and the Control Model as well as the Parameters+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65
Test Variation 4							
Error Level 1							
Means	-0.16	0.58	-0.035	1.13	*-0.043*	1.09	*0.23*
t-ratios	0.54	1.98	2.263	0.03	*12.323*	2.73	*10.88*
Error Level 2							
Means	-0.13	0.54	-0.030	1.08	*-0.024*	0.96	*0.02*
t-ratios	13.78	0.29	8.796	3.53	*56.79*	12.30	*85.27*

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

The above table shows that when at error level 1, only the mean values of the distributions of sample means for PK$$ and GRHcp are not significantly different from the parameters at 95% confidence level; while at error level 2, only the mean value of the distribution of sample means for PKspace is not significantly different from the parameters at 95% confidence level.

4.6.10 Test Variation 5

4.6.10.1 Test Design

The test tests the effect on estimated coefficients of design variables when the respondents both over and under estimate the values of VPWT and CPWT. Two error levels of VPWT and CPWT are designed as follows:

Error level 1: $VPWT_{new} = VPWT_{old} + 4.0 * (\sum_{j=1}^{12} x_j - 6.0)$

$$CPWT_{new} = CPWT_{old} + 2.0 * (\sum_{j=1}^{12} x_j - 6.0) \qquad (4.6.10.1.1)$$

Error level 2: $VPWT_{new} = VPWT_{old} + 12.0 * (\sum_{j=1}^{12} x_j - 6.0)$

$$CPWT_{new} = CPWT_{old} + 5.0 * (\sum_{j=1}^{12} x_j - 6.0) \qquad (4.6.10.1.2)$$

In equations 4.6.10.1 and 4.6.10.2, $CPWT_{new}$ represents the variable of CPWT whose value is measured with error, and

$CPWT_{old}$ represents the variable of CPWT whose value is measured with error.

Comparison between the true values and the error-involved values of CPWT as well as the statistical descriptive report of error-involved values of CPWT described in Table C-5 in Appendix-C.

4.6.10.2 Findings: Test Variation 5

Fifty runs were performed to obtain a central tendency and variability of the distribution of sample means for the design variables.

Table 30 below shows that:

1. For both error levels of CPWT and VPWT, the mean values of the distributions of sample means for all design variables except PKspace attenuate toward zero. Compared to the parameter, the mean value of the distribution of sample means for PKspace

increases by 1.3% at error level 1 and 4.3% at error level 2. The increases in PKspace at both error levels 1 and 2 are not significant.

Table 30: Comparison between the distributions of sample means for the Design Variables for under Control Model and Test Variation 5 as well as the Parameters in the Test of Errors in Variables+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Medians	-0.16	0.52	-0.038	1.15	-0.048	1.13	0.29
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave. % chg.*	0.8%	2.8%	20.0%	1.2%	20.00%	1.2%	2.3%
Test Variation 5							
Error Level 1							
Means	-0.15	0.55	*-0.033*	1.10	*-0.044*	1.09	0.27
Medians	-0.15	0.54	*-0.032*	1.09	*-0.043*	1.08	0.27
Std.Error	0.003	0.01	*0.0007*	0.01	*0.0005*	0.01	0.006
Ave. % chg.*	4.5%	1.3%	11.7%	2.7%	9.0%	3.3%	7.0%
Error Level 2							
Means	-0.12	0.56	*-0.02*	1.05	*-0.026*	0.90	0.17
Medians	-0.11	0.56	*-0.020*	1.05	*-0.026*	0.90	0.17
Std.Error	0.003	0.01	*0.0007*	0.01	*0.0004*	0.01	0.005
Ave. % chg.*	27.2%	4.3%	46.2%	7.4%	45.6%	20.7%	41.0%

* Ave. % chg. is calculated from: $|(\text{Mean Value} - \text{Parameter}) / \text{Parameter}|$.

\+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

Table 30 also shows that:

2. The standard errors of the distributions of sample means for CPWT and VPWT attenuate toward zero as the percentage error in VPWT and CPWT increases.

3. For both error levels of VPWT and CPWT, the median values of the distributions of sample means for all design variables are found within two standard errors

of the parameters. Therefore, the central tendency of the distribution of sample means for each design variable is not affected by outliers.

4. At error level 1 (see Table B-12-1 in Appendix-B), the distributions of sample means for PKspace, GRHcp, VPWT, GRHvp, and VPsub have slight positive skewnesses; the distributions of PK$$ and CPWT have slight negative skewnesses. At error level 2 (see Table B-12-2 in Appendix-B), the distributions of PKspace, CPWT, GRHvp, and VPsub have slight positive skewnesses; the distributions for PK$$, GRHcp, and VPWT have slight negative skewnesses.

5. For two error levels of VPWT and CPWT (see Tables B-12-1 and B-12-2 in Appendix-B), the distributions of sample means for all design variables are normal.

4.6.10.3 Statistical Test: Test Variation 5

The following table compares the mean values of the distributions of sample means for all design variables in test variation 4 to the control model as well as the parameters.

The table below shows that at both error level 1 and 2, only the mean values of the distributions of sample means for PKspace are not significantly different from the true parameters at 95% confidence level. It appears that the problem of errors in variables has a big impact on the mean values of distributions of sample means for most variables.

The following section performs a sensitivity analysis in order to examine the percentage change of the mean values of the variables due to one percent change of error in variable.

Table 31: Comparisons between the Mean Values of the Distributions of Sample Means for the Design Variables in Test Variation 5 and the Control Model as well as the Parameters+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
t-ratios	0.58	1.55	1.52	1.12	1.44	1.27	0.65
Test Variation 5							
Error Level 1							
Means	-0.15	0.55	*-0.033*	1.10	*-0.044*	1.09	0.27
t-ratios	2.84	0.46	*6.579*	2.60	*7.922*	2.54	3.15
Error Level 2							
Means	-0.12	0.56	*-0.020*	1.05	*-0.026*	0.90	0.17
t-ratios	15.60	1.65	*23.63*	6.97	*53.55*	15.86	21.69

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

4.6.11 Sensitivity Analysis

Sensitivity analysis quantify the impact of the errors in variables have on the mean values of the distributions of sample means for design variables when the percentage error in design variables is increased.

The following three tables present the average percentage change of errors in design variables, the resulting deviation of the mean values of the distributions of the sample means for design variables, and lastly the elasticity of the estimated coefficients of the design variables with respect to one percent change of error in the design variables.

From Table 32 below, the average percentage errors in VPWT increase from 29.0% at error level 1 to 88.0% at error level 2. The average percentage error in VPsub increase from 32.0% at error level 1 to 108.0% at error level 2. The average percentage error in CPWT increase from 16.0% at error level 1 to 40.0% at error level 2. Table 33 shows that as the average percentage errors in the design variables increase, the deviation of the mean values of the distributions of sample means for design variables measured with error from the parameters increase.

Table 32: Average Percentage Change of Errors in Design Variables at Error Level 1 and 2 for all Five Test Variations$^{\Delta}$

Variables with Error	Average Absolute Change*	Average % Change+
Error Level 1		
Test Variation 1		
VPWT	3.32min.	29.0%
Test Variation 2		
VPWT	3.32min.	29.0%
Test Variation 3		
VPWT	3.32min.	29.0%
Test Variation 4		
VPWT	3.32min.	29.0%
VPsub	$0.39	32.0%
Test Variation 5		
VPWT	3.32min.	29.0%
CPWT	1.60min.	16.0%
Error Level 2		
Test Variation 1		
VPWT	9.97min.	88.0%
Test Variation 2		
VPWT	9.97min.	88.0%
Test Variation 3		
VPWT	9.97min.	88.0%
Test Variation 4		
VPWT	9.97min.	88.0%
VPsub	$1.35	108.0%
Test Variation 5		
VPWT	9.97min.	88.0%
CPWT	4.00min.	40.0%

*Average absolute change is calculated from: $|(\text{error-involved values} - \text{true value})|$.

+Average % change is calculated from: $|(\text{error-involved value} - \text{true value}) / \text{true value}|$. For the average % change for VPsub and CPWT, the denominators are the average values of initial values of three levels for the design variable.

$^{\Delta}$The error-involved values of the design variable were taken from the output of the C program in which the seed values for uniformly distributed and normally distributed random numbers are 4 and 3 respectively (For detail, see Appendix-A).

Table 33: Comparison between the Mean Values of the Distributions of the Sample Means for all Design Variables in the Five Test Variations as well as the Control Model the Parameters+

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	VPWT wait-time	GRH for vanpool	Vanpool subsidy
Parameters	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Control Model							
Means	-0.16	0.52	-0.044	1.14	-0.057	1.14	0.30
Std.Error	0.002	0.01	0.005	0.01	0.007	0.01	0.01
Ave.%chg.*	0.8%	2.8%	20.0%	1.2%	20.0%	1.2%	2.3%
Test Model (Error Level 1)							
Test Variation 1							
Means	-0.15	0.54	-0.040	1.12	*-0.044*	1.09	0.27
Std.Error	0.003	0.02	0.005	0.01	*0.0005*	0.02	0.006
Ave.%chg.*	4.7%	0.5%	6.9%	0.9%	*8.4%*	3.8%	7.1%
Test Variation 2							
Means	-0.16	0.52	-0.037	1.09	*-0.043*	1.12	0.29
Std.Error	0.003	0.01	0.0007	0.01	*0.0005*	0.02	0.006
Ave.%chg.	3.0%	3.3%	0.9%	3.9%	*10.44%*	0.9%	1.5%
Test Variation 3							
Means	-0.12	0.68	-0.018	0.97	*-0.043*	1.12	0.29
Std.Error	0.003	0.01	0.002	0.01	*0.0006*	0.01	0.006
Ave.%chg.	27.0%	26.4%	52.0%	13.8%	*9.9%*	0.5%	1.4%
Test Variation 4							
Means	-0.16	0.58	-0.035	1.13	*-0.043*	1.09	*0.23*
Std.Error	0.002	0.02	0.0007	0.01	*0.0004*	0.01	*0.006*
Ave.%chg.	0.7%	6.5%	4.5%	0.03%	*9.9%*	3.5%	*2.2%*
Test Variation 5							
Means	-0.15	0.55	*-0.033*	1.10	*-0.044*	1.09	0.27
Std.Error	0.003	0.01	*0.0007*	0.01	*0.0005*	0.01	0.006
Ave.%chg.	4.5%	1.3%	*11.7%*	2.7%	*9.0%*	3.3%	7.0%
Test Model (Error Level 2)							
Test Variation 1							
Means	-0.12	0.55	-0.026	1.11	*-0.027*	0.89	0.17
Std.Error	0.002	0.01	0.0007	0.01	*0.0004*	0.01	0.006
Ave.%chg.	24.7%	2.4%	27.8%	1.7%	*42.8%*	21.5%	40.3%
Test Variation 2							
Means	-0.15	0.47	-0.035	1.04	*-0.033*	1.08	0.28
Std.Error	0.003	0.02	0.0008	0.01	*0.0004*	0.02	0.007
Ave.%chg.	5.1%	12.3%	4.2%	7.8%	*30.60%*	4.4%	2.3%
Test Variation 3							
Means	-0.12	0.60	-0.024	1.15	*-0.037*	0.85	0.16
Std.Error	0.002	0.02	0.0007	0.01	*0.0004*	0.02	0.009
Ave.%chg.	27.8%	10.5%	34.2%	1.9%	*22.7%*	25.2%	45.7%

Table 33: Comparison between the Mean Values of the Distributions of the Sample Means for all Design Variables in the Five Test Variations as well as the Control Model the Parameters+ (Continued)

Test Variation 4							
Means	-0.13	0.54	-0.030	1.08	*-0.024*	0.96	*0.02*
Std.Error	0.002	0.01	0.0008	0.01	*0.0004*	0.01	*0.003*
Ave.%chg.	20.8%	0.7%	18.1%	4.2%	*49.3%*	14.9%	*93.6%*
Test Variation 5							
Means	-0.12	0.56	*-0.020*	1.02	*-0.026*	0.90	0.17
Std.Error	0.003	0.01	*0.0007*	0.01	*0.0004*	0.01	0.005
Ave.%chg.	27.2%	4.3%	*46.2%*	7.4%	*45.6%*	20.7%	41.0%

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

* Ave. % chg. throughout the table is calculated from: |(Mean Value - Parameter) / Parameter|.

From table 33, we observe that:

1. The deviations of the mean values of the distributions of sample means from the respective parameters for all the design variables in test variations 1, 2, and 5 increase as the percentage error in design variables increase from error level 1 to error level 2.

2. In test variation 4, the mean values of the distributions of sample means for all design variables except PKspace from their parameters increase as the percentage error in design variables increase from error level 1 to error level 2. The deviation of mean value of the distribution of sample means for PKspace from its parameter decreases by about 5.8% as the percentage error in design variables increase from error level 1 to error level 2.

3. In test variation 3, the deviations of the mean values of the distributions of sample means for PKspace, CPWT and GRHcp from their parameters decrease as the percentage error in design variables increase. The deviations of mean values of

distributions of sample means for PK$$, VPWT, GRHvp, and VPsub increase as the average percentage error in design variables increases from error level 1 to error level 2.

4. For the design variables measured with error, the mean values of the distributions of sample means attenuate toward zero as the percentage error in design variables increase from error level 1 to error level 2. This supports hypothesis 1.

5. The standard errors of the distributions of sample means for the design variables measured with error attenuate toward zero. As the percentage error in the design variables increases, the deviation increases. This finding supports hypothesis 2.

The observations above confirm the hypothesis that the mean values of the distributions of sample means for most variables are affected when errors appear in the design variables. The following table examines the sensitivity of the mean values of the distribution of sample means for all design variables with respect to one percent change of the errors in the design variables. The table only deals with test variation 1, 2, and 3 in that each of which has only one design variable: VPWT, measured with errors. Both error levels are included.

From table 34 below, we observe that:

1. The elasticities of the mean values of the distributions of sample means for design variables are observed most stable in test variation 2 throughout the error levels 1 and 2. This shows a potential linear relationship for the relationship between percentage deviation of the mean values of the distributions of sample means for a design variable from its parameter and the percentage error in VPWT. The elasticities of the mean values

of the distributions of sample means for design variables in test variation 3 are observed most unstable throughout the error levels 1 and 2.

Table 34: Elasticity of the Mean Values of the Distributions of Sample Means for all Design Variables due to One Percent Change of Errors in VPWT at both Error Levels

Variables	Parking charge	Parking space	Carpool wait-time	GRH for carpool	Vanpool wait-time	GRH for vanpool	Vanpool subsidy
Error Level 1							
Test Variation 1							
Elasticities	0.16	0.02	0.24	0.03	*0.29*	0.13	0.25
Test Variation 2							
Elasticities	0.10	0.11	0.03	0.14	*0.36*	0.03	0.05
Test Variation 3							
Elasticities	0.93	0.91	1.79	0.47	*0.34*	0.02	0.05
Error Level 2							
Test Variation 1							
Elasticities	0.28	0.03	0.32	0.02	*0.49*	0.24	0.46
Test Variation 2							
Elasticities	0.06	0.14	0.05	0.09	*0.35*	0.05	0.03
Test Variation 3							
Elasticities	0.32	0.12	0.39	0.02	*0.26*	0.29	0.52

+ Italic values in the table represents that the corresponding design variable is constructed to be measured with error.

In addition, table 34 above shows that:

In test variation 2, for one unit change of errors in VPWT, the mean value of distribution of sample means change about 0.35 unit in a direction toward zero. However, the changes on mean values of distributions of sample means for other design variables due to one unit change of the errors in VPWT are rather minimal.

2. The elasticities of mean values of distributions of sample means for design variables except GRHcp in test variation 1 increase as the percentage error in VPWT increase. This implies a curvilinear relationship between the percentage difference of the

mean values of the distributions of sample means for a design variable from its parameter and the percentage error in VPWT.

In order to examine the impact on the estimation of the overall model due to one more design variable is measured with error, the average percentage difference of the mean values of the distributions of sample means over all seven design variables from their parameters is calculated, as shown in the following table.

Table 35: Average Percentage Deviation of Mean Values of Distributions of Sample Means over all Seven Design Variables in five Test Variations at both Error Levels

Error Levels	Error Level 1	Error Level 2
Overall Ave. % Chg.		
Test Variation 1	0.05%	0.23%
Test Variation 2	0.03	0.10
Test Variation 3	0.19	0.24
Test Variation 4	0.07	0.29
Test Variation 5	0.06	0.27

In the table 35 above, the overall Ave. % Chg. is calculated from:

$$\frac{1}{7}\sum_{j=1}^{7}\left|(\hat{\mu}_{mj}-\hat{\mu}_{pj})\,/\,\hat{\mu}_{pj}\right| \qquad (4.6.11.1)$$

In equation 4.6.11.1, j is the design variable in the overall model,

$\hat{\mu}_{mj}$ is the mean value of the distribution of sample means for jth design variable in the overall model, and

$\hat{\mu}_{pj}$ is the parameter of the jth design variable in the overall model.

Table 35 shows:

1. The overall average percentage deviation of the mean values of distributions of sample means over all seven design variables are not affected by adding one more design variable measured with error in the overall model. This finding does not support either hypothesis 5 or 6.

2. It is observed that the overall average percentage deviation of the mean values of distributions of sample means over all seven design variables in test variation 2 is lower than those in other test variations especially at error level 2. This finding does not support hypothesis 4.

3. The overall average percentage deviation of the mean values of distributions of sample means over all seven design variables increase as the percentage error in design variables increase from error level 1 to error level 2. This supports hypothesis 3.

4.7 Conclusions

Combining all the results from the comparison of the impact on the mean values of the distributions of sample means for design variables under control model and three test model designs as well as the parameters of the estimators in the simulation, it is concluded that under the condition of simulation:

1. The mean values of the distributions of sample means for the design variables are affected as the number of value levels of the design variables increases. However, the impact is not severe.

2. The mean values of the distributions of sample means for the design variables are not affected as the middle placement of level values within one design variable is varied.

3. Presence of errors in a single design variable affects most of the estimators significantly in the overall model.

4. With the presence of errors in a design variable, the mean value and the standard error of the distribution of sample means for that variable measured with error attenuate toward zero.

CHAPTER 5

APPLICATION

5.1 Introduction

This chapter applies the results of the simulations to an examination of two models. First, a model estimated using the RP approach is examined; this is followed with a discussion of a model using the SP approach. Section 5.2 describes the RP study known as the New Jersey Employee Commute Option (NJECO) program; emphasis is placed on the data generation process and the results of the parameter estimation process. Section 5.3 introduces the SP study performed on the employees of the Matsushita Electric Corporation of America's corporate headquarters site in Secaucus New Jersey. Section 5.4 briefs problems associated with NJECO and MECA. Section 5.5 concludes with a set of the consequences derived from the simulations in using SP and RP at a broader extent and suggestions for future research.

5.2 NJECO Model

5.2.1 General Background

The New Jersey's Employee Trip Reduction Program (ETRP) mandates every employer with more than 100 employees improve their Average Passenger Occupancy (APO) to an assigned target APO by November, 15, 1996. Questions on how much effect a specific Transportation Demand Management (TDM) policy will have on improving APO for a particular site are being asked by many affected employers. New Jersey DOT and COMSIS developed a model: NJECO, to help affected employers forecast the effect of a TDM strategy on improving the site's APO.

NJECO was estimated on a data base consisting of 2,437 employees working for 45 employers in southern California and the Sacramento metropolitan area (ETRP Calibration Report, 1994). Data generation for the RP study occurred through two surveys. The employer survey acquired information on the work location, and the type of TDM strategies that were offered at the site. The employee survey provided information on employee, his or her family, commuting trips, and the type of TDM offered to the employee. The two surveys were combined into a RP data base.

1. Merge files: Using the employee as the unit of observation, data for employers and employees were merged.

2. Traffic information among geographic zones: The employee's home and work locations were defined in terms of traffic analysis zones (TAZ). Information on employee's travel time and travel cost were obtained from several agencies in California.

3. Data revisions: Revisions to data were made and new variables were created such as auto operating costs. Data such as income were recoded as a binary variable. Daily parking cost was computed from an employee or an employer's response. Cost for vanpool and carpool were adjusted for the number of occupants in the vehicle.

4. Missing values: The whole record of an employee was dropped from the data base file when the missing value could not be replaced by the most frequently reported value or the variable's average value.

As a result of the preparation of data input into a database for mode choice model, 96 records out of 2,437 records were dropped.

Model estimation for the NJECO data uses the logit form of the disaggregate multinomial choice model:

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \tag{5.2.1.1}$$

Equation 5.2.1.1. states the individual's probability of choosing alternative i is the ratio of exponential of the systematic utility of selecting alternative i to the sum of the exponential of the utilities of choosing all other alternatives than alternative i. The estimation of the model was conducted by using ALOGIT program. The ALOGIT program uses Maximum Likelihood Estimation (MLE) to estimate the parameters of the design variables in the model. The mathematical derivation of the model and the Maximum Likelihood estimation have been discussed in Chapter 2.

The overall NJECO model includes independent variables in:

1. Transportation system variables: travel time, travel cost, parking cost,

2. Employee's socioeconomic variables: occupation, gender, age, household income, available vehicles, etc.,

3. Workplace variables: land use, parking spaces per employee, etc., and

4. TDM strategies.

Table 36: Final 5-mode Choice Model*

Variables	SOV	Carpool	Vanpool	Transit	Bike/Walk
Mode-specific constants		-1.517	-7.070	-3.048	-2.135
Transportation system variables					
In-vehicle time	-0.0399	-0.0399	-0.0399	-0.0110	
Out-of-vehicle time				-0.0165[a]	
Operating cost, fare	-0.0034	-0.0034	-0.0034	-0.0061	
Parking cost	-0.0086	-0.0086	-0.0086		
Bike lanes					1.220
Employee's socioeconomic variables					
Laborer?		0.3999			
Professional?		-0.2666	0.9054		
Manager?				-1.064	
Gender(1=male)					0.8727
Elderly?		0.5262	0.4355[b]	0.9089	
Midday business travel?		-0.7745			
Staggered work hours?				0.8148	
Part-time worker?				0.5377[b]	
1 worker/household?		-1.027			
Employee married?			0.9944		
Worksite variables					
Parking spaces/employee				-0.4155[b]	
SAC/Campus/Inst. LU?[d]		-0.8150			
No. of adjacent retail land uses		0.1069		0.1069	
TDM strategies					
Transportation coordinator and ridesharing matching program		0.0777[c]	0.0777[c]		
Preferential parking for ridesharers		0.1214[b]	0.1214[b]		
Transit info. center and bus pass sales			1.083		
Bike racks/showers/lockers					0.4056[b]
Guaranteed ride home		0.4476	0.4476	0.4476	0.4476
Modal subsidy		0.0125	0.0125		
Prizes, free meals, certificate[e]		0.0826	0.0826	0.0826	
Use of company vehicles by poolers		0.7861	0.7861		
Company-provided vans			2.586		

* Unless noted, all the coefficients are statistically significant at 95% confidence level.
? Variables shown with a question mark are binary variables, with value: 0=No, 1=Yes.
All time variables are in minutes, all cost variables are in cents (1992 dollars).
[a] Value constrained to equal 1.5 times the in-vehicle time coefficient.
[b] Coefficient value statistically significant at the 80% confidence level.
[c] Coefficient value not statistically significant at the 80% confidence level.
[d] Is work site a Suburban Activity Center, Campus, or Institutional land use?
[e] Coefficient value derived from other sources.

The final 5-mode model shown above was estimated with a ρ^2 of 0.183. The final mode choice model for 5 modes are listed in the following table.

5.2.2 Problems Associated with Data Generation in Estimating NJECO

As shown in chapter 2 and demonstrated in chapter 4, errors in variables lead to an attenuation towards zero for the parameter estimates of the independent variables. The data generation process incorporated in the NJECO model involves provides cases where errors in variables can exist.

In the data generation component for NJECO model, respondents were asked to report their actual travel behavior and socioeconomic characteristics. Recognition that reported actual travel behavior was not the same as actual behavior is of vital importance. Reported data of actual behavior can involve a large proportion of errors; size of error in the data depends on the nature of behavior, the type of data, and the variation in ability of respondents to report behavior that happened in the past. Data such as parking cost could be easily over or under reported.

Employee's travel cost was calculated by taking the distance and multiplying by 14 cents per mile. The distance traveled is represented by the centers of two geographical zones. This data will rarely reflect the actual commuting distance for an employee. Sometimes it will be far from the truth. For example, if the employee lives close to the boundary A of a geographic zone and works in another geographic zone but still close to boundary A, then the real distance between his or her home and work location is far different from the distance between two centers of two geographic zones. There are zones far off from the real distance. As a consequence, the distances for some employees between their home and work locations are overcounted while the distances for some employees are undercounted.

The typical travel time between all pairs of geographical zones is applied to every employee who lives in a geographical zone and works in another geographical zone. It is highly likely that the typical travel time between two geographical zones used in estimating NJECO involves a significant proportion of error.

From the results of the simulation in chapter 4, we know that the presence of one variable measured with error can disturb its parameter estimates as well as the estimates of other independent variables and attenuate their parameter estimate towards zero. This finding from the simulation suggests that most of the parameters in NJECO will be underestimated toward zero especially those variables measured with error. The degree of underestimation depends on the size of error. The simulation study suggests that the parameter estimates of travel time, travel cost, subsidy, parking cost, etc. are severely underestimated toward zero.

5.3 MECA Model

5.3.1 Introduction

The MECA model was estimated by Beaton, et. al. (1992) for the corporate headquarters of the Matsushita Electric Corporation of America (MECA) in the Hackensack Meadowlands of northern New Jersey. The model was developed to help MECA develop appropriate TDM strategies at the test site in order to improve APO from the current 1.08 to the target level of 1.73.

The design of the model on choice set, number of design variables, number of levels of the design variables, and values for each level of the design variables are determined after the discussion with focus groups. With little possibility to use public transportation and non-motorized modes to MECA, the stated choice model instrument was designed to support two mode choices: Single Occupancy Vehicle (SOV) and Ridesharers.

Data generation was completed using two surveys: an employee transportation survey and a stated choice experiment. The first survey collects information on employees' socioeconomic characteristics. The second survey contains two versions of a stated choice instrument in which high-level design variables were split into two different 3-level design variables and were presented to two groups of respondents. After the surveys were returned, these two versions were merged together to recover the high-level design variables.

In the structural model, 4 of 5 design variables are designed to have 5 levels and the other one is a binary variable. The following table displays the values and levels for each design variable.

Table 37: Values and Levels of the Design Variables for MECA Model

Variables	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Starting time(SOV)	8:00	8:30	9:00	10:00		
Parking charge (SOV)	$0.00	0.50	2.00	3.00	7.00	
Extra time to pick up rider (RS)	0.00min.	5.00	10.00	15.00	25.00	45.00
Guaranteed ride home program	1	0				
Rideshare coupon (RS)	$0.00	0.25	0.50	1.00	1.25	3.50

The model's structure is multinomial logit. This is essentially the same as in the case of the NJECO model. In the estimation process, ALOGIT program was also used for parameter estimation. The final model with 2-modes was expressed as follows:

Table 38: Final Model for MECA*

Attributes	SOV	Rideshare
Employee age	0.022	
	(3.17)**	
Unlikely to rideshare index	0.20	
	(3.5)	
Ridesharing is pleasant index	-0.14	
	(2.9)	
Trip length (natural log)	-0.35	
	(3.6)	
Drivers licenses per car	-0.50	
(clerical employee households)	(6.3)	
Date experiment held (week in 1991)	0.064	
	(4.9)	
Parking cost	-1.065	
	(9.4)	
Parking cost squared	0.074	
	(4.9)	
Flextime (early arrival in hours	0.31	
before 9:00 a.m.)	(1.4)	
Time lost ridesharing		-0.033
		(5.8)
Guaranteed Ride Home		1.33
		(9.0)
Rideshare coupon		0.85
		(3.0)
Rideshare coupon squared		-0.15
		(2.1)

* Analysis is based on 1,200 observations.
** "t" ratios displayed in parentheses. The flexible starting time estimators are not significant at the 0.05 level, however, they are reported to bring more information to bear on their use as TCM.

Attributes listed on the table below Parking cost are design variables. Information preceding Parking cost were collected via the first survey on employees' socioeconomic characteristics.

5.3.2 Problems Associated with MECA Model in terms of Data Generation and Design of Stated Choice

Evidence gained from the simulation studies showed that increasing the number of value levels can lead to biased estimates. Findings in the simulation on varying the number of value levels within the design variables suggests that the high-level design variables used in MECA to estimate the parameters could cause some estimates to be significantly different from their true parameters.

In the simulation on varying the middle placement within the design variables, I find that varying the middle placement within the quadratic form: PK$$, does not affect the ability of the program to recover the parameters of the design variable. This suggests no problems with the quadratic terms in MECA model which could be addressed by the results of simulation.

The values of the design variables are fixed, therefore, there is no errors in the design variables. However, problems with errors in variables still exist in MECA model. Data on employees' socioeconomic variables were reported by employees themselves. Thus, errors in these type of data are unavoidable. The results of the simulation on errors in variables in chapter 4 suggests that the presence of errors in one single variable could damage most of the estimates of the variables in the model; depending upon the level of error incorporated in the independent variable. Therefore, appearance of errors in the socioeconomic data could damage the estimates of the variables in the model to some extents.

Comparing the problems of errors in variables between MECA model and NJECO model, we expect that the consequences of errors in variables in NJECO are substantially more severe than those in MECA model. The reasons for the above statement are:

1. Fewer variables are measured with error in MECA than in NJECO. In NJECO model, all the data were either reported by employees and employers or were taken from some agencies. The critical transportation system variables in MECA are not measured with error. Data on every variable in NJECO could involve various proportion of errors. Overall, the size of errors in the independent variables in MECA is expected to be less than that in NJECO.

2. The variables measured with error in MECA are not design variables but socioeconomic variables. The design variables represent potential TDM policies to be implemented at the site in order to improve the Average Passenger Occupancy (APO) at the site. These design variables are being called upon to alter commuting behavior and are considered to have to a greater impact on improving APO than socioeconomic variables. This can be evidenced by comparing the coefficients of the design variables with the coefficients of the socioeconomic variables. The coefficients of the design variables are larger in absolute value than are the coefficients of the socioeconomic variables.

5.4 Application of findings to NJECO and MECA

Based on the results of the simulation study, the impact of TDM policies estimated by using NJECO model are underestimated due to errors in variables; While the impact of TDM policies by using MECA are likely to be overestimated. The degree of predicted underestimation of NJECO and overestimation of MECA can be demonstrated by using these two models in forecasting the improvement of APO at two sites in New Jersey due to change in TDM policies.

As partial confirmation of the coefficients of the parking charge design variable, comparison between two insurance companies in New Jersey suggests that a parking charge does help to improve the APO. Prudential and Mutual Benefit Life (MBL), have large office facilities in down town Newark New Jersey. Both facilities have good access to the city subway and comparable four block walk to the major transportation center in the city: Penn Station. Discussions with the Employee Transportation Coordinators at both sites suggests that the only difference in commuting conditions faced by employees is the cost of parking. Prudential Insurance provides all employees with fully subsidized parking; Mutual Benefit Life has for the past three years required a $2.00 per day parking charge for its on site parking. Alternative parking for MBL employees is available at a higher cost from nearby parking garages; no free parking exists that is within reasonable walking distance. The Compliance Plans show that Prudential Insurance has a current Average Passenger Occupancy level of 1.28; while, the MBL site's plan shows an APO of 1.58. Absent an in-depth disaggregate study at both sites, the initial conclusion is that parking charges do change commuting behavior. The APO for two sites identically placed and identically employed is over 20 percent higher where a \$2.00 per day parking charge is enforced.

Forecasting algorithms for both the NJECO and MECA models have been made available to the author. The impact of the implementation of a \$2.00 per day parking charge on APO will be forecast in both models. The current APO of Prudential is 1.28; it neither charges for SOV parking nor offers subsidies for ridesharers. On the other hand, MBL, has for three years charged employees a \$2.00 daily parking fee, their current APO is 1.58. Since the general background of Prudential and MBL are almost identical with each other, a \$2.00 daily parking charge imposed by Prudential should increase their APO to 1.58.

The NJECO model predicts an impact on APO derived from a \$2.00 parking charge from 1.28 to 1.41, this is a 57 % underestimation; the MECA model indicates under comparable conditions that the APO will change from 1.28 to 1.65, about 23% more than what it should have increased.

As we discussed in chapter 3, the scaling factor in the SP model comes into the forecasting component and will probably leads to an overestimation of parameters of the design variables.

The demonstration of the above example confirms the predicted underestimation of parameters in NJECO and overestimation of parameters in MECA. Furthermore, the degree of underestimation of NJECO shown in the above model almost doubles the degree of overestimation of the MECA model.

5.5 Conclusions

The purpose of my thesis is to examine sources of error within disaggregate discrete choice studies. Specifically, I have examined the variable design elements within the stated preference approach to discrete choice analysis and the errors in variables problem inherent to revealed preference studies. The conclusions relate solely to the estimation phase of a discrete choice analysis; forecasting issues, while briefly discussed, were not examined.

On the positive side, the logit model demonstrates its ability to produce consistent unbiased estimators under many conditions. Little in the way of problems in estimation occur due to the construction of stated preference design variables. Similarly, quadratic forms can be efficiently recovered using stated preference designs. However, a serious problem for discrete choice models in general has been found. Errors in variables will have major consequences for the estimation phase of discrete choice studies be the model either SP or RP. Errors in variables introduced into simulated discrete choice models drive parameter estimates toward zero in most cases. In all cases, the coefficient of the errors in variable will attenuated toward zero. The magnitude of the attenuation was measured through elasticity. Chapter 4 shows that the error elasticity ranged from 0.3 for low levels of error to 0.5 for high levels of error.

What are the consequences of these findings. First, the often stated assertion that Revealed Preference is the benchmark from which Stated Preference must be judged must be challenged. Neither approach is the absolute truth set from which the other can be measured. Second, in the estimation component, Stated Preference approach seems to provide stronger estimates than Revealed preference if the study of using SP is appropriately designed.

What is the next stage in this type of study? There are still many types of errors left unexamined which could occur in both RP and SP. The consequence of the different forms of specification error could be shown in future simulations. Bates (1988) stated that no problems in the estimation phase could be caused due to the response bias in SP study. A simulation of response bias can examine this statement.

The shape of the distribution of sample means near the tails of the distribution determines the probability of type I and type II error. Further understanding of these problems will require the number of simulation runs to be substantially increased. Additional research in errors in variables is needed. The tendency for the standard errors to attenuate toward zero undermines the hypothesis testing process and should be examined.

The issue of the scaling factor can also be brought to simulation. In Bates' view, it is pseudo-utility that is observed not true utility. The sum of the variances of the response bias and random utility is the variance used in estimation. While in forecasting, only the variance of the random utility term is used. The consequence of using different variances in estimation and forecasting is that the utility calculated in forecasting is biased. What is the magnitude of this bias? How to correct this bias? Knowledge of the answers to these two questions is lacking. A carefully-designed simulation can show the consequences of the scaling factor in the forecasting.

APPENDIX A

UNIX C PROGRAM FOR INPUT OF ALOGIT

Program A-1: Unix C Program for Input of Alogit Used in the Control Model

This program is written in Unix C and is used to generate inputs of the Alogit Program. The program performs the functions of generating random numbers with desired distribution, computing utilities for each alternative and assigning the mode choice with the highest utility to the choice variable. The output of the program is a data matrix containing 8 columns, of which one column is dependent variable and 7 columns are independent variables, and 800 observations. The program shown as follows is used in the control model.

```
#include <stdio.h>
#include <math.h>

#define RAND_MAX 2147483647
int whichmax(float,float,float);

main()
{
char inputfile[30],outputfile[30];
char iline[130];
long seed ;
int ii;
float a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w;
int na,nb,nc,nd,ne,nf,ng,nh;
FILE *fp1, *fp2;

printf("Please input the seed for random number generator (1-2147483647): ");
scanf("%ld",&seed);
```

```
for (ii=0;ii<30;++ii) {
    inputfile[ii] = '\0';
    outputfile[ii] = '\0';
}
printf("Please input file name : ");
scanf("%s",inputfile);
printf("Please output file name : ");
scanf("%s",outputfile);

if ((fp1 = fopen(inputfile,"r")) == NULL) {
     printf("inputfile does not exist!\n");
     return;
};
fp2 = fopen(outputfile,"w");

for (ii=0;ii<130;++ii)  iline[ii] = '\0';
fgets(iline,80,fp1);
for (ii=0;ii<130;++ii)  iline[ii] = '\0';
srandom(seed);
while(fgets(iline,80,fp1) != NULL) {
  sscanf(iline,"%f %f %f %f %f %f %f %f",&a,&b,&c,&d,&e,&f,&g,&h);

  i = (-1.0) * 0.16 * b + 0.54 * c;
  r = ((float) random())/ ((float) RAND_MAX);
  t = (-1.0) * log((-1.0)*log(r))/(1.28/0.7);
  l = i + t;

  j = (-0.037) * d + 1.13 * e;
  q = ((float ) random())/((float) RAND_MAX);
  u = (-1) * log(-(log(q)))/ (1.28/0.7);
  m = j + u;

  k = (-0.048) * f + 1.13 * g + 0.29 * h;
  p = ((float ) random())/((float) RAND_MAX);
  v = (-(log(-(log(p))))) / (1.28/0.7);
  n = k + v;
  printf("r%f,q%f,p%f",r,q,p);

  if ((na=whichmax(l,m,n)) == 0) {
     printf("error occur.\n");
     return;
  };
  printf("l%f, m%f, n%f, na%d", l,m,n,na);
```

```
  nc = (int) c;
  nd = (int) d;
  ne = (int) e;
  nf = (int) f;
  ng = (int) g;
  nh = (int) h;
  fprintf(fp2,"%d   %d   %d   %d   %d   %d   %d   %d\n",
        na,nb,nc,nd,ne,nf,ng,nh);
  for (ii=0;ii<130;++ii)  iline[ii] = '\0';
}
fclose(fp1);
fclose(fp2);

return;
}

int whichmax(float x, float y, float z)
{
 if ((x >= y) && ( x >= z)) return 1;
 if ((y >= x) && ( y >= z)) return 2;
 if ((z >= x) && ( z >= y)) return 3;
 return 0;
}
```

Program A-2: Unix C Program for Inputs of Alogit in the Test of Errors in Variables

This program basically performs the same function as it does in the other simualations than in the test of errors in variables. Program A-1 displays a program used in the control model and other simulations than in the test of errors in variables with minor modifications. In the test of errors in variables, another error term with normal distribution is added to the independent variables in addition to the regression error term in the utility function. The original program was modified by adding lines to generate the errors with normal distribution at the end. The program shown as follows assumes the all the respondents in the sample over estimate extra time over single driving when vanpool is used.

```
#include <stdio.h>
#include <math.h>

#define RAND_MAX  2147483647
int whichmax(float,float,float);
float normal_rand();

main()
{
char inputfile[30],outputfile[30];
char iline[130];
long ranseed, seed ;
int ii;
float a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,nf;
int na,nb,nc,ne,ng,nh,nd;
FILE *fp1, *fp2;

printf("Please input the seed for random number generator (1-2147483647): ");
scanf("%ld",&seed);
printf("Please input the ranseed for normal distribution (1-2147483647): ");
scanf("%ld",&ranseed);
for (ii=0;ii<30;++ii) {
   inputfile[ii] = '\0';
   outputfile[ii] = '\0';
}
printf("Please input file name : ");
scanf("%s",inputfile);
printf("Please output file name : ");
scanf("%s",outputfile);

if ((fp1 = fopen(inputfile,"r")) == NULL) {
    printf("inputfile does not exist!\n");
    return;
};
fp2 = fopen(outputfile,"w");

for (ii=0;ii<130;++ii)  iline[ii] = '\0';
fgets(iline,80,fp1);
for (ii=0;ii<130;++ii)  iline[ii] = '\0';
srandom(seed);
```

```
srand(ranseed);
while(fgets(iline,80,fp1) != NULL) {
 sscanf(iline,"%f %f %f %f %f %f %f %f",&a,&b,&c,&d,&e,&f,&g,&h);

 i = (-1.0) * 0.16 * b + 0.54 * c;
 r = ((float) random())/ ((float) RAND_MAX);
 t = (-1.0) * log((-1.0)*log(r))/(1.28/0.7);
 l = i + t;

 j = (-0.037) * d + 1.13 * e;
 q = ((float ) random())/((float) RAND_MAX);
 u = (-1) * log(-(log(q)))/ (1.28/0.7);
 m = j + u;

 k = (-0.048) * f + 1.13 * g + 0.29 * h;
 p = ((float ) random())/((float) RAND_MAX);
 v = (-(log(-(log(p))))) / (1.28/0.7);
 n = k + v;

 if ((na=whichmax(l,m,n)) == 0) {
    printf("error occur.\n");
    return;
 };

 nf = f + 4 * fabs(normal_rand(1.0));
 nb = (int) b;
 nd = (int) d;
 nc = (int) c;
 ne = (int) e;
 ng = (int) g;
 nh = (int) h;
 fprintf(fp2,"%d   %d   %d   %d   %d   %f   %d   %d\n",
        na,nb,nc,nd,ne,nf,ng,nh);
 for (ii=0;ii<130;++ii) iline[ii] = '\0';
}
fclose(fp1);
fclose(fp2);

return;
}

int whichmax(float x, float y, float z)
{
 if ((x >= y) && ( x >= z)) return 1;
 if ((y >= x) && ( y >= z)) return 2;
```

```
 if ((z >= x) && ( z >= y)) return 3;
 return 0;
}

/************************************************************************/

float normal_rand(ro)
     float ro;

    {int i, j;
     float x;

     x=0.0;

   for (i=1;i<=12; i++)
         x += rand()/2147483647.0;
         x = ro*(x-6.0);
         return (x);
    }
/************************************************************************/
```

APPENDIX B

PARAMETERS AND STATISTICAL SUMMARY REPORT DERIVED FROM SIMULATIONS

Table B-1: Parameters and Descriptive Statistics obtained from 100 Simulations for 6 Designed Variables and one Alternative-specific Constant in a 3 Alternative Mode Choice Study for the control model in the Test of Varying the Number of Levels within Design Variables

Variables	Pkspace	Pk $$	CPWT	GRHcp	VPWT	GRHvp	Const.vp
Parameter	0.54	-0.16	-0.037	1.13	-0.048	1.13	0.29
Mode	0.6064	-0.1465	-0.0370	0.9557	#N/A	1.2630	0.3192
Mean	0.5414	-0.1600	-0.0402	1.1261	-0.0476	1.1422	0.3062
Median	0.5473	-0.1590	-0.0367	1.1137	-0.0490	1.1617	0.3154
Std.Error	0.0093	0.0016	0.0035	0.0102	0.0011	0.0118	0.0102
Std. Devi.	0.0932	0.0156	0.0347	0.1018	0.0112	0.1176	0.1023
Variance	0.0087	0.0002	0.0012	0.0104	0.0001	0.0138	0.0105
Skewness	-0.3224	-0.3170	-9.703	0.54	7.8934	-0.4421	-0.3453
Kurtosis	0.2644	-0.0143	95.9949	1.11	72.50	-0.4838	-0.1521
Range	0.4784	0.0789	0.3540	0.6036	0.1127	0.4994	0.4870
Minimum	0.2818	-0.202	-0.3799	0.9071	-0.0578	0.8705	0.0416
Maximum	0.7603	-0.1231	-0.0260	1.5107	0.0549	1.3700	0.5285
Count	100	100	100	100	100	100	100

Table B-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 6 Designed Variables and one Alternative-specific Constant in a 3 Alternative Mode Choice Study for Test Variation 1* in the Test of Varying the Number of Levels within the Design Variables

Variables	Pk space	Pk $$$	CPWT	GRHcp	VPWT	GRHvp	Const.vp
Parameter	0.54	-0.16	-0.037	1.13	-0.048	1.13	0.29
Mode	#N/A	#N/A	#N/A	#N/A	#N/A	1.1533	#N/A
Mean	0.5369	-0.1568	-0.0373	1.144	-0.065	1.1438	0.2840
Median	0.5250	-0.1522	-0.0363	1.1475	-0.049	1.155	0.2653
Std.Error	0.0147	0.0026	0.0008	0.0135	0.0117	0.0167	0.0155
Std. Devi.	0.1036	0.0182	0.0057	0.0952	0.0827	0.1178	0.1099
Variance	0.011	0.0003	3.3E-05	0.0091	0.0068	0.0139	0.0121
Skewness	0.2362	-0.8413	-0.1828	0.0183	-4.871	0.2950	0.2124
Kurtosis	-0.8066	0.1633	-0.7606	-0.069	22.889	-0.1188	-0.514
Range	0.4296	0.0768	0.0239	0.4432	0.4585	0.5289	0.4333
Minimum	0.3432	-0.2037	-0.050	0.9382	-0.4974	0.9404	0.0729
Maximum	0.7728	-0.1269	-0.02612	1.3814	-0.0389	1.4693	0.5062
Count	50	50	50	50	50	50	50

*The first test variation design includes two 3-level variables, three 2-level variables and one 4-level variable. PK$$ is the only 4-level design variable whose values at different levels are 0, 3, 7 and 12. Fifty runs have been applied to obtain the mean values of sampling distributions for design variables.

Table B-3: Parameters and Descriptive Statistics obtained from 50 Simulations for 6 Designed Variables and one Alternative-specific Constant in a 3 Alternative Mode Choice Study for Test Variation 2* in the Test of Varying the Number of Levels within the Design Variables

Variables	Pk space	Pk $$$	CPWT	GRHcp	VPWT	GRHvp	Const.vp
Parameter	0.54	-0.16	-0.037	1.13	-0.048	1.13	0.29
Mode	0.614	-0.1796	#N/A	1.2837	#N/A	#N/A	#N/A
Mean	0.5323	-0.1621	-0.0374	1.1417	-0.0566	1.1428	0.2852
Median	0.5279	-0.1633	-0.0377	1.1432	-0.0482	1.1317	0.2637
Std.Error	0.0154	0.0025	0.00061	0.0153	0.0085	0.0158	0.0167
Std. Devi.	0.1086	0.0174	0.0043	0.1078	0.0599	0.1119	0.1180
Variance	0.0118	0.0003	1.85E-05	0.0116	0.0036	0.0125	0.0139
Skewness	-0.389	0.083	-0.1908	-0.0950	-6.999	0.0807	-0.1328
Kurtosis	0.9353	-0.1753	0.03749	0.0358	49.3132	-0.0346	0.3577
Range	0.5733	0.0843	0.020	0.5349	0.433	0.5403	0.5718
Minimum	0.188	-0.2053	-0.0484	0.8476	-0.47	0.8814	-0.0465
Maximum	0.7614	-0.1211	-0.0285	1.3825	-0.037	1.4218	0.5253
Count	50	50	50	50	50	50	50

*Test variation 2 tests how well the parameters are recovered when two 4-level design variables are involved in the model. PK$$ still keeps the 4-level values and CPWT is increased from 3-level to 4-level. The values for carpool wait-time are 0, 10, 20, 30. The values for PK$$ are 0, 3, 7, and 12. Fifty runs were performed to derive the mean values of the distributions of sample means for the design variables.

Table B-4: Parameters and Descriptive Statistics obtained from 50 Simulations for 6 Designed Variables and one Alternative-specific Constant in a 3 Alternative Mode Choice Study for Test Variation 3* in the Test of Varying the Number of Levels within the Design Variables

Variables	Pk space	Pk $$$	CPWT	GRHcp	VPWT	GRHvp	Const.vp
Parameter	0.54	-0.16	-0.037	1.13	-0.048	1.13	0.29
Mode	0.5463	#N/A	-0.0375	1.1467	#N/A	1.1773	#N/A
Mean	0.5428	-0.1668	-0.0369	1.106	-0.0570	1.111	0.308
Median	0.5466	-0.1673	-0.0366	1.1191	-0.0492	1.127	0.2925
Std.Error	0.01657	0.0028	0.0007	0.0117	0.0083	0.0198	0.0156
Std. Devi.	0.1171	0.0197	0.0049	0.0829	0.0586	0.140	0.1103
Variance	0.0137	0.0004	2.38E-05	0.0069	0.0034	0.0195	0.0121
Skewness	-0.1905	-1.058	-0.7519	-0.1421	-6.978	-0.388	0.7093
Kurtosis	1.063	3.927	1.057	-0.2711	49.1041	-0.3298	1.077
Range	0.6370	0.1152	0.0234	0.3728	0.4218	0.6026	0.5541
Minimum	0.2368	-0.2449	-0.0519	0.9104	-0.4616	0.7707	0.0992
Maximum	0.8738	-0.1297	-0.0285	1.2831	-0.0398	1.3732	0.6533
Count	50	50	50	50	50	50	50

*Test variation 3 contains three 2-level design variables, three 4-level variables and one alternative-specific constant for vanpool. Parking charge, carpool wait-time and vanpool wait-time are three 4-level variables. The values for CPWT are 0, 10, 20, 30. The values for PK$$ are 0, 3, 7, and 12. The values for VPWT are 5, 25, 35, and 45. Fifty runs were performed to obtain the mean values of the distributions of sample means distributions of six design variables and one alternative-specific constant.

Table B-5: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for the Control Model in the Test of Varying the Middle Placement of Variable of Parking Charge

Variables	Pk space	Pk $$$ (squared)	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	0.54	-0.21	-0.037	1.13	-0.048	1.13	0.29
Mode	0.5944	-0.2474	#N/A	#N/A	-0.045	1.096	#N/A
Mean	0.5572	-0.2145	-0.0374	1.140	-0.0483	1.13	0.289
Median	0.5572	-0.2151	-0.0374	1.132	-0.0481	1.112	0.2859
Std.Error	0.0165	0.0028	0.00084	0.01556	0.00068	0.0146	0.0079
Std. Devi.	0.1166	0.0195	0.0060	0.11	0.0048	0.1031	0.0559
Variance	0.0136	0.00038	3.55E-05	0.012	2.28E-05	0.011	0.0031
Skewness	0.2772	-0.32	-0.1684	0.1056	-0.523	0.4882	0.0606
Kurtosis	-0.6113	-0.2742	-0.4049	1.5427	0.6978	0.9140	-0.0675
Range	0.4492	0.0814	0.0265	0.6282	0.0251	0.5436	0.2576
Minimum	0.3630	-0.2544	-0.052	0.8127	-0.0624	0.912	0.1494
Maximum	0.8121	-0.1730	-0.025	1.44	-0.0373	1.456	0.407
Count	50	50	50	50	50	50	50

* PK$$ is entered in the utility function as a quadratic term instead of a linear term. The parameter of the PK$$ was changed from -0.16 to -0.21. The control model was designed to maintain all the current level values of all seven design variables constant. The test model was designed to change the middle value of PK$$ from $3.00 to $1.00.

Table B-6: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for the Test Model in the Test of Varying the Middle Placement of Variable of Parking Charge

Variables	Pk space	Pk $$$	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	0.54	-0.21	-0.037	1.13	-0.048	1.13	0.29
Mode	#N/A	#N/A	#N/A	1.0364	-0.0478	1.194	#N/A
Mean	0.5461	-0.2056	-0.0369	1.1440	-0.0483	1.1211	0.297
Median	0.5424	-0.2093	-0.0367	1.1527	-0.0483	1.1208	0.2926
Std.Error	0.0115	0.0122	0.0007	0.0137	0.0006	0.0166	0.0083
Std. Devi.	0.0815	0.0862	0.005	0.0968	0.0043	0.117	0.059
Variance	0.0067	0.0074	2.5E-05	0.0094	1.85E-05	0.0137	0.0035
Skewness	0.0874	-0.5982	-0.3065	0.1276	-0.5917	-0.179	0.5646
Kurtosis	-0.5602	0.2712	-0.039	-0.4502	0.6427	-0.3809	0.9128
Range	0.3187	0.3817	0.0241	0.4153	0.020	0.5054	0.2842
Minimum	0.4028	-0.4643	-0.050	0.9556	-0.0603	0.846	0.1816
Maximum	0.7215	-0.0826	-0.026	1.371	-0.040	1.3514	0.4659
Count	50	50	50	50	50	50	50

*The test examines how well the parameters of the seven design variables are recovered when the middle value of the quadratic term: PK$$, is changed from $3.00 to $1.00. Fifty runs were performed to obtain the mean values of the distributions of sample means for seven design variables.

Table B-7: Parameters and Descriptive Statistics obtained from 100 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study in the Test of Errors in Variables

Variables	Pk $$$	Pk space	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mode	-0.1665	0.5774	-0.03756	1.1232	-0.0486	1.1396	0.2865
Mean	-0.1613	0.5246	-0.04427	1.1432	-0.0575	1.144	0.2966
Median	-0.1611	0.5241	-0.0375	1.150	-0.0481	1.1287	0.2918
Std.Error	0.00216	0.00985	0.00475	0.01168	0.00658	0.01098	0.01012
Std. Devi.	0.0216	0.09846	0.04753	0.1168	0.0658	0.1098	0.10123
Variance	0.00047	0.09694	0.00226	0.0136	0.00433	0.01206	0.01025
Skewness	-0.2290	0.5927	-7.581	0.2633	-7.067	0.4954	6.700
Kurtosis	1.8529	1.3428	60.3683	0.4572	49.8592	0.1578	58.673
Range	0.1491	0.5846	0.4271	0.6375	0.5284	0.5185	1.0073
Minimum	-0.2421	0.3050	-0.4521	0.8820	-0.5665	0.9213	0.1704
Maximum	-0.093	0.8896	-0.025	1.52	-0.0381	1.44	1.178
Count	100	100	100	100	100	100	100

Table B-8-1: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 1* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.152	0.54	-0.039	1.12	-0.044	1.086	0.27
Mode	-0.137	0.574	#N/A	#N/A	-0.0435	1.043	#N/A
Median	-0.15	0.53	-0.035	1.12	-0.044	1.073	0.27
Std.Error	0.0026	0.015	0.0054	0.0145	0.00047	0.016	0.0055
Std. Devi.	0.0185	0.109	0.038	0.103	0.0033	0.115	0.039
Variance	0.0003	0.012	0.0014	0.010	1.11E-05	0.013	0.0015
Skewness	0.0011	0.57	-6.86	-0.023	-0.15	0.36	-0.18
Kurtosis	-1.24	-0.12	47.98	0.06	1.37	0.097	0.413
Range	0.062	0.45	0.27	0.48	0.018	0.54	0.18
Minimum	-0.184	0.37	-0.30	0.89	-0.05	0.87	0.17
Maximum	-0.122	0.82	-0.023	1.36	-0.035	1.42	0.35
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly over and under estimate the values of VPWT. The mean value and the variance of the error term attached to the variable of VPWT are zero and 16 respectively.

Table B-8-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 1* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1205	0.5529	-0.0267	1.1106	-0.0275	0.8868	0.1732
Mode	-0.1464	#N/A	#N/A	1.043	#N/A	0.8263	#N/A
Median	-0.1197	0.5274	-0.028	1.111	-0.0277	0.8806	0.1762
Std.Error	0.0023	0.0146	0.0007	0.0123	0.00039	0.0140	0.0060
Std. Devi.	0.0163	0.1033	0.0048	0.087	0.0027	0.0987	0.0423
Variance	0.00027	0.0107	2.32E-05	0.0075	7.46E-06	0.0097	0.00179
Skewness	-0.1946	0.0016	0.4959	-0.1107	1.000	-0.0379	-0.0923
Kurtosis	-0.7546	0.0613	0.1581	-0.6705	2.0702	-0.1436	0.1126
Range	0.0651	0.4972	0.0225	0.3553	0.0149	0.4344	0.1942
Minimum	-0.1552	0.2631	-0.0356	0.9087	-0.0326	0.6397	0.0783
Maximum	-0.0902	0.7603	-0.0132	1.264	-0.0178	1.074	0.2725
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly overestimate and underestimate the values of VPWT. The mean value of the error term attached to the variable of VPWT is zero while the variance of the error term changes to 144.

Table B-9-1: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 2* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.155	0.522	-0.037	1.086	-0.043	1.12	0.29
Mode	-0.159	#N/A	#N/A	1.07	#N/A	1.056	0.32
Median	-0.159	0.51	-0.036	1.086	-0.043	1.11	0.29
Std.Error	0.0035	0.013	0.00067	0.012	0.0005	0.017	0.006
Std. Devi.	0.025	0.094	0.005	0.086	0.004	0.122	0.043
Variance	0.0006	0.0088	2.26E-05	0.0075	1.48E-05	0.015	0.0018
Skewness	3.5	0.21	-0.50	0.11	-0.05	0.25	0.19
Kurtosis	18.74	-0.56	-0.16	0.155	0.20	-0.89	-0.19
Range	0.17	0.39	0.022	0.43	0.018	0.51	0.19
Minimum	-0.19	0.33	-0.05	0.87	-0.05	0.90	0.20
Maximum	-0.019	0.72	-0.028	1.30	-0.034	1.41	0.39
Count	50	50	50	50	50	50	50

*All the respondents are assumed to overestimate the extra time over single occupancy driving when vanpool is used. The mean value of the error term attached to the variable of VPWT is zero and the variance of the error term is 16.

Table B-9-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 2* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1519	0.4735	-0.035	1.042	-0.033	1.08	0.283
Mode	-0.152	#N/A	#N/A	1.02	#N/A	0.963	#N/A
Median	-0.152	0.4675	-0.0359	1.037	-0.034	1.049	0.28
Std.Error	0.0028	0.022	0.0008	0.0129	0.0005	0.018	0.0067
Std. Devi.	0.0198	0.154	0.0058	0.0913	0.0035	0.13	0.047
Variance	0.0004	0.0236	3.38E-05	0.0083	1.22E-05	0.0169	0.0022
Skewness	0.2005	-3.27	-0.0177	0.067	0.359	0.334	0.45
Kurtosis	-0.188	17.69	-0.63	-0.2002	-0.565	-0.162	-0.432
Range	0.08	1.066	0.025	0.389	0.0147	0.604	0.176
Minimum	-0.189	-0.3649	-0.049	0.844	-0.04	0.816	0.205
Maximum	-0.109	0.7013	-0.024	1.233	-0.0258	1.42	0.38
Count	50	50	50	50	50	50	50

*All the respondents are assumed to overestimate the extra time over single occupancy driving when vanpool is used. The mean value of the error term attached to the variable of VPWT is zero and the variance of the error term is 144.

Table B-10-1: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 3* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1168	0.682759	-0.01775	0.974395	-0.04327	1.124465	0.293971
Mode	-0.11423	0.636389	-0.01085	1.044638	-0.03994	1.182722	0.356181
Median	-0.11606	0.680052	-0.01422	0.9685	-0.04352	1.118319	0.285693
Std.Error	0.002538	0.014394	0.002358	0.013419	0.000643	0.014687	0.006266
Std. Devi.	0.017949	0.101781	0.016673	0.09489	0.004545	0.103853	0.044304
Variance	0.000322	0.010359	0.000278	0.009004	2.07E-05	0.010786	0.001963
Skewness	-0.12359	0.043763	-2.38673	-0.40257	0.090383	0.094722	0.132461
Kurtosis	-0.14851	-0.68711	8.231057	0.461032	-0.76877	0.779748	-0.43132
Range	0.08165	0.412343	0.09188	0.477564	0.017198	0.558887	0.194955
Minimum	-0.15997	0.486024	-0.09377	0.692059	-0.05153	0.849792	0.1943
Maximum	-0.07832	0.898367	-0.00189	1.169623	-0.03433	1.408678	0.389256
Count	50	50	50	50	50	50	50

*The respondents are assumed to underestimate the extra time over single occupancy vehicle when vanpool is used and vanpool subsidy. The mean value and the variance of the error term attached to the variable of VPWT are zero and 16 respectively.

Table B-10-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 3* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1155	0.5965	-0.0244	1.1516	-0.037	0.8456	0.1574
Mode	-0.1027	0.668	#N/A	1.073	-0.0376	0.9289	#N/A
Median	-0.1129	0.6107	-0.0243	1.131	-0.0375	0.855	0.1577
Std.Error	0.0021	0.0171	0.0007	0.013	0.0004	0.015	0.0087
Std. Devi.	0.0148	0.121	0.0049	0.0924	0.0031	0.106	0.0617
Variance	0.0002	0.0146	2.37E-05	0.0085	9.84E-06	0.011	0.0038
Skewness	-0.574	0.2524	-0.0319	0.7142	-0.18	0.2107	3.9
Kurtosis	0.5915	-0.102	-0.2146	0.124	-0.5383	-0.3259	22.48
Range	0.0732	0.577	0.0214	0.395	0.0125	0.453	0.438
Minimum	-0.1567	0.3617	-0.0355	1.0026	-0.043	0.6511	0.074
Maximum	-0.0835	0.9388	-0.014	1.39	-0.031	1.104	0.5125
Count	50	50	50	50	50	50	50

*All the respondents are assumed to underestimate the extra time over single occupancy driving when vanpool is used. The mean value of the error term attached to the variable of VPWT is zero and the variance of the error term is 144.

Table B-11-1: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 4* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.158	0.57	-0.035	1.13	-0.043	1.09	0.23
Mode	-0.17	0.645	#N/A	1.033	#N/A	1.02	0.23
Median	-0.16	0.56	-0.035	1.124	-0.043	1.093	0.23
Std.Error	0.0022	0.018	0.0007	0.013	0.00038	0.014	0.0057
Std. Devi.	0.0159	0.125	0.005	0.095	0.0027	0.102	0.04
Variance	0.00025	0.0156	2.67E-05	0.089	7.22E-06	0.010	0.0016
Skewness	0.886	2.519	-0.73	0.144	0.27	0.388	-0.314
Kurtosis	1.79	12.55	0.56	1.72	0.62	0.114	-0.338
Range	0.085	0.84	0.025	0.58	0.013	0.47	0.168
Minimum	-0.195	0.36	-0.051	0.85	-0.0487	0.886	0.134
Maximum	-0.11	1.205	-0.026	1.43	-0.035	1.35	0.302
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly over and under estimate the extra time over single occupancy driving when vanpool is used and vanpool subsidy. The mean value and the variance of the error term attached to the variable of VPWT are zero and 16 respectively. The mean value and the variance of the variable of VPsub are zero and 0.25 respectively.

Table B-11-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 4* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1267	0.544	-0.030	1.082	-0.024	0.9616	0.0185
Mode	#N/A	0.5938	-0.032	1.05	#N/A	0.994	#N/A
Median	-0.1285	0.5478	-0.03	1.073	-0.024	0.953	0.0137
Std.Error	0.0024	0.0138	0.00075	0.0134	0.0004	0.0136	0.003
Std. Devi.	0.0169	0.097	0.0053	0.095	0.0029	0.096	0.022
Variance	0.00029	0.0095	2.83E-05	0.009	8.53E-06	0.0092	0.0005
Skewness	0.3242	-0.65	-1.347	0.1785	-0.56	0.211	0.3239
Kurtosis	-0.6053	0.42	5.55	-0.5368	-0.192	-0.135	-0.738
Range	0.0644	0.4313	0.033	0.4131	0.0112	0.434	0.092
Minimum	-0.1564	0.2815	-0.053	0.878	-0.031	0.7368	-0.021
Maximum	-0.092	0.713	-0.0195	1.29	-0.0196	1.17	0.07
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly over and under estimate the extra time over single occupancy vehicle when vanpool is used and vanpool subsidy. The mean value and the variance of the error term attached to the variable of VPWT are zero and 144 respectively. The mean value and the variance of the error term attached to the variable of VPsub are zero and 4 respectively.

Table B-12-1: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 5* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1528	0.5468	-0.0327	1.099	-0.044	1.092	0.2698
Mode	-0.1373	#N/A	#N/A	1.133	-0.039	1.08	0.2286
Median	-0.1525	0.5405	-0.032	1.088	-0.043	1.08	0.265
Std.Error	0.0025	0.0145	0.00065	0.0117	0.0005	0.0147	0.0063
Std. Devi.	0.0178	0.1023	0.0046	0.083	0.0038	0.104	0.045
Variance	0.00032	0.0105	2.12E-05	0.00689	1.47E-05	0.0108	0.0020
Skewness	-0.051	0.16	-1.36	0.625	0.083	0.27	0.093
Kurtosis	0.313	-0.395	2.878	0.086	0.022	-0.07	-0.337
Range	0.077	0.436	0.023	0.357	0.018	0.475	0.21
Minimum	-0.199	0.328	-0.048	0.966	-0.05	0.85	0.166
Maximum	-0.12	0.76	-0.025	1.32	-0.033	1.32	0.38
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly over and under estimate the extra time over single occupancy driving when vanpool and carpool are used. The mean value and the variance of the error term attached to the variable of VPWT are zero and 16, respectively. The mean value and the variance of the error term attached to the variable of CPWT are zero and 4, respectively.

Table B-12-2: Parameters and Descriptive Statistics obtained from 50 Simulations for 7 Designed Variables in a 3 Alternative Mode Choice Study for Test Variation 5* in the Test of Errors in Variables

Variables	PK$$	Pkspace	CPWT	GRHcp	VPWT	GRHvp	VPsub
Parameter	-0.16	0.54	-0.037	1.13	-0.048	1.13	0.29
Mean	-0.1164	0.5634	-0.02	1.046	-0.026	0.896	0.171
Mode	-0.1319	0.585	#N/A	#N/A	#N/A	0.96	0.195
Median	-0.1134	0.5638	-0.02	1.05	-0.026	0.8986	0.169
Std.Error	0.0028	0.014	0.0007	0.012	0.0004	0.0146	0.0054
Std. Devi.	0.02	0.099	0.005	0.084	0.0029	0.103	0.038
Variance	0.0004	0.0098	2.56E-05	0.007	8.18E-06	0.01	0.0015
Skewness	-0.21	0.159	0.097	-0.31	-0.34	0.0156	0.558
Kurtosis	-0.4345	1.929	-0.126	0.156	-0.504	-0.2672	1.0874
Range	0.089	0.605	0.0243	0.39	0.0117	0.446	0.19
Minimum	-0.158	0.26	-0.032	0.825	-0.032	0.685	0.103
Maximum	-0.069	0.865	-0.0076	1.216	-0.02	1.13	0.294
Count	50	50	50	50	50	50	50

*The respondents are assumed to randomly over and estimate the extra time over single occupancy driving when vanpool is used. In addition, the respondents randomly over and estimated the extra time over single occupancy driving when carpool is used. The mean value of the error term attached to the variable of VPWT are zero and 144 respectively. The mean value and variance of the error term attached to the variable of CPWT are zero and 25.

APPENDIX C

SUMMARY REPORTS FOR CHARACTERISTICS OF ERROR-INVOLVED VALUES OF DESIGN VARIABLES IN TEST OF ERRORS IN VARIABLES

Table C-1: Summary Report for Error-Involved Values of VPWT when all Respondents in the Sample randomly under or overestimate the value of VPWT (Seed values of 4 and 3 are used for Uniformly and Normally Distributed Random Numbers Respectively)

Variance	16	16	16	144	144	144
True Value	5	25	35	5	25	35
Mean	5.16	24.75	35.11	22.28	22.69	22.09
Std.Error	0.28	0.21	0.30	0.82	0.59	0.79
Median	5.40	24.76	35.21	24.43	24.97	24.49
Std.Devi.	3.95	4.13	4.18	11.66	11.76	11.24
Variance	15.61	17.08	17.54	135.91	138.25	126.24
Kurtosis	-0.52	-0.29	-0.41	-0.73	-0.70	-0.74
Skewness	-0.14	-0.06	-0.102	-0.46	-0.49	-0.54
Range	18.97	20.90	19.47	47.45	49.94	43.73
Minimum	-5.48	14.61	24.95	-3.20	-5.48	-1.84
Maximum	13.50	35.51	44.46	44.25	44.46	41.90
Sum	1031.4	9898.62	7021.97	4455.83	9078.00	4418.17
Count	2200	400	200	200	400	200
Ave.abs. Chg.	3.25	3.33	3.38	3.25	3.51	3.02
Ave.% Chg. *	0.65	0.13	0.096	0.25	0.27	0.23

* Average % change is calculated by $(VPWT_{new} - VPWT_{old}) / VPWT_{old}$.

Table C-2: Summary Report for Error-Involved Values of VPWT when all Respondents in the Sample overestimate the value of VPWT (Seed values of 4 and 3 are used for Uniformly and Normally Distributed Random Numbers Respectively)

Variance	16	16	16	144	144	144
True Value	5	25	35	5	25	35
Mean	8.25	28.33	38.38	14.76	35.01	45.14
Std.Error	0.16	0.12	0.17	0.47	0.37	0.52
Median	7.78	27.87	37.75	13.35	33.63	43.24
Std.Devi.	2.24	2.44	2.46	6.71	7.33	7.39
Variance	5.00	5.97	6.07	45.00	53.71	54.62
Kurtosis	-0.11	0.049	-0.35	-0.11	0.049	-0.35
Skewness	0.65	0.80	0.70	0.65	0.80	0.70
Range	10.44	10.50	10.00	31.32	31.51	30.01
Minimum	5.04	25.00	35.00	5.11	25.01	35.03
Maximum	15.48	35.51	45.01	36.43	56.53	65.04
Sum	1650.7	11335.51	7676.08	2952.13	14006.53	9028.25
Count	200	400	200	200	400	200
Ave.abs. Chg.	3.25	3.33	3.38	9.76	10.01	10.14
Ave.% Chg. *	0.65	0.13	0.096	1.95	0.40	0.29

* Average % change is calculated by $(VPWT_{new} - VPWT_{old}) / VPWT_{old}$.

Table C-3: Summary Report for Error-Involved Values of VPWT when all Respondents in the Sample underestimate the value of VPWT (Seed values of 4 and 3 are used for Uniformly and Normally Distributed Random Numbers Respectively)

Variance	16	16	16	144	144	144
True Value	5	25	35	5	25	35
Mean	1.75	21.67	31.62	-4.76	14.98	24.86
Std.Error	0.16	0.12	0.17	0.47	0.37	0.52
Median	2.22	22.12	32.25	-3.35	16.37	26.76
Std.Devi.	2.24	2.44	2.46	6.7	7.33	7.39
Variance	5.00	5.97	6.07	45.00	53.70	54.62
Kurtosis	-0.11	0.049	-0.35	-0.11	0.048	-0.35
Skewness	-0.65	-0.80	-0.70	-0.65	-0.80	-0.70
Range	10.44	10.50	10.00	31.32	31.51	30.01
Minimum	-5.48	14.49	24.98	-26.42	-6.53	4.96
Maximum	4.96	25.00	35.00	4.89	24.98	34.97
Sum	349.29	8664.49	6323.92	-952.13	5993.47	4971.76
Count	200	400	200	200	400	200
Ave.abs. Chg.	3.25	3.39	3.38	9.76	10.01	10.14
Ave.% Chg. *	8.08	0.17	0.11	3.52	3.99	0.63

* Average % change is calculated by $(VPWT_{new} - VPWT_{old}) / VPWT_{old}$.

Table C-4: Summary Report for Error-Involved Values of VPsub when all Respondents in the Sample underestimate the value of VPsub (Seed values of 4 and 3 are used for Uniformly and Normally Distributed Random Numbers Respectively)

Variance	0.25	0.25	0.25	4	4	4
True Value	0	1	3	0	1	3
Mean	0.40	1.00	2.99	2.03	2.10	2.06
Std.Error	0.02	0.024	0.035	0.12	0.075	0.11
Median	0.33	1.00	2.97	1.69	1.87	1.78
Std.Devi.	0.29	0.49	0.50	1.63	1.50	1.54
Variance	0.08	0.24	0.25	2.67	2.27	2.37
Kurtosis	0.09	-0.45	-0.23	1.77	0.17	-0.23
Skewness	0.82	0.13	0.04	1.27	0.82	0.72
Range	1.30	2.48	2.58	8.25	7.41	6.93
Minimum	0.5E-05	0.0013	1.73	0.009	0.22E-04	0.001
Maximum	1.30	2.48	4.31	8.26	7.41	6.94
Sum	79.69	400.29	597.03	405.86	838.50	412.30
Count	200	400	200	200	400	200
Ave.abs. Chg.	0.40	0.40	0.40	1.32	1.34	1.38
Ave.% Chg.*	0.32	0.32	0.32	1.06	1.07	1.11

* Average % change is calculated by $(VPsub_{new} - VPsub_{old}) / 1.25$, where 1.25 is the average value of the original set of values for CPWT.

Table C-5: Summary Report for Error-Involved Values of CPWT when all Respondents in the Sample underestimate the value of CPWT (Seed values of 4 and 3 are used for Uniformly and Normally Distributed Random Numbers Respectively)

Variance	4	4	4	25	25	25
True Value	0	10	20	0	10	20
Mean	10.02	10.03	9.83	-0.65	10.15	20.11
Std.Error	0.54	0.37	0.52	0.35	0.25	0.36
Median	10.02	10.40	9.67	-0.79	10.19	19.98
Std.Devi.	7.60	7.44	7.39	4.91	5.04	5.10
Variance	57.75	55.37	54.55	24.12	25.40	25.97
Kurtosis	-0.86	-0.90	-0.77	-0.13	-0.44	0.02
Skewness	0.031	-0.11	0.08	0.037	-0.076	0.066
Range	29.63	28.72	31.01	25.40	27.21	27.83
Minimum	-5.00	-4.6	-5.07	-12.68	-4.07	7.02
Maximum	24.62	24.11	25.94	12.72	23.13	34.85
Sum	2003.63	4010.78	1966.6	-129.36	4059.27	4022.62
Count	200	400	200	200	400	200
Ave.abs. Chg.	1.58	1.64	1.57	3.94	4.11	3.94
Ave.% Chg.*	0.16	0.16	0.16	0.39	0.42	0.39

* Average % change is calculated by $(CPWT_{new} - CPWT_{old}) / 10$, where 10 is the average value of the original set of values for CPWT.

BIBLIOGRAPHY

ALOGIT Users' Guide, prepared by Hague Consulting Group, September, 1992.

Bates, John, "Econometric Issues in Stated Preference Analysis," *Journal of Transport Economics and Policy,* vol. 22, no. 1, pp. 59-69, 1988.

Beaton, W. P., Meghdir, H., and Carragher, F. J., "Assessing the Effectiveness of Transportation Control Measures: Use of Stated Preference Models to Project Mode Split for Work Trips," *Transportation Research Record 1346,* 1992.

Bonsall, P.W., "Transfer Price Data - its Use and Abuse." *Transportation Planning Method.* proceedings from PTRC conference, 1983.

Button, Kenneth J., *Transportation Economics.* Edward Elgar Publishing Limited, 1993.

Bradley, M.A. and Kroes, E.P., "Simulaneous Analysis of Stated Preference and Revealed Preference Information," Paper for the PTRC 18th Summer Annual Meeting Seminar on Transportation Planning Methods, University of Sussex, England, 1990.

Clean Air Act, as amended by Pub.L.101-549 (42 U.S.C. (7401 et seq.)).

Employee Trip Reduction Program Forecasting Software: Model Calibration Report and Users' Guide, prepared for New Jersey Department of Transportation, Comsis Corporation, 1994.

Ferguson, C. E., *Microeconomic Theory.* Richard D. Irwin, Inc., 1972.

Fowkes, Tony and Mark Wardman, "The Design of Stated Preference Travel Choice Experiments." *Journal of Transport Economics and Policy,* vol. 22, no. 1, pp. 59-69, 1988.

Fowkes, Tony, Mark Warkman, and Holden, D. L., *Nonorthogonality in SP Design.*, Working Paper, 1994.

Hensher, D. and Johnson, L.W., *Applied Discrete Choice Modelling.* New York: Halsted Press, 1981.

Jones, Peter, "An Overview of Stated Preference Techniques." *An Introduction to Stated Preference Techniques,* proceedings from PTRC conference, 1990.

Iida, Y., Akiyama, T. and Uchida, T., "Experimental Analysis of Dynamic Route Choice Behavior," *Transport Research-B,* Vol. 26B, no. 1, pp. 17-32, 1992.

Law, A. M. and Kelton, W. D., *Simulation Modelling and Analysis.* McGraw-Hill, Inc., 1991.

Dormencich, Thomas and McFadden, D., *Urban Travel Demand - a Behavioral Analysis.* North-Holland Publishing Company, 1975.

M. Ben Akiva and S. Lerman, *Discrete Choice Analysis.* MIT Press., Cambrige, Massachusetts, 1985.

Menon, Geeta, "The Effects of Accessibility of Information in Memory on Judgements of Behavioral Frequencies," *Journal of Consumer Research,* Vol. 20, pp. 431-440, 1993.

New Jersey Traffic Congestion and Air Pollution Control Act (P.L.1992, C.32, N.J.S.A. 27:26A-1).

Philips, Louis, *Applied Consumption Analysis.* North-Holland Publishing Company, 1974.

Pindyck, R.S. and Rubinfeld, D.L., *Econometric Models & Economic Forecasts.* McGraw-Hill, Inc., 1991.

Rosner, B., Spiegelman, D. and Willett, W., "Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: the Case of Multiple Covariates Measured with Error," *American Journal of Epidemiology,* Vol. 132, no. 4, pp. 734-745, 1990.

Wardman, M., "A Comparison of Revealed Preference and Stated Preference Models of Travel Behavior," *Jouranl of Transport Economics and Policy,* vol. 22, no. 1, pp. 71-91, 1988.

Willett, W., "An Overview of Issued Related to the Correction of Non-differential Exposure Measurement Error in Epidemiologic Studies," *Statistics in Medicine,* Vol. 8, no. 9, pp. 1031-1040, 1989.

Yatchew, A. and Griliches, Z., "Specification Error in Probit Models," *The Review of Economics and Statistics,* vol. 67, no. 1, pp. 134-139, 1985.