CONFIDENCE INTERVALS FOR WILLINGNESS-TO-PAY AND BEYOND: A COMPARATIVE ANALYSIS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Woon Lyn Esther Chiew

January 2015

CONFIDENCE INTERVALS FOR WILLINGNESS-TO-PAY AND BEYOND: A COMPARATIVE ANALYSIS

Woon Lyn Esther Chiew, Ph.D.
Cornell University 2015

The aim of this research is to investigate and develop methods for building confidence intervals (CIs) for parameter functions of discrete choice models, with a special focus on the CIs for willingness-to-pay measures. CIs are more than simply statistical measures. Rather, they are a convenient and easily understood means by which the variability of a parameter or sample statistic can be reported, especially because they can be presented graphically. CIs should be reported for all random statistics, and especially so in applied work where one cannot assume that the estimated parameter would exactly equal the true (unknown) parameter. Yet, when presenting willingness-to-pay values, the CIs are often neglected. This is partially because building CIs for willingness-to-pay values is not a trivial task, due to the possibility of discontinuity in the willingness-to-pay measure and its unknown probability distribution *a priori*. In addition, the methods used to build these intervals are debated greatly, with no consensus as to the best method to use. This research consolidates the contradictory results and presents reasons for the disparity currently present in the literature. It also extends the work of building CIs beyond willingness-to-pay measures to other parameter functions; in particular, this research demonstrates how CIs can be built for the probability that an airline passenger cancels his ticket.

The methods of building CIs are studied using Monte Carlo simulations and

case studies. Results indicate that when sample sizes or the price parameter is large (i.e. there are fewer chances for discontinuity to occur), all the preference space methods studied work equally well. However, under weak identification (when the price parameter is small), the Fieller method performs best. Hence, in general, the Fieller method should be the preferred method for building CIs for willingness-to-pay values.

This research also proposes the use of the Bayesian post-processing method to build CIs. This method, though a viable option, is not often discussed. The Bayesian method also has an edge over the other methods studied for several reasons, including the ease of constructing individual CIs and the ability to incorporate factors such as historical data into the model.

BIOGRAPHICAL SKETCH

Esther Chiew is currently completing a Ph.D. in Transportation Systems Engineering at Cornell University. She has previously completed a Masters of Engineering in Operations Research and Information Engineering at Cornell University, and a Bachelor of Science in Mathematics at University of Illinois at Urbana-Champaign. Esther is interested in the application of choice models to a wide variety of transportation systems. In particular, she is interested in factors that influence passenger decisions within the airline and alternative fuel vehicle industries. She has also worked for the World Bank as a consultant and economic modeling specialist. She can be reached at wc437@cornell.edu

ACKNOWLEDGEMENTS

I would like to thank my advisors at Cornell University: Dr. Ricardo Daziano and Dr. Linda Nozick, Department of Civil and Environmental Engineering, Dr. Huseyin Topaloglu, Department of Operations Research and Information Engineering, and Dr. Shanjun Li, Department of Applied Economics Management.

My main advisor, Dr. Daziano, has provided invaluable insight, direction and support for this work. His expertise in discrete choice modeling not only grew my own interest in this topic but also encouraged me to pursue different ways in which they could be used. Without his help and suggestions, this thesis would never have been possible.

Dr. Nozick, Dr. Topaloglu and Dr. Li also assisted greatly through the sharing of their knowledge, suggestions and pushing me to question different aspects of my work even further. I am grateful for the time they spent despite their very busy schedules.

I would like to thank Dr. Laurie Garrow of Georgia Institute of Technology for her ideas and advice that resulted in a large portion of this dissertation. Her provision of data for my use, expertise in the current models already used and great encouragement grew my interest in the airline industry.

To my parents and brother, who advised me when I first considered pursuing a Ph.D., encouraged me throughout, and listened to my complaints and problems. Thank you for all the love and care you have given me.

To Vineet John, current research analyst at Embarq, for his help with organizing data, proof-reading and searching the internet for useful papers. His assistance especially during the last months before my defense is much appreciated.

To him and all my friends at Cornell and CICF, thank you for your prayers and support, meals, teas and coffees, and all your care for me. I would not have made it through without you all being my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Confidence intervals (CIs) have long been used in all types of data analysis and statistical studies. Even before they were formally introduced by Jerzy Neyman in 1937, statisticians were reporting parameter estimates together with an interval created using estimated variances of the parameter. CIs have been applied to all types of research, from science and engineering to the social sciences.

As described by Neyman (1937), CIs first arose out of necessity. If one considers a true parameter θ and its estimate $\tilde{\theta}$, it is impossible to assume that these would be exactly equal. Hence a statistician would require some way to measure the accuracy of estimate $\tilde{\theta}$. The generally accepted method was to calculate the estimate $s_{\tilde{\theta}}^2$ of the variance σ^2 of $\tilde{\theta}$, and write the estimated result as $\tilde{\theta} \pm s_{\tilde{\theta}}$. The underlying understanding was that the true value θ would fall within this interval a majority of the time. Seeing how often this $\pm s_{\tilde{\theta}}$ was reported, it could be said that statisticians were in fact not trying to estimate one unique value $\tilde{\theta}$, but two estimates having the form $\underline{\theta} = \tilde{\theta} - k_1 s$ and $\bar{\theta} = \tilde{\theta} + k_2 s$, where k_1, k_2 are constants to be defined, and $\underline{\theta}, \bar{\theta}$ indicates the limits within which θ could be found.

Following this very practical use of intervals, Neyman (1937) formalized the theory of CIs to be as follows. Consider some confidence level α where $0 < \alpha < 1$. Then, $(\underline{\theta}, \bar{\theta})$ forms a CI if

$$\Pr(\underline{\theta} \leq \theta \leq \bar{\theta} | \theta) = 1 - \alpha \tag{1.1}$$

As a statistical means of interpreting results and evaluating model estimates, CIs are just one type among many different options. Why, then, are CIs important? In fact, CIs have often been disregarded in favor of hypothesis testing, where p-values are reported and the researcher concludes whether or not to reject a predefined hypothesis at a certain confidence level. However, CIs can not only be used to answer the same questions as a hypothesis test, they go beyond the test to provide other useful information as well. This was the conclusion within the medical field, where for many years hypothesis testing was used to formulate some conclusion about the effect of the factor being studied in a clinical trial. Several researchers began advocating for CIs, proposing several reasons for the use of CIs rather than other statistical measures (Gardner and Altman, 1986; Borenstein, 1994; Thompson, 2002; Masson and Loftus, 2003). Their conclusions can be summarized in the following points:

1. Hypothesis tests can only give a binary conclusion: whether there *is* an effect, or *not*. They cannot be used to determine magnitude of the effect (or whatever parameter is being studied). In contrast, CIs give a range of plausible values for the parameter of interest.
2. Sample statistics are imprecise due to both a degree of variability of the parameter and the limited sample size. Hypothesis tests are unable to capture the effect of sample size, as it is possible for any parameter to meet the criterion for significance with a large enough sample size. In contrast, both of these causes are reflected using CIs.
3. There is a tendency to equate statistical significance with medical importance or biological relevance.
4. CIs are able to perform the same hypothesis tests as p-values (by observing whether the CI contains 0).
5. CIs can be presented graphically, which tends to allow for better under-

stand on the parts of readers.

In 1999, the American Psychological Association (APA) Task Force on Statistical Inference issued a report in which they recommended the reporting of CIs (Wilkinson, 1999). In addition, the 2001 APA Publication Manual suggests that CIs represent ”in general, the best reporting strategy”.

The medical field is not unique in its support of the use of CIs. Within the education sector, the use of CIs came into focus when schools and districts had to report performance statistics after the implementation of the *No Child Left Behind* policy. There was no uniformity to the use of CIs in these reports, and Coladarci (2003) commented on this, arguing for the use of CIs based on many of the same reasons as those listed above. He gave special focus to the importance of CIs when the sample size is small. If only mean values were reported, small schools were more likely to be penalized for not having reached target performance levels. However, a small sample size results in higher variability in the data, which should be captured and reported. This can be done through the use of a CI, whose width increases as sample size decreases.

In the field of transportation, CIs have also been applied. In Chapter 4, I look more closely at the literature that advocates for or studies CIs for willingness-to-pay measures. However, CIs have been reported for many other parameter estimates as well. Chen et al. (2006) discuss the use of CIs for origin-destination (OD) demand estimates, and propose a method by which to build these CIs. According to the authors, CIs are especially relevant for OD demand estimates due to the uncertain nature of these estimates: demand is often estimated in advance to be used for planning, and there can be incomplete information from which the OD flows are inferred. This uncertainty makes it all the more important for CIs to be built. Other transportation studies in which CIs are used include analyzing the trade off between travel time and the availability and quality of bicycle lanes (Tilahun et al., 2007), calculating transportation mode share (Clark and McKimm, 2005), and forecasting the market share of alternative fuel vehicles (Mau et al., 2008). However, many of these and other studies simply list the CIs in the results table with at most a cursory mention in the text, and no explanation of its value or interpretation.

As can be seen, CIs are not only used in a wide variety of research areas, they are also seen in some as the best means by which to interpret the results of a study. Due to its importance and the mounting interest in building these intervals, this research looks at the specific application of CIs within the context of discrete choice models, with a majority focus on the CIs of willingness-to-pay.

1.1 Focus on Willingness-to-pay

Willingness-to-pay (WTP) measures provide a means to understanding how much value consumers place on the unit improvement of any particular attribute. They are important when reporting results of discrete choice models (McFadden, 2001), and in applications to policy analysis, especially aiming at welfare-improving scenarios. They are also especially useful in analyzing the feasibility and potential outcomes of implementing a proposed policy, and for this reason are often reported in cost-benefit analysis studies.

As a result of their widespread use and importance in applied work, interest in the accurate estimation and portrayal of WTP measures is high. For a linear-in-attributes discrete choice model, the WTP is the ratio of two parameter values (this formulation will be further discussed in chapter 2). It is thus easy to calculate and report this value along with other results of discrete choice experiments. Using WTP measures also allows for the easy comparison of attributes. For example, individuals often do not regard the different aspects of time (such as travel and waiting time) equally, but trying to observe this inequality simply from the values of marginal utility is difficult (what does it mean to raise one's utility by 0.1?). By using WTP, these aspects of time gain a monetary value, allowing for a more understandable comparison (being willing to pay, for instance, $5 to reduce one's waiting time by one hour versus $3 to reduce one's travel time by one hour).

WTP measures are also often applied in policy analysis, especially those aiming at welfare-improving scenarios. In cost-benefit analyses, the marginal cost of an improvement is compared to the WTP for said improvement, and decisions are made based on this comparison. For example, consider a proposal to improve the driving range of battery electric vehicles, which at current technology contain batteries which allow the car to travel an average of 100-200 miles before it needs to be recharged (U.S. Department of Energy, 2014). Suppose the current marginal cost of producing a battery with an additional mile is estimated to be $175/mile, and the WTP for this one-mile improvement in driving range is estimated to be $130/mile. Cost-benefit analysis would conclude that since the marginal cost outweighs the WTP, the proposal should be rejected.

However, since WTP is itself a random variable, reporting its standard errors or CIs is an important task, though often overlooked in applied work. Returning to the example of the battery electric vehicle, the conclusion to reject the proposal should be questioned if the CIs of the marginal cost and WTP overlap. This would imply that consumers could be willing to pay for an improvement to the driving range, and further analysis is required as to the viability of the proposal. In addition, different policy options being evaluated using mean WTP values could potentially result in inaccurate conclusions. Two WTP values, while numerically different, might have overlapping CIs, indicating that they are not statistically significantly different (cf. Park et al., 1991, for a similar argument concerning benefit estimates).

As a result, this work begins the analyses of building CIs with building those for WTP measures. Building CIs for WTP measures, and ratio measures in general, contain their own set of difficulties. I then move beyond WTP measures to analyze how CIs can be built for other functions of model parameters.

CHAPTER 2

RANDOM UTILITY MAXIMIZATION AND CHOICE BEHAVIOR MODELS

The purpose of this chapter is to give an introduction to random utility maximization (RUM) models, how they are structured, and how they can be solved. In particular, RUM models will be discussed in the context of the choice behavior of consumers facing discrete alternatives. Much of the content of this chapter is taken from the Nobel Laureate lecture delivered by McFadden (2001).

Prior to the 1960s, consumer theory was rarely applied empirically, and when used it was applied to national-level or market-level data. At such high levels of application, the theory was developed for a representative agent, and market level behavior was described according to this representative. If any observations were found to deviate from the representative's behavior, it was attributed to disturbances or data measurement errors, rather than to the unobserved differences among individuals. Clearly this was not an adequate means by which to understand individual behavior, as even individuals with similar demographics would have seemingly unaccounted-for differences in choices and preferences.

With the advent of digital computers and an growing amount of data concerning individual behavior in the 1960s, more emphasis was placed on the variation between individuals. It became increasingly important to explore and model these variations rather than attributing them to errors or disturbances. Daniel McFadden, working in the area of transportation, was the first to introduce many of the solutions and tools we use in solving these models today. McFadden also investigated and showed how the choice behavior models were consistent with RUM models.

In discrete choice, behavior models differ from the standard consumption theory in the demand function used. Standard consumption theory has a continuous demand function, but in discrete choice models, individuals choose a discrete quantity of goods. For example a consumer can only buy a discrete quantity of cars, and not some fraction of a car. An individual's satisfaction, or utility, increases upon acquiring goods, and more precisely, is dependent on the characteristics or attributes of the good. This is called the *hedonic approach*. Such attribute differentiation is necessary because often times goods can perform the same function. For example, all cars can fulfill the need for transportation. If choice were based purely on functionality of a good, then all goods fulfilling the same function would be chosen with equal probability. Yet it is clear from data that certain goods are chosen over others. Hence the hedonic approach is necessary in order to differentiate between products which an individual might be facing.

The discrete choice model is thus characterized as follows. An individual n faces a choice set C_n of of J_n different alternatives. He thus chooses alternative $i \in C_n = \{1, \ldots, J_n\}$. Each alternative has a set of attributes $\mathbf{x}_{in}(\mathbf{q}_{in}, c_{in})$, where $\mathbf{q}_{in}$ are the characteristics of the alternative and c_{in} is the cost of the alternative. Socio-economic characteristics of the individual, $\mathbf{s}_n$, may also be added to account for individual heterogeneity. Finally, the individual obtains utility U_{in}, which is the indirect utility conditional on the choice. U_{in} is a function of consumer preferences β, $\mathbf{x}_{in}$ and $\mathbf{s}_n$.

Discrete choice models are shown to be consistent with RUM models in two ways. Firstly, an individual makes his choice based on utility maximization.

RUM works on the idea of rational consumer behavior, in which a rational individual would work to maximize his satisfaction, or utility. Hence, in order for an alternative i to be chosen, its utility U_{in} must be greater than (or equal to) the utility U_{jn} obtained from each of the other alternatives in the choice set C_n.

RUM also introduces the notion that the utility is random. The basis of the RUM model came from the work of Thurstone in 1927, who wrote on psychophysical discrimination. In his seminal paper, Thurstone (1927) theorized that an alternative i, with true stimulus level V_i, would instead be perceived with a normal error as $V_i + \sigma\varepsilon_i$. When this 'stimulus' is replaced by levels of utility, the model can be interpreted as an economic choice model. Marschak (1960) introduced this work into the economics literature, calling it the RUM model and using it to explore the theoretical implications for the choice probabilities of the maximization of utilities that contained random elements.

Another important study in the choice behavior literature was by Luce, who introduced an Independence of Irrelevant Alternatives (IIA) axiom. IIA states that for every choice set C that contains both alternatives i and j, the ratio of the choice probabilities of these alternatives is the same. Luce (1959) showed that for positive choice probabilities, IIA implies strict utilities U_i such that the probability of choosing alternative i, $P_C(i) = \frac{U_i}{\sum_{k \in C} U_k}$. Marschak (1960) then linked this theory with the concept of random utility.

In choice behavior models, the notion of random utility is important for two reasons. Firstly, individual choice behavior contains variation and is intrinsically probabilistic. Secondly, the modeler often obtains only incomplete information. That is, the analyst is unable to fully observe all variables that influence the decision of the individual. These two reasons make it essential to view individual utility as random rather than deterministic.

The remainder of the chapter will be focused on introducing the different models, in particular fixed coefficient and random coefficient models. The solutions for these discrete choice models will also be outlined, differing according to the distributional assumptions of the error term. In addition, how WTP measures are formed from the coefficient estimates will be discussed.

2.1 Fixed Coefficient Models

As outlined above, consider a standard choice situation in which an individual n chooses an alternative $i \in C_n = \{1, \ldots, J\}$ by maximizing the utility U_{in} among the random utility vector $\mathbf{U}_n = [U_{1n} \ldots U_{Jn}]$. Because the utility is random, the analyst can decompose this into a deterministic utility $\mathbf{V}_n$ and an error term $\boldsymbol{\varepsilon}_n$. In practice, the deterministic utility is often further assumed to be linear-in-attributes:

$$\begin{aligned} \mathbf{V}_n &= \mathbf{X}_n \boldsymbol{\beta} \\ &= \mathbf{x}'_{cn}\beta_c + \mathbf{x}'_{1n}\beta_1 + \cdots + \mathbf{x}'_{Kn}\beta_K \end{aligned} \tag{2.1}$$

where $\mathbf{X}_n$ is a matrix of observable exogenous attributes, $\mathbf{x}'_{cn}$ is the attribute of cost (or purchase price) and $\mathbf{x}'_{kn}, k \in \{1, \ldots, K\}$ are the remaining K attributes. Note that in this fixed coefficient model, all individuals are assumed to have the same consumer preferences, i.e., $\boldsymbol{\beta}$ does not differ by individual.

An individual chooses alternative i if $U_{in} \geq \max_{j \neq i} U_{jn} \; \forall \; i, j \in C_n$. However, since the $\boldsymbol{\varepsilon}_n$ terms are random, the analyst can only ascertain the choice probability of the individual. The probability that an individual chooses alternative i

is given by

$$
\begin{aligned}
P_{in} = \mathbb{P}(i|C_n) &= \mathbb{P}(U_{in} > U_{jn}, \forall\, j \in C_n, j \neq i) \\
&= \mathbb{P}(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}, \forall\, j \in C_n, j \neq i) \qquad (2.2) \\
&= \mathbb{P}(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \forall\, j \in C_n, j \neq i)
\end{aligned}
$$

Supposing that $\boldsymbol{\varepsilon}_n$ has probability density function $f(\boldsymbol{\varepsilon}_n)$, the choice probability can be written as

$$
P_{in} = \int_{\boldsymbol{\varepsilon}} I(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \forall\, j \in C_n, j \neq i) f(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n, \qquad (2.3)
$$

where $I(\cdot)$ is the indicator function.

There are two important aspects of the choice decision process that affect the specification and estimation of choice models. Firstly, only utility differences matter. This can be seen from equation (2.2), where the choice probability depends on the difference in utilities V_{in} and V_{jn}. In practical terms, this also makes sense. The numerical value of U_{in} does not matter to an individual; what matters is whether one good is providing a higher utility than another good, i.e. the difference in these utilities. That only utility differences matter implies that only attributes that capture differences across alternatives can be estimated. For example, if the same socio-economic variable enters the utility of all alternatives, then the effect of that variable on each alternative cannot be estimated. Instead, only the relative differences can be estimated. The same applies for alternative-specific constants that enter into the utility function. Train (2009) contains specific examples of these problems.

Secondly, the overall scale of the utility does not matter. That is, multiplying the utilities of every alternative by the same constant does not change the individual's choice. As a result, the analyst must normalize the scale of the utility.

In practice, this is usually done by normalizing the variance of the error term, or setting the scale of the utility to equal 1.

Different discrete choice models are derived from different assumptions over the error term. Sections 2.1.1 and 2.1.2 assume that the error term is distributed extreme value and normal, and are called the multinomial logit model (MNL) and multinomial probit model (MNP) respectively.

2.1.1 Multinomial Logit Model

Assuming that each error term ε_{in} is independently and identically distributed extreme value, i.e. $\varepsilon_{in} \overset{iid}{\sim} \text{EV}(0, \lambda)$, gives the MNL model. In practice, the scale is typically normalized by setting $\lambda = 1$. With this normalization, the probability of individual n choosing alternative i can be expressed by the following closed form equation:

$$P_{in} = \frac{\exp(\mathbf{x}'_{in}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{jn}\boldsymbol{\beta})}, \qquad \forall\, i \in C_n \tag{2.4}$$

The most common frequentist, or classical, method of solving the MNL model is via maximum likelihood estimation (MLE). Specifically, consider that we have choice indicators $\mathbf{y}$ such that

$$y_{in} = \begin{cases} 1, & \text{if individual } n \text{ chooses alternative } i \\ 0, & \text{otherwise} \end{cases} \tag{2.5}$$

Then the probability of individual n actually choosing alternative i can be expressed as $\prod_{i \in C_n} (P_{in})^{y_{in}}$. Assuming that each individual makes his choice independently from all other individuals, then the probability of each individual actually choosing the alternative which he was observed to have chosen (also called the likelihood), is given by

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N} \prod_{i \in C_n} (P_{in})^{y_{in}} \tag{2.6}$$

MLE searches for the $\boldsymbol{\beta}$ that maximizes the value of the likelihood function, or alternatively, the log-likelihood function which is given by the log of equation (2.6):

$$\begin{aligned} LL(\boldsymbol{\beta}) &= \ln \left(\prod_{n=1}^{N} \prod_{i \in C_n} (P_{in})^{y_{in}} \right) \\ &= \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_{in} \end{aligned} \tag{2.7}$$

MNL models are useful for a number of reasons. Firstly, it is by far the simplest model to use and estimate. Since the choice probabilities can be expressed as a closed-form equation, MLE can be easily applied in order to solve for the required $\boldsymbol{\beta}$ values. Secondly, MNL models exhibit the IIA property (Luce, 1959). Note that for any two alternatives i and k,

$$\begin{aligned} \frac{P_{in}}{P_{kn}} &= \frac{\exp(V_{in}) / \sum_{j \in C_n} \exp(V_{jn})}{\exp(V_{kn}) / \sum_{j \in C_n} \exp(V_{jn})} \\ &= \frac{\exp(V_{in})}{\exp(V_{kn})} \\ &= \exp(V_{in} - V_{kn}) \qquad , \end{aligned}$$

that is the ratio of the choice probabilities of two alternatives is independent of the choice probability of any other alternative.

When IIA accurately represents the reality of the situation being modeled, then MNL models provide significant advantages. IIA makes it possible to consistently estimate the model parameters on a subset of alternatives for each individual. This can greatly improve the computational time, especially if the original data set contains so many alternatives as to be impossible to solve. In addition, if the analyst is only interested in the choices between a subset of alternatives (even though others might exist), then she can save time and effort by only collecting the data for the alternatives in question, rather than for the entire spectrum of possible alternatives.

However, there are also three main reasons for which MNL models are limited. Firstly, IIA is often not an accurate representation of the choice situation. A oft-quoted example of a situation in which it is inappropriate to assume IIA is the red-bus-blue-bus problem. Suppose that an individual has a choice of travel modes between a red bus and a car, and that the choice probabilities are equal, i.e. $P_{\text{red bus}} = P_{\text{car}} = \frac{1}{2}$. Now suppose a blue bus, which is identical to the red bus in every way except color, is introduced as a travel mode option. In this choice situation, one would expect (logically) that the choice probabilities would become $P_{\text{red bus}} = P_{\text{blue bus}} = \frac{1}{4}$, $P_{\text{car}} = \frac{1}{2}$. However, due to IIA, the MNL model would predict that the choice probabilities be $P_{\text{red bus}} = P_{\text{blue bus}} = P_{\text{car}} = \frac{1}{3}$. This situation, and many others, do not follow the IIA property, and thus using the MNL model imposes inaccurate representations of the choice situation.

Secondly, MNL models cannot represent random taste variation. In particular, if consumer preferences $\boldsymbol{\beta}$ varies randomly, the randomness will be subsumed by the error term ε, which would then contradict the IID assumptions of the error term. As such, random taste variation cannot be modeled using MNL.

Finally, MNL models cannot be used with panel data where, for each individual, there are unobserved variables which are correlated over time. This is because the unobserved variables are also subsumed, as with random taste variation, by the error term, which is assumed to be independently distributed. As a result, the time correlation will not be accounted for by the MNL model.

In order to benefit from the closed form equations of the MNL model yet account for the fact that IIA is often not an accurate representation of the choice situation, an analyst can use a generalized extreme value (GEV) model. GEV models contain all models which assume that the error terms ε are jointly distributed generalized extreme value. Apart from this rule, GEV models can allow for any variety of substitution patterns between the alternatives. In this way, the often incorrect assumption of IIA can be overcome. Note that if there are no correlations between alternatives, then the GEV model is the standard MNL model.

The most commonly used GEV model is the Nested Logit (NL) model. This model allows alternatives to be placed into different groups, or nests. Alternatives within the same nest are better substitutes of each other, and IIA holds within nests. However, alternatives in different nests are not good substitutes, or are correlated to other alternatives in the two nests. Hence IIA does not exist between nests. Consider, for example, an individual who is deciding how to travel to work. He has the alternative of either driving alone, carpooling, taking the bus, or taking the train. One way to partition these alternatives into nests would be to place the first two into one nest and the next two into a different nest, as shown in Figure 2.1. Since both driving alone and carpooling involves cars, while taking the bus or train are transit options, it makes sense that the alternatives within each nest would be good substitutes of each other, but not alternatives between nests.

While GEV models can relax the assumption of IIA, the other limitations of the MNL model cannot be so easily overcome will still maintaining the extreme value distribution of the error terms. The MNP model, as described in the fol-

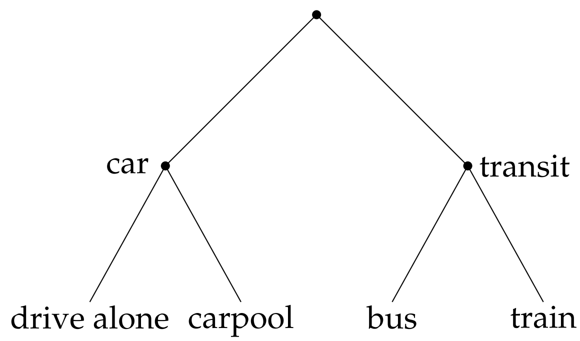


Figure 2.1: Example of 2-level nested logit model

lowing section, is a model that can overcome these restrictions.

2.1.2 Multinomial Probit Model

MNP models, unlike GEV models, handle all three restrictions of MNL models. Unfortunately this comes with a price, because MNP models assume that error terms follow a normal distribution. As a result, the choice probabilities depend on an integral for which there is no closed form, making estimation more difficult.

The MNP model is derived as follows. Assume that the error term ε is distributed multivariate normal, i.e. $\varepsilon \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is the general variance-covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2J} \\ \vdots & & \ddots & \vdots \\ \sigma_{J1} & \sigma_{J2} & \cdots & \sigma_J^2 \end{bmatrix} \tag{2.8}$$

Following equation (2.3), the probability of individual n choosing alternative

i then becomes

$$P_{in} = \int_{\boldsymbol{varepsilon}} I(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \forall\ j \in C_n, j \neq i)\phi(\boldsymbol{\varepsilon}_n)d\boldsymbol{\varepsilon}_n, \tag{2.9}$$

where $\phi(\cdot)$ is the probability density function of the multivariate normal distribution.

In order to estimate this integral, we express the choice probability in the following way. Since only utility differences matter, consider (without loss of generality) the model in differences with respect to the first alternative; i.e., consider the model

$$\tilde{U}_n = \Delta_1 \mathbf{U}_n = \begin{bmatrix} U_{2n} - U_{1n} \\ \vdots \\ U_{Jn} - U_{1n} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_{2n} - \mathbf{x}'_{1n} \\ \vdots \\ \mathbf{x}'_{Jn} - \mathbf{x}'_{1n} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \varepsilon_{2n} - \varepsilon_{1n} \\ \vdots \\ \varepsilon_{Jn} - \varepsilon_{1n} \end{bmatrix}, \tag{2.10}$$

where Δ_1 is the matrix difference operator that creates the vectors above. Note that with this specification, the probability that individual n chooses alternative 1 is now $P_{1n} = \mathbb{P}(U_{1n} > U_{jn}, \forall\ j \in C_n, j > 1) = \mathbb{P}(\Delta_1 U_{jn} < 0, \forall\ j \in C_n, j > 1)$.

Since the difference between two normals is normal, $\Delta_1 \varepsilon_n$ is multivariate normally distributed. It can be shown that the covariance matrix of the model becomes $\tilde{\boldsymbol{\Sigma}}_1 = \Delta_1 \boldsymbol{\Sigma} \Delta'_1$. Then the choice probability becomes a $(J-1)$-dimensional integral

$$\begin{aligned} P_{in} &= \mathbb{P}(U_{in} > U_{jn}, \forall\ j \in C_n, j \neq i) \\ &= \mathbb{P}(U_{in} - U_{1n} > U_{jn} - U_{1n}, \forall\ j \in C_n, j \neq i) \\ &= \mathbb{P}(\tilde{U}_{in} > \tilde{U}_{jn}, \forall\ j \in C_n, j \neq i) \\ &= \mathbb{P}(\tilde{V}_{in} + \tilde{\varepsilon}_{in} > \tilde{V}_{jn} + \tilde{\varepsilon}_{jn}, \forall\ j \in C_n, j \neq i) \\ &= \mathbb{P}(\tilde{\varepsilon}_{jn} - \tilde{\varepsilon}_{in} < \tilde{V}_{in} - \tilde{V}_{jn}, \forall\ j \in C_n, j \neq i) \\ &= \int I(\tilde{\varepsilon}_{jn} - \tilde{\varepsilon}_{in} < \tilde{V}_{in} - \tilde{V}_{jn}, \forall\ j \in C_n, j \neq i)\phi(\tilde{\varepsilon}_n)d\tilde{\varepsilon}_n \end{aligned} \tag{2.11}$$

In order to estimate the MNP model using MLE, the choice probabilities have to be effectively calculated. $(J-1)$-dimensional integrals, with no closed form, are difficult to evaluate, and numerical evaluation is possible for only up to 3 alternatives. Hence, simulation is used to calculate the choice probabilities. Rather than maximizing the likelihood as similar to equation (2.7), we instead maximize the simulated log-likelihood (SLL), which uses simulated choice probabilities, i.e.

$$SLL = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln \hat{P}_{in} \tag{2.12}$$

This is maximized over both β and Σ.

The simulated choice probabilities $\hat{P}_{in}$ can be obtained using a variety of methods, but the most widely used is the GHK simulator, named after Geweke (1989, 1991), Hajivassiliou (as reported in Hajivassiliou and McFadden (1998)), and Keane (1990, 1994), who are credited for having developed this simulator. This simulator works on utility differences, as described above. In addition, which utility is subtracted from the others depends on the choice probability being simulated, i.e. if P_{1n} is being simulated, than utility U_{1n} is subtracted from all other utilities $U_{2n}, \ldots, U_{Jn}$, and so forth. Hence, without loss of generality, suppose one is simulating the choice probability for alternative 1. Then the model in differences is as described above in equation (2.10). One obtains the Cholesky decomposition $\mathbf{L}_1$ of the new covariance matrix $\tilde{\mathbf{\Sigma}}_1 = \mathbf{L}_1 \mathbf{L}_1^{'}$, where

$$\mathbf{L}_1 = \begin{pmatrix} c_{11} & 0 & \ldots\ldots & 0 \\ c_{21} & c_{22} & 0 \quad \ldots & 0 \\ \vdots & \vdots & \vdots \quad \vdots & \vdots \end{pmatrix} \tag{2.13}$$

With the Cholesky decomposition, $\tilde{\mathbf{U}}_n = \tilde{V}_n + \tilde{\boldsymbol{\varepsilon}}_n = \tilde{V}_n + \mathbf{L}_1 \boldsymbol{\eta}_n$, where $\boldsymbol{\eta}_n$ is a $J_n - 1$ vector whose elements are IID standard normal. Written explicitly, this becomes

$$\tilde{U}_{2n} = \tilde{V}_{2n} + c_{11}\eta_{1n}$$
$$\tilde{U}_{3n} = \tilde{V}_{3n} + c_{12}\eta_{1n} + c_{22}\eta_{2n}, \tag{2.14}$$

and so on. The probability of individual n choosing alternative 1 is then

$$\begin{aligned} P_{1n} &= \mathbb{P}(\tilde{U}_{jn} < 0, \forall\, j \in C_n, j > 1) \\ &= \mathbb{P}\left(\eta_{1n} < \frac{-\tilde{V}_{2n}}{c_{11}}\right) \times \mathbb{P}\left(\eta_{2n} < \frac{-\tilde{V}_{3n} - c_{12}\eta_{1n}}{c_{22}} \middle| \eta_{1n} < \frac{-\tilde{V}_{2n}}{c_{11}}\right) \times \cdots \end{aligned} \tag{2.15}$$

Using this probability, the GHK simulator proceeds as follows:

1. Calculate $\mathbb{P}\left(\eta_{1n} < \frac{-\tilde{V}_{2n}}{c_{11}}\right) = \Phi\left(\frac{-\tilde{V}_{2n}}{c_{11}}\right)$.
2. Draw η_{1n}^s from a truncated standard normal distribution, truncated at $\frac{-\tilde{V}_{2n}}{c_{11}}$.
3. Calculate $\mathbb{P}\left(\eta_{2n} < \frac{-\tilde{V}_{3n}-c_{12}\eta_{1n}}{c_{22}} \middle| \eta_{1n} = \eta_{1n}^s\right) = \Phi\left(\frac{-\tilde{V}_{3n}-c_{12}\eta_{1n}^s}{c_{22}}\right)$.
4. Draw η_{2n}^s from a truncated standard normal distribution, truncated at $\frac{-\tilde{V}_{3n}-c_{12}\eta_{1n}^s}{c_{22}}$.
5. Continue for all alternatives $j \in C_n, j > 1$.
6. Simulated probability for s^{th} draw is $\hat{P}_{1n}^s = \Phi\left(\frac{-\tilde{V}_{2n}}{c_{11}}\right) \times \Phi\left(\frac{-\tilde{V}_{3n}-c_{12}\eta_{1n}^s}{c_{22}}\right) \times \cdots$.
7. Repeat above steps for $s = 1, \ldots, S$.
8. Simulated choice probability $\hat{P}_{1n} = \frac{1}{S}\sum_{s} \hat{P}_{1n}^s$.

2.2 Random Coefficient Model

Motivation for a random coefficient model is rather logical. It is not possible to claim that all individuals in a sample, or population, have exactly the same preferences, which is the assumption of the fixed coefficient model. Instead, variations exist between groups of people for almost all products, and these are called *taste variations*. These differences are often the concern of analysts trying to perform market segmentation, or understand the preferences of certain groups of people.

There are several ways to analyze taste variations. The first is in a deterministic way, in which certain market segments are pre-identified by the analyst, and fixed coefficients are added to the model to account for this. For example, supposed the analyst thinks that men and women will have different preferences for some attribute x_k. One common strategy is to use a different coefficient for men and women, i.e. β_{km} and β_{kw}. An alternative strategy is to add a coefficient representing the taste variation of one group relative to another, e.g. using $(\beta_k + \beta_{kw} I_{wn}) x_{kn}$, where I_{wn} is the indicator representing whether individual n is a woman. Then the marginal utility of attribute k for men is β_k, while that for women is $\beta_k + \beta_{kw}$.

While deterministic taste variations can test the validity of pre-identified market segments, it cannot on its own identify these segments. As can be imagined, there are many possible segments that could exist, and it would be extremely time consuming for the analyst to individually test each of these segments deterministically. The analyst would more likely test the segments which are either observable, or which she knows would exist from past experience (such as by gender, age, income group). However, there may also be many unobservable taste variations that should be accounted for.

To account for these, random taste variations can be applied. This assumes that variations in taste cannot be explained in some systematic way (as with deterministic taste variations). A common approach to this problem is to apply a parametric assumption to the distribution of taste preferences, i.e. assume that the coefficients are random variables following some parametric distribution, $\beta_k \sim f(\beta_k|\boldsymbol{\theta}_{\beta_k})$, where $\boldsymbol{\theta}_{\beta_k}$ are the parameters describing the parametric distribution. With this approach, each individual n has taste preferences β_{kn} for attribute k, and each β_{kn} is drawn from the distribution of β_k. Hence, the utility of individual n for choosing alternative i becomes

$$U_{in} = \mathbf{x}_{in}^{'}\boldsymbol{\beta}_n + \varepsilon_{in}, \tag{2.16}$$

where $\mathbf{x}_{in}$ is the vector of attributes associated with alternative i and individual n, ε_{in} is the random term, and $\boldsymbol{\beta}_n$ the vector of coefficients representing the taste preferences unique to individual n.

In line with RUM, the individual, who knows his own preferences, chooses alternative i if and only if $U_{in} > U_{jn} \forall\ j \neq i$. However, the analyst cannot actually observe the individual's β_n. If she did, in fact, know the individual's taste preferences, then the choice probability would be identical to that of the fixed parameter case in equation (2.3). Thus, in order to obtain the choice probability, it is necessary to consider all possible values of β_n:

$$P_{in} = \int \left(\int_{\boldsymbol{\varepsilon}} I(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \forall\ j \in C_n, j \neq i) f(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n \right) f(\boldsymbol{\beta}) d\boldsymbol{\beta} \tag{2.17}$$

There are often no closed forms for equation (2.17). In addition, depending on how many parameters are assumed to be random, the integral can contain many dimensions. As a result, random coefficient models are most often solved using simulation methods.

As with fixed coefficient models, the different discrete choice models are derived from different assumptions over the error term ε. Subsections 2.2.1

and 2.2.2 assume that the random term is distributed extreme value and normal respectively.

2.2.1 Mixed Logit Model

The mixed multinomial logit (MMNL) model, also known as the mixed logit (ML) model, is derived by assuming that each error term ε_{in} is IID distributed extreme value, i.e. $\varepsilon_{in} \overset{iid}{\sim}$ EV$(0, \lambda)$. As before, the scale is typically normalized by setting $\lambda = 1$. If the analyst could observe the individual's taste preferences (i.e., conditional on taste preferences β), then the probability that an individual n chooses alternative i is

$$P_n(i|\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_{in}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{jn}\boldsymbol{\beta})}, \qquad \forall i \in C_n \tag{2.18}$$

Hence the unconditional choice probability thus becomes

$$\begin{aligned} P_{in} &= \int P_n(i|\boldsymbol{\beta}) f(\boldsymbol{\beta}) d\boldsymbol{\beta} \\ &= \int \frac{\exp(\mathbf{x}'_{in}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{jn}\boldsymbol{\beta})} f(\boldsymbol{\beta}) d\boldsymbol{\beta}, \qquad \forall i \in C_n \end{aligned} \tag{2.19}$$

The ML model is an extremely flexible model which, under regularity conditions, is able to approximate any RUM model to any degree of accuracy (McFadden and Train, 2000). It also overcomes the three limitations of the MNL model, while maintaining the assumption of a logit error distribution, which is why the conditional choice probability can be expressed in a closed form as with the MNL model. This eliminates one stage of simulation needed in the estimation process.

The flexibility of the ML model can be seen in the expression of $f(\boldsymbol{\beta})$. The analyst can specify any distribution for the taste preferences, and then estimate the parameters of this distribution. Taste preferences are often assumed to be distributed normal or lognormal. Lognormal distributions are advantageous when the coefficient is known to have the same sign over all individuals. For example, the cost coefficient can be assumed to distributed lognormal because all individuals should have a negative sign for this coefficient (a positive sign would indicate that paying more gives higher utility). Other distributions including triangle and uniform distributions have also been used. The ML model also allows the analyst to apply different distribution assumptions to different coefficients in the model.

Estimation of the ML model again proceeds via MLE. Using the same notation as before, the log-likelihood function is given by

$$\begin{aligned} LL(\beta) &= \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_{in} \\ &= \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln \left(\int \frac{\exp(\mathbf{x}'_{in}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{jn}\boldsymbol{\beta})} f(\boldsymbol{\beta}) d\boldsymbol{\beta} \right) \end{aligned} \tag{2.20}$$

The analyst wants to find the β that maximizes this equation. However, there are two problems to doing this directly. Firstly, there is no closed form equation to the integral, and thus simulation is necessary to calculate the choice probabilities. Secondly, β is random, and thus the analyst needs to specify some distribution $f(\beta|\theta)$ for β. She is then trying to find the values of θ that maximizes the log-likelihood function. For some given value of θ, the choice probabilities are simulated as follows:

1. Draw β^s from $f(\beta|\theta)$.
2. Calculate $\hat{P}^s_{in} = \frac{\exp(\mathbf{x}'_{in}\boldsymbol{\beta}^s)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{jn}\boldsymbol{\beta}^s)}$.
3. Repeat above steps for $s = 1, \ldots, S$.

4. Simulated choice probability $\hat{P}_{in} = \frac{1}{S} \sum_{s} \hat{P}^s_{in}$.

These simulated choice probabilities are then used in the simulated log-likelihood equation, which is maximized over θ.

$$SLL = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln \hat{P}_{in} \tag{2.21}$$

2.2.2 Random Coefficient Probit Model

The random coefficient probit model works similarly to the ML model, except that the error term is assumed to follow a normal distribution. The conditional choice probability is thus that of the MNP model (equation (2.9)). As with the MNP model, this means that there is no closed form expression for the conditional choice probability, let alone the unconditional choice probability.

Little work has been done with the random coefficient probit model. While an extremely useful and highly flexible model, it faces the difficulties of estimation. Even without considering the additional integral dimensions arising from the random coefficients, estimation of the MNP model requires a high computational cost. The MSL method for estimating the MNP model as described in subsection 2.1.2 has been criticized for several reasons (Bhat, 2011). Firstly, accuracy of any simulation technique is known to degrade as the number of dimensions of integration increases, which also results in simulation noise rising substantially. In addition, the covariance matrix of the estimator is often numerically estimated with low accuracy. Evaluating this with a high accuracy further increases computational cost to the point of being infeasible for research.

Despite these difficulties, using a random coefficient probit model gives advantages even beyond those of the ML model. Bhat and Sidharthan (2012) highlight several of these advantages. Firstly, in cases when the utility of individuals have a spatial dependency component, the parametric covariance that results is infeasible, or at best extremely inefficient to incorporate over a restrictive EV kernel covariance surface. Secondly, if the random coefficients take on an assumption of being normally distributed, then the resulting random parameter model collapses into a regular MNP model. However, this is still difficult to solve using the MSL method. As such, Bhat (2011) proposes a method called the "Maximum Approximate Composite Marginal Likelihood (MACML)" method as an alternative to the MSL method. The MACML procedure uses an analytic approximation method rather than simulation to evaluate the cumulative distribution function of the multivariate normal, which improves on the accuracy of the parameter and covariance matrix estimates.

The Bhat (2011) version of the MACML procedure was only applicable for the special case when the random parameters were distributed normally. However, Bhat and Sidharthan (2012) derived the model for cases when the random parameters can take on other distribution assumptions, in particular the multivariate skew-normal distribution function. It is clear that much more work can be done in this area, and might happen as technology improves to accommodate the high computational costs necessary for estimation.

2.3 Willingness-to-pay Measures

In all the models discussed above, the deterministic utility V_{in} is assumed to be linear-in-attributes. As a result, the derivative of U_{in} with respect to changes in attribute x_{kin} and cost x_{cin} is given by $\mathrm{d}U_{in} = \beta_k \mathrm{d}x_{kin} + \beta_c \mathrm{d}x_{cin}$. Equating this to zero and solving for $\mathrm{d}x_{cin}/\mathrm{d}x_{kin}$ gives the change in cost that would allow total utility to remain constant given a change in some attribute x_{kin}, or the WTP for an improvement in attribute x_{kin}. This is also known in Economics as the marginal rate of substitution.

$$\frac{\mathrm{d}x_{cin}}{\mathrm{d}x_{kin}} = \mathrm{WTP}_k = -\frac{\beta_k}{\beta_c} \tag{2.22}$$

As has been discussed above, parameters of a discrete choice model are usually estimated using MLE. Using this estimation method implies that the estimated parameters are asymptotically distributed multivariate normal. This means that the WTP is a random variable, since it is a ratio of two random variables. In addition, as the ratio of two normally distributed variables, WTP is governed by an unknown probability distribution *a priori*. This results in a host of complications when estimating these measures and building CIs for them, as will be detailed in Chapter 3.

CHAPTER 3

METHODOLOGY

This chapter gives an introduction to the various method of building CIs, in particular for WTP measures. The standard equation for CIs uses the estimated mean and standard errors of the random variable obtained from the model. Hence for some random variable θ that is assumed to follow a normal distribution, the CIs are given by

$$(\hat{\theta} - z_{\alpha/2} s_\theta, \hat{\theta} + z_{\alpha/2} s_\theta), \tag{3.1}$$

where $\hat{\theta}$ and s_θ are the estimated mean and standard error of θ respectively, α is the confidence level chosen and $z_{\alpha/2}$ is the inverse of the cumulative standard normal distribution at probability level $(1 - \frac{\alpha}{2})$; i.e. the value $z_{\alpha/2}$ such that $\mathbb{P}(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

However, WTP measures are computed as the ratio of the model parameter estimates, rather than estimated directly from the model. As a result, the standard equation above cannot be applied wholesale to the building of their CIs. One can easily see a problem posed by these ratio measures: a ratio will have a singularity where its denominator is equal to zero. Hence if the denominator is close to zero (i.e. when weak identification exists), it becomes difficult to accurately estimate the model parameters, consequently resulting in inaccurate ratio estimates and CIs (Dufour, 1997). In addition, parameter estimates obtained using MLE result in a random WTP variable with an unknown *a priori* probability distribution. Thus, it is not a trivial matter to obtain the mean and standard error of the WTP variable.

Two situations exist in which the WTP measure does follow a known probability distribution *a priori*. First, WTP follows a Cauchy distribution when the parameters are independently distributed standard normal (Arnold and Brockett, 1992). However, a Cauchy distribution has no moments, and thus standard methods of interval estimation cannot be used. In addition, it is unlikely for the parameters to be independently distributed standard normal. Second, when the coefficient of variation of the denominator (for the WTP, this is the cost coefficient β_c) is negligible, i.e., when the standard error of β_c is small compared to its mean, then the distribution of WTP is likely to be approximately normal (Fieller, 1932; Hinkley, 1969). This is still a strong assumption that might not always hold; however, in deriving the CIs for WTP measures, it is usually assumed that the WTP measure is distributed normal, as will be made clear below.

The most prevalent methods of constructing CIs of WTP in the fields of Economics and Transportation Science are the Delta method, the Fieller method, and the Krinsky-Robb method. This chapter will explain these three methods, and also include a variety of other methods, namely the use of WTP-space, the Bayesian post-processing method, and the use of the Bayes Factor. These three methods have not been widely used in the literature to construct CIs. Working in WTP-space has been shown to produce more accurate results than working in preference space (Sonnier et al., 2007), but has not been used before to build CIs. The Bayesian post-processing method is rarely used or considered in the existing literature, despite the fact that it involves only at most as much work as the other commonly used methods. The Bayes factor is a test statistic used in Bayesian hypothesis testing to compare null and alternative hypotheses. While it has been used in this capacity for many years, it has only recently been used in building CIs (Guerron-Quintana et al., 2013).

The remainder of this chapter describes the different basic assumptions, formulation, and advantages and disadvantages of each method. The list below describes the notation used in the equations following:

- $\boldsymbol{\beta}$: mean of asymptotic normal distribution of parameter estimates
- $\boldsymbol{\Sigma}_\beta$: covariance matrix of asymptotic normal distribution of parameter estimates
- $\hat{\beta}_k$: point estimate of the parameter of attribute x_k
- $\hat{\beta}_c$: point estimate of the parameter for cost x_c
- $\widehat{\text{WTP}}_k$: willingness-to-pay for attribute x_k, calculated as $\frac{\hat{\beta}_k}{\hat{\beta}_c}$
- $\hat{\boldsymbol{\Sigma}}_{kc}$: submatrix of $\hat{\boldsymbol{\Sigma}}_\beta$ corresponding to $\hat{\beta}_k$ and $\hat{\beta}_c$, written explicitly as $\begin{bmatrix} \hat{\nu}_k & \hat{\nu}_{kc} \\ \hat{\nu}_{kc} & \hat{\nu}_c \end{bmatrix}$
- α: confidence level of intervals produced

The methods for building CIs can be split into those used for fixed or random coefficient models, and then further into frequentist and Bayesian methods. The following sections classify the methods according to these categories as in Table 3.1.

Table 3.1: Different methods of building confidence intervals

	Frequentist	**Bayesian**
Fixed Parameter	Delta (Mean) Fieller Krinsky-Robb WTP-Space	Post-Processing WTP-Space Bayes Factor
Random Parameter	Delta (Mean) Delta (Median) Krinsky-Robb	Post-Processing

3.1 Fixed Coefficient Models

There are many well established methods in the literature for building CIs of ratio measures that can be applied to the WTP measures obtained from fixed coefficient models. Recall that the fixed coefficient model assumes *a priori* that all individuals have the same (fixed) taste preferences. As a result, there is also only one WTP measure to account for, rather than WTP values that vary across individuals.

3.1.1 Frequentist Methods

Within the Transportation Science and Economics literature, the Delta method, Fieller Method, and Krinsky-Robb methods are the most often used. Also included in this section of frequentist methods is the WTP-space method.

Delta Method

In general, the Delta method estimates the variance of a non-linear function of two or more random variables by first taking the first-order Taylor expansion around the mean value of the variables, then calculating the variance of this expression (see example in Greene, 2003). By using this method, the variance of the WTP is thus:

$$\begin{aligned}\text{var}(\widehat{\text{WTP}}_k) &= (\widehat{\text{WTP}}_{\beta_k})^2\text{var}(\hat{\beta}_k) + (\widehat{\text{WTP}}_{\beta_c})^2\text{var}(\hat{\beta}_c) \\ &\quad + 2\widehat{\text{WTP}}_{\beta_k}\widehat{\text{WTP}}_{\beta_c}\text{covar}(\hat{\beta}_k, \hat{\beta}_c) \\ &= \left(\frac{-1}{\hat{\beta}_c}\right)^2 \hat{\nu}_k + \left(\frac{\hat{\beta}_k}{\hat{\beta}_c^2}\right)^2 \hat{\nu}_c + 2\left(\frac{-1}{\hat{\beta}_c}\right)\left(\frac{\hat{\beta}_k}{\hat{\beta}_c^2}\right)\hat{\nu}_{kc},\end{aligned} \tag{3.2}$$

where $\widehat{\text{WTP}}_{\beta_k}$ and $\widehat{\text{WTP}}_{\beta_c}$ are the partial derivatives of WTP_k with respect to β_k and β_c respectively, evaluated at the point estimates.

The Delta method then assumes that the WTP is normally distributed, and thus symmetrical about its mean. As such, confidence intervals are created in the conventional manner:

$$\text{CI}_{\text{delta}} = \widehat{\text{WTP}}_k \pm z_{1-\alpha/2}\sqrt{\text{var}(\widehat{\text{WTP}}_k)}, \tag{3.3}$$

The Delta method is simple to use due to its linear approximation. It often produces CIs that are comparatively narrower than those of other methods, and so is seen as more accurate. In addition, though the variance of the WTP is given here (and in much of literature) as the first order Taylor expansion, Daly et al. (2012) argues that this variance obtained is in fact the accurate, exact variance of the ratio of parameters. However, as discussed above, the assumption that the WTP is normally distributed only holds when a sufficiently large sample is used, and the standard error of β_c is small compared to its mean. As these conditions might not necessarily hold, the assumption of normality of the WTP is strong. The Delta method also depends on a continuous cost coefficient, and fails in the presence of weak identification. In addition, there is an assumption that the CI is symmetric about its mean, which often does not occur in practice. Finally, a narrow CI might not necessarily be beneficial as it might result in low coverage.

Fieller Method

As with the Delta method, the Fieller method (Fieller, 1944, 1954) assumes the consistent asymptotic normal distribution of the parameter values. However, it does not assume that the WTP is itself normally distributed, but that the coefficients are joint normally distributed. The Fieller method relies on the fact that a linear combination of normal random variables is itself normal. In particular, $\text{WTP}_k = \delta = \frac{\hat{\beta}_k}{\hat{\beta}_c}$ implies that $\hat{\beta}_k - \delta\hat{\beta}_c = 0$, and thus

$$\hat{\beta}_k - \delta\hat{\beta}_c \sim \text{N}(0, \sigma^2_{WTP}) \tag{3.4}$$

where $\sigma^2_{WTP} = \delta^2\nu_c - 2\delta\nu_{kc} + \nu_k$ and δ is used in place of WTP_k for brevity.

Dividing by the estimator of the standard deviation of WTP gives the statistic:

$$T_0 = \frac{\hat{\beta}_k - \delta\hat{\beta}_c}{\sqrt{(\hat{\delta}^2_k\hat{\nu}_c - 2\hat{\delta}_k\hat{\nu}_{kc} + \hat{\nu}_k)}} \tag{3.5}$$

which follows approximately or exactly a Student t-distribution with df degrees of freedom. In most cases, however, the relationship is approximate, with the t-distribution corresponding to a normal distribution with d$f = \infty$. Franz (2007) describes the following conditions that need to be met for an exact relationship:

1. $(\hat{\beta}_k, \hat{\beta}_c)$ is exactly normally distributed
2. The covariance matrix is known up to a proportionality constant σ^2
3. σ^2 can be estimated by $\hat{\sigma}^2$ independent of $(\hat{\beta}_k, \hat{\beta}_c)$, such that $\frac{df\hat{\sigma}^2}{\sigma^2}$ is distributed chi-square with df degrees of freedom

If these three conditions are met, then the t-distribution in equation (3.5) has df degrees of freedom. However, discrete choice models contain no finite-sample counterpart for this test, hence the relationship is approximate and the normal distribution is used.

The $(1 - \alpha)$ CI corresponds to inverting the test $\hat{\beta}_k - \delta\hat{\beta}_c = 0$ with respect to δ, i.e. finding the values of δ such that $T_0^2 \leq z^2_{\alpha/2}$. Let $A = \hat{\beta}^2_c - z^2_{\alpha/2}\hat{\nu}_c$,

$B = -\hat{\beta}_k\hat{\beta}_c + z^2_{\alpha/2}\hat{\nu}_{kc}$ and $C = \hat{\beta}^2_k - z^2_{\alpha/2}\hat{\nu}_k$. Then there are three cases which must be considered.

$\boldsymbol{A > 0}$. This is equivalent to saying that $\hat{\beta}_c$ is significantly different from 0 (i.e. $\frac{\hat{\beta}^2_c}{\hat{\nu}_c} > z^2_{\alpha/2}$). Then the following bounded CI is obtained:

$$(\underline{\delta}, \bar{\delta}) = (\frac{-B - \sqrt{B^2 - AC}}{A}, \frac{-B + \sqrt{B^2 - AC}}{A}) \tag{3.6}$$

$\boldsymbol{A < 0}$ and $\boldsymbol{B^2 - AC > 0}$. If $\hat{\beta}_c$ is not significantly different from 0, then we can differentiate between two cases. This first case results in a CI that includes all values except those between $\underline{\delta}$ and $\bar{\delta}$ as they are defined above.

$\boldsymbol{A < 0}$ and $\boldsymbol{B^2 - AC < 0}$. This final case produces a CI that is the entire real line.

Note that no other case would exist, i.e. $A > 0$ and $B^2 - AC < 0$ cannot occur at the same time (Bolduc et al., 2010). This is because of the Cauchy-Schwartz inequality (Rao, 1973, pg. 55), where $\hat{\nu}^2_{kc} - \hat{\nu}_k\hat{\nu}_c < 0$. In particular, consider

$$\begin{aligned}
B^2 - AC < 0 \;&\Rightarrow (-\hat{\beta}_k\hat{\beta}_c + z^2_{\alpha/2}\hat{\nu}_{kc})^2 - (\hat{\beta}^2_c - z^2_{\alpha/2}\hat{\nu}_c)(\hat{\beta}^2_k - z^2_{\alpha/2}\hat{\nu}_k) \\
&\qquad = z^4_{\alpha/2}(\hat{\nu}^2_{kc} - \hat{\nu}_k\hat{\nu}_c) + z^2_{\alpha/2}(\hat{\beta}^2_k\hat{\nu}_c + \hat{\beta}^2_c\hat{\nu}_k - 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_{kc}) < 0 \\
&\Leftrightarrow z^2_{\alpha/2}(\hat{\beta}^2_k\hat{\nu}_c + \hat{\beta}^2_c\hat{\nu}_k - 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_{kc}) < -z^4_{\alpha/2}(\hat{\nu}^2_{kc} - \hat{\nu}_k\hat{\nu}_c) \\
&\Leftrightarrow \frac{\hat{\beta}^2_k\hat{\nu}_c + \hat{\beta}^2_c\hat{\nu}_k - 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_{kc}}{\hat{\nu}_k\hat{\nu}_c - \hat{\nu}^2_{kc}} < z^2_{\alpha/2}
\end{aligned} \tag{3.7}$$

Let $z^* = \frac{\hat{\beta}_k^2\hat{\nu}_c + \hat{\beta}_c^2\hat{\nu}_k - 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_{kc}}{\hat{\nu}_k\hat{\nu}_c - \hat{\nu}_{kc}^2}$, and consider

$$\begin{aligned}
\frac{\hat{\beta}_c^2}{\hat{\nu}_c} - z^* &= \frac{\hat{\beta}_c^2}{\hat{\nu}_c} - \frac{\hat{\beta}_k^2\hat{\nu}_c + \hat{\beta}_c^2\hat{\nu}_k - 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_{kc}}{\hat{\nu}_k\hat{\nu}_c - \hat{\nu}_{kc}^2} \\
&= \frac{-\hat{\beta}_c^2\hat{\nu}_{kc}^2 + 2\hat{\beta}_k\hat{\beta}_c\hat{\nu}_c\hat{\nu}_{kc} - \hat{\beta}_k^2\hat{\nu}_c^2}{\hat{\nu}_c(\hat{\nu}_k\hat{\nu}_c - \hat{\nu}_{kc}^2)} \\
&= -\frac{(\hat{\beta}_c\hat{\nu}_{kc} - \hat{\beta}_k\hat{\nu}_c)^2}{\hat{\nu}_c(\hat{\nu}_k\hat{\nu}_c - \hat{\nu}_{kc}^2)} < 0 \\
\Rightarrow & \frac{\hat{\beta}_c^2}{\hat{\nu}_c} < z^*
\end{aligned} \tag{3.8}$$

Thus $B^2 - AC < 0 \Rightarrow \frac{\hat{\beta}_c^2}{\hat{\nu}_c} < z^* < z_{\alpha/2}^2 \Rightarrow \frac{\hat{\beta}_c^2}{\hat{\nu}_c} < z_{\alpha/2}^2 \Rightarrow \hat{\beta}_c^2 - z_{\alpha/2}^2\hat{\nu}_c < 0 \Rightarrow A < 0$. As a result, $A > 0 \Rightarrow B^2 - AC > 0$.

The Fieller method is not constrained by the assumption that the WTP is normally distributed, nor does it assume that the point estimate must be at the center of the CI. However, because it relies on an asymptotic distribution, its reliability might be questionable under finite samples. In addition, there is a big debate regarding the practical interpretation of the unbounded CIs produced when $\hat{\beta}_C$ is not significantly different from 0. Section 4.1 will further analyze the question of bounded versus unbounded CIs.

In his original paper, Fieller (1954) discusses the general case of inverting a Wald-type test associated with a hypothesis of more than one degree, i.e. when looking at equations of the type

$$\beta_1 F_1(\alpha) + \beta_2 F_2(\alpha) + \cdots = 0 \tag{3.9}$$

When looking at a single WTP measure, we are thus inverting the univariate Wald test.

Some authors (Armstrong et al., 2001; Bernard et al., 2007) have also suggested the inversion of the Likelihood Ratio test (which I will call the LR method), as an alternative to the Fieller method. Armstrong et al. (2001) found that the Fieller and LR method gives comparable results. However, recall that the Wald test, LR test and the Lagrange Multiplier (LM) test are asymptotically equivalent. It can thus be argued that the LR method is essentially equivalent to, or an extension of, the Fieller method.

Krinsky-Robb Method

The Krinsky-Robb (KR) method (Krinsky and Robb, 1986, 1990) creates a multivariate normal distribution using the mean and covariance matrix of the estimated coefficients, namely $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{kc}$. A large number of draws is taken from this distribution, with WTP estimates calculated from each draw. Sorting the estimates and taking the $100(\alpha/2)^{th}$ and $100(1-\alpha/2)^{th}$ percentile values give the KR $(1-\alpha)$ CI.

While the KR method also assumes a joint normal distribution for the coefficients, it is versatile as one can apply this method to any utility function specification. It can also be used for any linear or nonlinear function of the estimated parameters, and accounts for (co-)variability associated with all coefficients. However, this is perhaps the most computationally demanding of all frequentist methods studied in this paper due to resampling.

The KR method is in fact just one sampling technique classified under the family of bootstrap methods. In general, bootstrap methods simulate a distribution of the variable in question, here the WTP measure. The most common bootstrap sampling method does not assume a multivariate normal distribution for the coefficients, but draws a large number of samples, say N, from the data with replacement. Each sample is then solved and the WTP measure is calculated using equation (2.22). The CI is then calculated using percentiles. This general bootstrap has a clear advantage in that it assumes nothing about the symmetry of the WTP distribution, or the distribution of the estimated coefficients. However, it is even more computationally demanding than the KR method, since the model has to be estimated N times. In addition, research within the Economics and Transportation Science fields have found the bootstrap to be either worse or no different from other methods being used in this work (Armstrong et al., 2001; Hole, 2007; Bolduc et al., 2010).

WTP-Space Method

Working in WTP-space involves re-parameterizing the full conditional likelihood so that WTP can be directly measured. Consider the indirect utility function:

$$U_{in} = V_{in} + \varepsilon_{in} = x_{cin}\beta_c + x_{1in}\beta_1 + \ldots + x_{Kin}\beta_K + \varepsilon_{in} \tag{3.10}$$

Rewriting this equation gives the utility model in WTP-space, i.e.

$$\begin{aligned} U_{in} &= x_{cin}\beta_c + x_{1in}\frac{\beta_1}{\beta_c}\beta_c + \ldots + x_{Kin}\frac{\beta_K}{\beta_c}\beta_c + \varepsilon_{in} \\ &= \beta_c(x_{cin} + x_{1in}\text{WTP}_1 + \ldots + x_{Kin}\text{WTP}_K) + \varepsilon_{in} \end{aligned} \tag{3.11}$$

Note that this is equivalent to the consumer surplus model, in which equation (3.10) is divided throughout by the coefficient of cost β_c to obtain the consumer surplus of individual n for alternative i, $\mathbb{C}_{in}$:

$$\begin{aligned} \frac{U_{in}}{\beta_c} &= x_{cin} + x_{1in}\frac{\beta_1}{\beta_c} + \ldots + x_{Kin}\frac{\beta_K}{\beta_c} + \frac{\varepsilon_{in}}{\beta_c} \\ \mathbb{C}_{in} &= x_{cin} + x_{1in}\text{WTP}_1 + \ldots + x_{Kin}\text{WTP}_K + \eta_{in} \end{aligned} \tag{3.12}$$

As before, the individual chooses an alternative that maximizes his utility (or surplus) among the J alternatives available. The MNL choice probability associated with this model is given by

$$P_{in} = \left[\frac{\exp\{\beta_c(x_{cin} + \mathbf{x}'_{in,-c}\mathbf{WTP})\}}{\sum_{j=1}^{J} \exp\{\beta_c(x_{cjn} + \mathbf{x}'_{jn,-c}\mathbf{WTP})\}} \right], \tag{3.13}$$

where $\mathbf{x}'_{in,-c}$ are all the attributes excluding the attribute of cost. Equation (3.13) allows the WTP measures to be directly estimated from the model. The frequentist CIs are then obtained using standard CI equations:

$$\text{CI}_{\text{WTP Space}} = \widehat{\text{WTP}}_k \pm z_{\alpha/2}\sqrt{\text{var}(\widehat{\text{WTP}}_k)} \tag{3.14}$$

where $\widehat{\text{WTP}}_k$ is the point estimate of the WTP of attribute k estimated directly from the consumer surplus model.

An obvious benefit to re-parameterizing the model into WTP-space is the ability to directly implement a prior for the WTP measure. No additional steps are needed to calculate WTP from model estimates, and the CIs are easy to build. Sonnier et al. (2007) show that solving a model in WTP-space also gives a more accurate recovery of the true WTP than solving it in preference space, even if the data was originally produced using the preference space. However, since to our knowledge this method has not been used to build CIs, it is as yet unknown how beneficial this method will prove to be.

3.1.2 Bayesian Methods

Bayesian methods are not often discussed or used in estimating discrete choice models, let alone in building CIs for WTP measures. Yet the portrayal of CIs in the Bayesian context is in fact more intuitive than that in the frequentist context. The Bayesian model assumes that the parameter is random, which differs from the frequentist context where the parameter is assumed to be a fixed, true parameter. This plays into the interpretation of the CI that is built. Consider $\alpha = 0.05$, or a 95% confidence level. In the frequentist context, the "95% confidence interval" means that over repeated sampling, 95% of the CIs built will contain the fixed parameter. However, in the Bayesian context, the random parameter has a 95% probability of falling into the interval (called *credible interval*). This latter interpretation is actually how many people think of all CIs, though it is not an accurate portrayal of the CI in the frequentist context.

There are some who might argue that the frequentist CI and the Bayesian credible interval should not be compared. In this work I compare between Bayesian and frequentist methods by using a "frequentist" aproach, i.e. the simulation is repeated multiple times, and a coverage is calculated over this repetitions. This simulated coverage gives us one basis of comparison between the different methods.

There are two main Bayesian methods being discussed in this section. The first is the Bayesian post-processing method, while the second is the method of inverting the Bayes Factor.

Bayesian Post-Processing Method

The Bayesian post-processing method can be used either in preference space or WTP-space. In either space, the respective model is solved using the Bayesian method, which uses Markov Chain Monte Carlo (MCMC) simulations. The

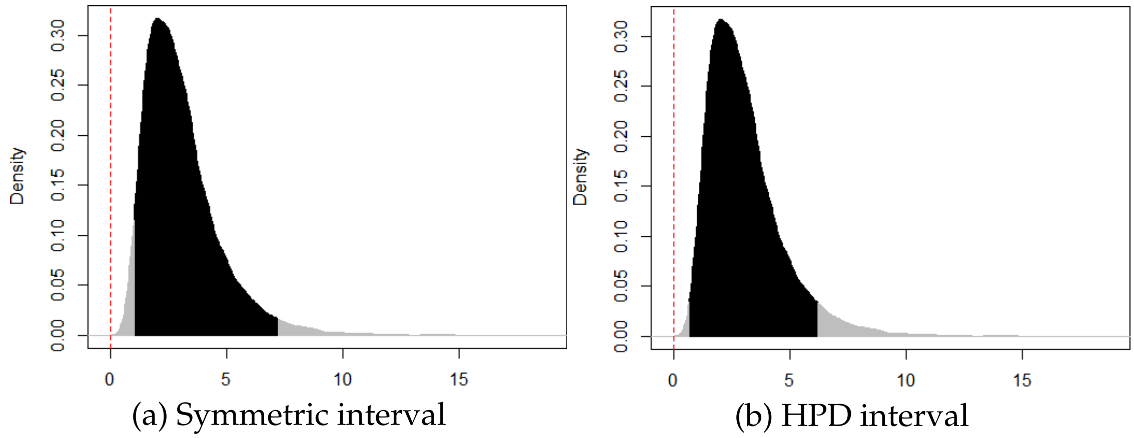


Figure 3.1: Comparing symmetric and HPD intervals

posterior solutions are generated by R iterations of the MCMC sample. In the preference space, the WTP estimate for each realization of the MCMC chain $r = 1, \ldots, R$ can be calculated as $WTP_k^r = \frac{\tilde{\beta}_k^r}{\tilde{\beta}_c^r}$, while in the WTP-space, the WTP estimates are directly solved by the model. The credible interval is then calculated by taking the highest probability density (HPD) $1 - \alpha$ interval of the WTP estimates.

This method of building CIs can be thought of as the Bayesian counterpart to the KR method. However, the Bayesian method is less involved because estimating the model parameters uses realizations of the MCMC sample. Hence, no additional sampling is needed as in the KR method; instead the MCMC sample can be used directly. In addition, no assumptions are made regarding the distribution of WTP, and like the KR method, the Bayesian method can be applied to any utility function specification and any linear or nonlinear function of the estimated parameters. Finally, the MCMC chain is sampled directly from the posterior distribution of the parameters, and so does not require a large sample size in order to obtain an asymptotic distribution as with estimates obtained through MLE.

Note that there are many ways in which intervals can be formed. The standard method of deriving the credible interval would be to take the $100(\alpha/2)$th and $100(1-\alpha/2)$th percentile values of the sorted WTP estimates. This method is accurate when the distribution of the random variable is symmetric. However, because it chooses the interval bounds such that there is an equal amount of probability mass on either side of the interval (here $\alpha/2$), it gives a less plausible interval for when the distribution itself is not symmetric. Instead, the HPD interval can be used, which accounts for the asymmetry and has been shown to give the shortest interval for any specific probability $1-\alpha$ (Box and Tiao, 1992). As can be seen from Figure 3.1, the HPD interval is shifted further to the right as compared to the symmetric credible interval, due to the fact that the right side of the probability distribution has a greater probability mass. In addition, if the distribution is in fact symmetric, then the HPD and symmetric intervals will be identical. Hence in order to account for possible asymmetrical variable distributions, and also to compare with the shortest interval possible, I construct HPD intervals in this work.

Inverting the Bayes Factor[1]

The Bayes factor was first developed by Jeffreys (1935, 1961), and well explained and reviewed by Kass and Raftery (1995). Supposing, as in the Bayesian context, we have data D that is assumed to have come about under either hypothesis H_1 or H_2 according to probability densities $\mathbb{P}(D|H_1)$ and $\mathbb{P}(D|H_2)$ respectively (these are also known as the marginal likelihoods). Given prior probabilities $\mathbb{P}(H_1)$ and $\mathbb{P}(H_2)$, we can obtain the posterior probabilities using Bayes Theo-

[1]Note that this method is here explained as a possible method for building CIs, but is not studied in this research due to computational constraints.

rem,

$$\mathbb{P}(H_1|D) = \frac{\mathbb{P}(D|H_1)\mathbb{P}(H_1)}{\mathbb{P}(D)} = \frac{\mathbb{P}(D|H_1)\mathbb{P}(H_1)}{\mathbb{P}(D|H_1)\mathbb{P}(H_1) + \mathbb{P}(D|H_2)\mathbb{P}(H_2)} \tag{3.15}$$

Taking the ratio of the posterior probabilities gives

$$\frac{\mathbb{P}(H_1|D)}{\mathbb{P}(H_2|D)} = \frac{\mathbb{P}(D|H_1)}{\mathbb{P}(D|H_2)}\frac{\mathbb{P}(H_1)}{\mathbb{P}(H_2)} \quad , \tag{3.16}$$

where the Bayes factor in favor of H_1 is the ratio of the marginal likelihoods, i.e. $B_{12} = \frac{\mathbb{P}(D|H_1)}{\mathbb{P}(D|H_2)}$. In words, posterior odds $=$ Bayes factor $\times$ prior odds.

Note that one of the advantages of the Bayes factor is that, in contrast to frequentist hypothesis testing, the Bayes factor provides a way of evaluating evidence in favor of the null hypothesis. In frequentist hypothesis testing, an analyst can only conclude one of two things: either she rejects the null hypothesis in favor of the alternative, or she fails to reject the null hypothesis. However, the Bayes factor allows the analyst to favor the null hypothesis.

Consider a case where the two hypotheses being tested are simple, i.e. a hypothesis that is a single distribution with no free parameters. An example would be a test where $H_0 : \theta_1 = 0.2$ vs. $H_1 : \theta_1 = 0.3$. In this simplest case, the Bayes factor is simply the likelihood ratio. When either one or both hypotheses are not simple, then the probability density $\mathbb{P}(D|H_k)$ is obtained by integrating (rather than maximizing) over the parameter space, as in

$$\mathbb{P}(D|H_k) = \int \mathbb{P}(D|\theta_k, H_k)\mathbb{P}(\theta_k|H_k)d\theta_k \tag{3.17}$$

where θ_k is the parameter under hypothesis H_k.

Obtaining the marginal likelihood is often computationally intractable, and many authors have written and proposed methods with which to calculate this value. Oft-used methods include Laplace's method of approximation (de Bruijn, 1970, sec. 4.4; Tierney and Kadane, 1986) and the weighted likelihood bootstrap (Newton and Raftery, 1994). The accurate computation of the marginal likelihood is beyond the scope of this paper, and a good review can be found in Kass and Raftery (1995).

Specific to this problem of building CIs for WTP values, we consider the following hypotheses:

$$H_0 : \frac{\hat{\beta}_k}{\hat{\beta}_c} = \delta \qquad \text{vs.} \qquad H_1 : \frac{\hat{\beta}_k}{\hat{\beta}_c} \neq \delta \tag{3.18}$$

The asymptotic distribution of $-2\ln(B_{01})$ is $\chi^2_{1,1-\alpha}$. Assuming that this Bayes factor has been obtained, the CI is then built by inverting the test statistic with respect to δ, i.e. finding the values of δ such that $B_{01} \leq e^{\chi^2_{1,1-\alpha}/2}$. The minimum and maximum values of δ that follow this criterion are the lower and upper bounds of the CI respectively.

While in theory this method appears to work well, studying all possible values of δ over the entire parameter space is impossible. Hence the space from which δ is chosen can be taken from the posterior MCMC sample to reduce the computational burden. Using the MCMC realizations is justified because these become dense in the parameter space as the number of draws increases (Guerron-Quintana et al., 2013). The steps for building a CI using the Bayes factor proceeds as follows:

1. Estimate the unrestricted (H_1) model $\mathbf{U}_n = \mathbf{x}^{'}_{cn}\beta_c + \mathbf{x}^{'}_{1n}\beta_1 + \ldots + \mathbf{x}^{'}_{kn}\beta_k + \ldots + \mathbf{x}^{'}_{Kn}\beta_K$ and obtain posterior sample $\tilde{\boldsymbol{\beta}}^r$ for realizations $r = 1, \ldots, R$.
2. Calculate $WTP^r_k = \frac{\tilde{\beta}^r_k}{\tilde{\beta}^r_c}$.

3. Estimate the null (H_0) model $\mathbf{U}_n = \beta_c(\mathbf{x}_{cn}^{'} + WTP_k^r \mathbf{x}_{kn}^{'}) + \mathbf{x}_{1n}^{'}\beta_1 + \ldots + \mathbf{x}_{k-1,n}^{'}\beta_{k-1} + \mathbf{x}_{k+1,n}^{'}\beta_{k+1} + \ldots + \mathbf{x}_{Kn}^{'}\beta_K$.
4. Calculate the Bayes factor $B_{01} = \frac{\mathbb{P}(D|H_0)}{\mathbb{P}(D|H_1)}$.
5. If $B_{01} \leq e^{\chi^2_{1,1-\alpha/2}}$, then keep WTP_k^r. Else discard.
6. Repeat steps 2 - 5 for all $r = 1, \ldots, R$.
7. $\text{CI}_{\text{BF}} = (\min_{(r)} WTP_k^{(r)}, \max_{(r)} WTP_k^{(r)})$ for all (r) that were kept in step 5.

Similar to the Fieller method, building CIs this way involves inverting a test statistic, and hence this method can be seen as equivalent to the Fieller method, or the LR method. In fact, under certain conditions, inverting the Bayes factor is exactly the LR test. One might actually think of this method as a frequentist one, due to the type of hypothesis being studied in this problem (equation 3.18). Under Bayesian assumptions, where model parameters are assumed random, the probability that the null hypothesis holds is 0. Hence, it does not actually make sense to invert this test statistic from a Bayesian point of view. One is thus applying a frequentist method to a Bayesian statistic with this method.

Having said that, I classify this as a Bayesian method since it not only utilizes the Bayes factor, which is a Bayesian statistic used to compare models and hypotheses, but it also uses the the MCMC realizations of the posterior sample as described above.

3.2 Random Coefficient Models

This section investigates the methods by which CIs can be built for WTP measures obtained under assumptions for unobserved heterogeneity. Using random coefficients adds difficulty to the problem, as WTP is no longer a single ratio of coefficient estimates, and hence the methods used must be modified appropriately. The Delta method has been modified and proposed in the literature, while other modifications (such as to the Krinsky-Robb and Bayesian post-processing methods) are easily applied, though computationally intensive.

3.2.1 Frequentist Methods

Delta Method

The Delta method for random parameter models has been applied in the Transportation Science primarily by Bliemer and Rose (2012), who adapt the Delta method for random coefficient logit models. Suppose we have independently distributed random coefficients, and let each β_k and β_c follow some distribution with a vector of parameters θ_k. For example, if $\beta_k \sim N(\mu_k, \sigma_k^2)$, then $\theta_k = (\mu_k, \sigma_k^2)$. Estimating the random coefficient model will then give estimates of these distributional parameters, $\hat{\theta}_k$.

In order to use the Delta method, we first map the standard errors (and covariances) of $\hat{\theta}_k$ to a standard error of β_k, and determine the standard error of ${}^{\beta_k}/_{\beta_c}$. This is done by rewriting the coefficients β_k and β_c into functions of θ_k and θ_c using parameter free distributions (i.e. standard distributions), as in

$$\beta_k = \beta_k(a_k|\theta_k), \tag{3.19}$$

where a_k is a standard probability distribution. In the example where $\beta_k \sim N(\mu_k, \sigma_k^2)$, we can write $\beta_k = \mu_k + \sigma_k a_k$, where $a_k \sim N(0,1)$. Then, the WTP can be written as

$$\text{WTP}_k(a_k, a_c|\theta_k, \theta_c) = \frac{\beta_k(a_k|\theta_k)}{\beta_c(a_c|\theta_c)} \tag{3.20}$$

Conditional on some draw of a_k, a_c, we apply the Delta method to obtain

$$\widehat{\text{WTP}}_k(a_k, a_c) \overset{D}{\to} N\left(\text{WTP}_k, \begin{pmatrix} \nabla_{\theta_k}\text{WTP}_k \\ \nabla_{\theta_c}\text{WTP}_k \end{pmatrix}^T \Omega_{\theta_{kc}} \begin{pmatrix} \nabla_{\theta_k}\text{WTP}_k \\ \nabla_{\theta_c}\text{WTP}_k \end{pmatrix}\right), \tag{3.21}$$

where $\nabla_{\theta_k}\text{WTP}_k$ and $\nabla_{\theta_c}\text{WTP}_k$ are the Jacobians of the WTP with respect to θ_k and θ_c, evaluated at the true values of the parameters, respectively, and $\Omega_{\theta_{kc}}$ is the submatrix of the variances and covariances of distributional parameters θ_k and θ_c.

The Jacobians can be calculated

$$\begin{aligned} \nabla_{\theta_k}\text{WTP}_k &= \frac{1}{\beta_c}\nabla_{\theta_k}\beta_k \\ \nabla_{\theta_c}\text{WTP}_k &= -\frac{\text{WTP}_k}{\beta_c}\nabla_{\theta_c}\beta_c \end{aligned}, \tag{3.22}$$

and thus (3.21) can be written as

$$\widehat{\text{WTP}}_k(a_k, a_c) \overset{D}{\to} N\left(\text{WTP}_k, \frac{1}{\beta_c^2}\begin{pmatrix} \nabla_{\theta_k}\beta_k \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c \end{pmatrix}^T \Omega_{\theta_{kc}} \begin{pmatrix} \nabla_{\theta_k}\beta_k \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c \end{pmatrix}\right) \tag{3.23}$$

Since this is the WTP estimate conditional on a_k, a_c, the unconditional expected WTP estimate is then defined by integrating over the whole parameter space of a_k, a_c, as follows

$$\widehat{\text{WTP}}_k = \int_{a_k}\int_{a_c} \widehat{\text{WTP}}_k(a_k, a_c) dF_k(a_k) dF_c(a_c), \tag{3.24}$$

where $dF_k(a_k)$ and $dF_c(a_c)$ are the cumulative distribution functions of a_k and a_c respectively. Depending on the assumed parameter distribution, this can be a multidimensional integral that is difficult to evaluate. As such, it is often approximated by the Monte Carlo simulation

$$\widehat{\text{WTP}}_k \approx \frac{1}{R}\sum_{r=1}^{R}\widehat{\text{WTP}}_k(a_k^{(r)}, a_c^{(r)}) \tag{3.25}$$

Since, by the assumptions of the Delta method, $\widehat{\text{WTP}}_k(a_k^{(r)}, a_c^{(r)})$ is asymptotically normally distributed, $\widehat{\text{WTP}}_k$ will be normally distributed with the following simulated variance:

$$\text{var}(\widehat{\text{WTP}}_k) \approx \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{(\beta_c^{(r)})^2}\begin{pmatrix}\nabla_{\theta_k}\beta_k^{(r)} \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c^{(r)}\end{pmatrix}^T \Omega_{\theta_{kc}}\begin{pmatrix}\nabla_{\theta_k}\beta_k^{(r)} \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c^{(r)}\end{pmatrix}\right), \tag{3.26}$$

and the $(1-\alpha)$ confidence interval then becomes

$$\text{CI}_{\text{Delta, Mean}} = \widehat{\text{WTP}}_k \pm z_{1-\alpha/2}\sqrt{\text{var}(\widehat{\text{WTP}}_k)} \tag{3.27}$$

Due to the normality assumption of the WTP value, the integral in equation (3.24) becomes undefined at $\beta_c = 0$. This problem can be avoid either by assuming a parameter distribution for cost that has no probability mass at $\beta_c = 0$ (e.g. lognormal distribution), or by using the median rather than the mean to obtain the WTP estimate, i.e. equations (3.25) and (3.26) respectively become

$$\widehat{\text{WTP}}_k \approx \underset{r}{\text{median}}\left(\widehat{\text{WTP}}_k(a_k^{(r)}, a_c^{(r)})\right) \tag{3.28}$$

$$\text{var}(\widehat{\text{WTP}}_k) \approx \underset{r}{\text{median}}\left(\frac{1}{(\beta_c^{(r)})^2}\begin{pmatrix}\nabla_{\theta_k}\beta_k^{(r)} \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c^{(r)}\end{pmatrix}^T \Omega_{\theta_{kc}}\begin{pmatrix}\nabla_{\theta_k}\beta_k^{(r)} \\ -\text{WTP}_k\nabla_{\theta_c}\beta_c^{(r)}\end{pmatrix}\right) \tag{3.29}$$

The assumptions of the Delta method for the fixed coefficient model hold also for the random coefficient model. In addition to those assumptions, the

Delta method requires less simulation than other methods (in particular, the KR method). In the example where the coefficients are distributed normal, the Delta method requires only two dimensions of simulation (simulating normal variates z_k and z_c), while the KR method requires six (as described below). As such, the Delta method is computationally less intensive than the KR method.

Krinsky-Robb Method

Adapting the KR method for random coefficient models is not difficult. As with the Delta method for random coefficient models, the coefficients β_k and β_c are written in terms of the distributional parameters θ_k, θ_c such that the WTP measure can be expressed as equation (3.20).

The KR method then creates a multivariate normal distribution using the mean and covariance matrix of the estimated distributional parameters, i.e. $\hat{\boldsymbol{\theta}}$ and $\hat{\Sigma}_{\boldsymbol{\theta}}$. Note that this includes the estimated means and covariance matrix of all the distributional parameters. As with the fixed coefficient model, a large number of draws are taken from this distribution. In addition, random draws from the standard probability distribution must be taken for each draw from the multivariate normal distribution. The WTP measure is then calculated for each draw as in equation (3.20), and as before these estimates are sorted and the $100(\alpha/2)^{th}$ and $100(1-\alpha/2)^{th}$ percentile values give the KR$(1-\alpha)$ CI.

Note how, as argued by Bliemer and Rose (2012), the KR method is much more computationally intensive. For the example where the coefficients are normally distributed, the KR method requires six dimensions of simulation (2 for each coefficient, and 2 for the simulated normal variates).

3.2.2 Bayesian Methods

Bayesian Post-Processing Method

As with the KR method, adapting the Bayesian post-processing method for random coefficient models is not difficult. Individual draws are taken from the MCMC posterior sample for those attributes that are assumed to be random. That is, the Bayesian estimation would take R draws per individual for each random attribute. The WTP measure can then be calculated using these individual draws following equation (3.20), resulting in $N \times R$ WTP values.

There is no fixed way by which to generate the aggregate CI from these values. Following the example of Daziano and Achtnicht (2014), I treat all WTP samples as having come from one Markov Chain sample of the same posterior. Hence the HPD interval is taken over all $N \times R$ WTP samples to obtain the aggregate CI.

As ca be seen, it is also straightfoward to build the CIs on the individual level. The HPD interval can be taken over the R draws for each individual to obtain the individual level CI. This will be demonstrated in Chapter 6.

CHAPTER 4

LITERATURE REVIEW

The interval estimation problem has been studied in a variety of fields and research areas, including biomedical science, information science, transportation and economics. This chapter gives an in-depth review to some of these studies that have been undertaken. In applied work, each field appears to use a specific method of building CIs by default, though few papers give a reason for using that method as opposed to others. Table 4.1 gives a few examples to highlight the wide range of methods used.

Table 4.1: Methods used to build confidence intervals

Author	Field	Method Primarily Used
Park et al. (1991)	Land Economics	Krinsky-Robb
Armstrong et al. (2001)	Transportation	Fieller
Franz (2003)	Psychology	"index" method
Beyene and Moineddin (2005)	Biomedical	Generalized linear modeling

There is an even greater contrast within the literature that compare different methods through simulation. Table 4.2 gives a quick outline of the conclusions reached by a number of authors who, through simulation, have studied the different methods of building CIs for WTP measures. It is clear that while many of them study the same methods of building CIs, few of them agree with each other. These differences are not due to the preference for a certain method within a field, as even within areas of study, the authors do not agree on the best method.

Other authors aim to introduce fresh interpretations to methods that have been disregarded, and often include reasons as to why this method is superior.

Table 4.2: Conclusions from studies of confidence intervals. BS = bootstrap, D = Delta, F = Fieller, KR = Krinsky-Robb, LR = likelihood-ratio test, GLM = generalized linear model

Author	Field	Methods Used	Conclusion
Armstrong et al. (2001)	Transportation	BS, D, F, KR, LR	KR, F, LR give similar results; superior to BS, D
Beyene and Moineddin (2005)	Biomedical Science	D, F, GLM	All give similar results for large samples; GLM performs best for small samples
Hole (2007)	Health Economics	BS, D, F, KR	Any method can be used in most situations
Bernard et al. (2007)	Econometrics	D, F, LR	F performs better than D
Franz (2007)	Psychology	BS, D, F, others	Fieller or Hwang-bootstrap for insignificant cost coef; any (of these 3) for significant cost coef
Bolduc et al. (2010)	Transportation	BS, D, F	F performs extremely well; BS, D have poor coverage, should be avoided
Gatta et al. (2013)	Statistical Science	BS, D, F, KR, LR	F is generally the best; use LR if model is known to be correctly specified; with large sample size, all methods perform equally

Table 4.3 cites some of these papers, and the conclusions drawn from them. Of particular interest are the claims of Daly et al. (2012) and Bliemer and Rose (2012) in support of the Delta method, which prior to this had mostly been disregarded for the reasons listed in section 3.1.1 (most of the authors in Table 4.2 conclude that the Delta method performs poorly, or at best as well as the other methods being compared). However, Daly et al. (2012) supports the Delta method, showing through the work of Cramer (1986) that the Delta method gives exact standard errors, rather than approximate as is often claimed. They also demonstrated the superiority of the Delta method to simulation methods of obtaining the standard errors. Bliemer and Rose (2012) build on this argument, stating that the Delta method is preferable not only for its exact standard errors, but also because it typically produces narrower CIs (indicating greater accuracy) and it requires less simulation (as compared to the KR method).

Table 4.3: Proposals for new methods of building confidence intervals / new interpretations of methods. BS = bootstrap, D = Delta, F = Fieller, KR = Krinsky-Robb, WTP-Sp = WTP-space

Author	Field	Method Proposed	Additional claims
Siani and de Peretti (2003)	Health Economics	Geometric interpretation of F	F is applicable in all situations
Sonnier et al. (2007)	Marketing & Economics	WTP-Sp	WTP-Sp is superior to preference space
von Luxburg and Franz (2009)	Psychology	Geometric interpretation of F	F is superior to BS
Daly et al. (2012)	Transportation	D	D is superior to KR
Bliemer and Rose (2012)	Transportation	D for mixed logit model	D is easier to implement than KR

Adding to the confusion in the literature is the fact that what one author claims to be a benefit for a method is seen as a disadvantage by other authors. For example, Bliemer and Rose (2012) praise the Delta method for having narrow CIs, while Armstrong et al. (2001) claim that the Delta method produces CIs that are, in contrast, *too* narrow, and Bernard et al. (2007) agree, saying that the Delta method may have "zero coverage probability", i.e. the probability that the CI does not contain the true parameter may be one (Dufour, 1997).

Perhaps the most comprehensive discussion concerning this topic to date is by Gatta et al. (2013). These authors conduct an excellent review and analysis of various methods of building CIs, though only for the frequentist methods. In particular, the authors explain and analyze the Delta, Fieller (or asymptotic *t*-test), likelihood ratio, as well as eight different bootstrap methods (Table 4.4). In addition, they also explain the different algorithms of drawing samples for the bootstrap methods; these are the parameteric bootstrap, the non-parametric bootstrap, and the Krinsky-Robb algorithms. In order to analyze these methods of building CIs, the authors conducted simulations similar to that of Hole (2007). They simulated a number of individuals each facing 16 choice situations where they had to choose between 2 alternatives, each containing 3 attributes

Table 4.4: Summary of findings by Gatta et al. (2013)

Method	Findings
Delta	Produces symmetric CIs by construction, thus cannot account for asymmetry of WTP distribution
Fieller (*t*-test)	Good performance, not affected by small cost parameter, simple, time-saving to calculate
Likelihood-ratio	Good performance, not affected by small cost parameter, simple, time-saving to calculate; usually narrower than Filler CIs, but more sensitive to heteroscedasticity
non-Studentized bootstrap	Inaccurate with low coverage rates; might be affected by small cost parameter
Bootstrap-*t*	Inaccurate with low coverage rates; might be affected by small cost parameter
Normal-theory bootstrap	Produces symmetric CIs by construction, thus cannot account for asymmetry of WTP distribution; gives low coverage rates, and might be affected by small cost parameter
Bootstrap percentile	Accurate, generally performed well but might be affected by small cost parameter
Bias-corrected bootstrap percentile	Accurate, generally performed well but might be affected by small cost parameter; better than general bootstrap percentile
Bias-corrected, accelerated bootstrap percentile	Accurate, generally performed well but might be affected by small cost parameter; better than general bootstrap percentile
Test-inversion bootstrap	Not affected by small cost parameter but sometimes does not give satisfactory results
Studentized test-inversion bootstrap	Not affected by small cost parameter but sometimes does not give satisfactory results

(including cost). The utilities were generated based on the MNL model. Their simulations analyzed the different cases of correct model specification, incorrect model specification and weak identification. The methods were compared based on the coverage, length, and shape of the CIs produced. In addition, they also applied these methods to a dataset to analyze what CIs were built. Table 4.4 gives a brief summary of their conclusions.

Gatta et al. (2013) conclude that the Fieller method is the best method to use in general, while the LR method is a good alternative if it is known that the model is being specified correctly. In this, they support the findings of Armstrong et al. (2001) and Bolduc et al. (2010). Note however that the conclusions concerning weak identification (i.e. the effect of a small cost parameter) is extrapolated from their actual findings and that of Bolduc et al. (2010). The authors were unable to observe an effect of weak identification, concluding that this was due to their not actually using small enough cost parameter values.

While much has already been analyzed in the literature, and especially by Gatta et al. (2013), there are still areas for further investigation which this work aims to address. Firstly, I maintain that Bayesian methods are also an equally viable and useful method for building CIs, and propose that they be included as a mainstream method for interval estimation and construction. Secondly, I look deeper into the effect of weak identification, specifically looking at which methods perform well, and how the CIs built are affected. Thirdly, I study how these CI intervals can be built under assumptions of unobserved heterogeneity, i.e. when random parameter models are assumed. Finally, I investigate the building of CIs for functions of parameter estimates other than ratio measures, i.e. looking beyond WTP measures to other functions of parameter estimates, such as choice probabilities.

4.1 Bounded and Unbounded Confidence Intervals

One of the biggest contentions with the Fieller method is the possibility of constructing unbounded CIs. Certain authors see this as a reason for not using the Fieller method (Hole, 2007). Others say that any bounded CI can have zero coverage probability, and thus a method for building CIs must be able to produce unbounded CIs by necessity (Bernard et al., 2007). Even if unbounded CIs are accepted, there still remains the practical problem of interpreting these unbounded CIs.

While bounded CIs might be more useful in practical applications, the fact that unbounded CIs exist is merely a consequence of studying a ratio function. Several researchers (Gleser and Hwang, 1987; Koschat, 1987; Hwang, 1995) have shown that any method which cannot produce unbounded CIs for a ratio might instead produce arbitrarily large deviations from the intended confidence level. In fact, among the methods studied in this paper, only the Fieller method is able to produce unbounded CIs and will not result in these large deviations. (While the Bayesian or KR method can potentially also lead to unbounded CIs, the probability that a sample will have a cost parameter estimate of exactly 0 is 0. However, in such situations, these methods might potentially produce seemingly "unreasonably" large CIs.)

As discussed in section 3.1.1, the Fieller method will only produce unbounded CIs if the denominator of the ratio is not significantly different from 0. For a discrete choice model, this implies that the parameter of cost should not be significantly different from 0, i.e. a consumer is insensitive to changes in cost. Mathematically, this is sensible; if the cost parameter is not significantly different from zero, then the value of WTP can be arbitrarily large, resulting in an unbounded CI. To understand this empirically, consider separately the cases of having a CI that is bounded on one side (case 2 in section 3.1.1) and having a CI that is the entire real line (case 3 in section 3.1.1). Following the argument of Siani and de Peretti (2003), the former case implies that the WTP is not statistically distinguishable from infinity, and the Fieller method is useful in detecting this. For the latter case, obtaining the entire real line implies that the WTP ratio is poorly defined; this results either from the fact that having an improvement to the attribute in question is equivalent to not having said improvement, or that the sample size is not large enough to distinguish between these two situations (i.e., we can learn nothing from our data).

Even with these interpretations, there is difficulty in understanding how to report the CI and its coverage. In particular, if the CI is the entire real line, then it will definitely contain the true parameter ratio, while not contributing much qualitative meaning as a CI. There are several methods that have been adopted to solve this problem. The first is to ignore the unbounded CIs, and only report those which are bounded; doing this effectively gives us a conditional confidence level (Buonaccorsi and Iyer, 1984). However, this would result in a low coverage, as demonstrated in the results following. Another solution, as proposed by Tsao and Hwang (1998), is to estimate the confidence as 1 in the unbounded case, and as $1 - \alpha$ in the bounded case. In this paper, the percentage of "real line" CIs produced by the Fieller method is reported separately from the coverage rate of the remaining CIs produced. In addition, examples are also given to demonstrate the coverage rates that would occur if using a conditional confidence level.

CHAPTER 5

BUILDING CONFIDENCE INTERVALS FOR WILLINGNESS-TO-PAY

In this chapter I describe the simulation and case studies conducted to compare and contrast the different methods of building CIs for WTP (ratio) measures. In each simulation and case study, I build CIs using each method described in Chapter 3. In addition, the performance of these methods are compared under different model assumptions (fixed parameter versus random parameter models) and in the presence of weak identification.

5.1 Simulation Studies

This simulation extends the work of Bolduc et al. (2010), who consider a simple choice situation in which an individual n chooses from two alternatives $i \in \{1, 2\}$, with each alternative containing three attributes. The attributes consist of one constant value and two independently drawn from the standard normal distribution. Parameter values, $\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 3.3 \\ \beta_3 \end{pmatrix}$, where $\beta_3 \in \{2,\ 1,\ 0.5,\ 0.4,\ 0.3,\ 0.2,\ 0.1,\ 0.01,\ 0.001,\ 0.0001\}$. The choice model is as follows:

$$\begin{aligned} \mathbf{U}_{in} &= \mathbf{X}_{in}\boldsymbol{\beta} + \varepsilon_{in} \\ \varepsilon_n &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xi_n \\ \xi_n &\overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \end{aligned} \tag{5.1}$$

which gives the simple binary probit model. Again, choice is made by maximizing utility, such that

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn} \text{ for } j = 1, 2 \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

Bolduc et al. (2010) use their simulation to compare the Delta, Fieller and Bootstrap methods over four different sample sizes. In this study, I compare all the methods listed in chapter 3 (other than the method of inverting the Bayes factor), and also study the effect of one more sample size, choosing individual sample sizes from $N \in \{100, 250, 1000, 5000, 100000\}$. This provides a total of 50 different simulation cases. For each case, CIs for the ratio β_2/β_3 is constructed for $M = 1000$ trials. Only 1000 trials were conducted after comparing the results obtained using 1000 trials and 10000 trials (as conducted by Bolduc et al. (2010)), and finding very minor differences between them.

Supposing that for each trial $m \in 1, \ldots, 1000$ we build CI $(\widehat{WTP}_L^{(m)}, \widehat{WTP}_U^{(m)})$. Then in order to access the performance of the different methods, the following statistics are evaluated:

Coverage: This calculates the fraction of CIs that contain the true WTP value. Since we begin with a known WTP value (denoted $WTP = \beta_2/\beta_3$), coverage can be computed as

$$\text{Coverage } = \frac{1}{M} \sum_{m=1}^{M} I(\widehat{WTP}_L^{(m)} \leq WTP \leq \widehat{WTP}_U^{(m)}) \tag{5.3}$$

where $I(\cdot)$ is the indicator function. The ideal coverage depends on the confidence level chosen for the simulation. In this study, the confidence level is set at $\alpha = 0.05$, and thus an accurate method would give a coverage of 0.95. Having a coverage either too low or too high indicates some inaccuracy in the means by which the CI is being constructed; in particular, while one might assume a higher coverage to be better, this might indicate that the CIs being constructed are too wide and thus every CI built contains the true WTP value.

Average Width: In general, the narrower the width of the CI, the more accurate it is said to be. However, CIs might be built so narrow as to exclude the true WTP value, thus losing out in coverage. This is the complaint that a number of previous studies have with the Delta Method. Average width is computed as follows:

$$\text{Average Width } = \frac{1}{M}\sum_{m=1}^{M}(\widehat{WTP}_U^{(m)} - \widehat{WTP}_L^{(m)}) \tag{5.4}$$

Shape: This gives an insight into the asymmetry of the CI built. Shape can be computed as follows:

$$\text{Shape } = \frac{1}{M}\sum_{m=1}^{M}\frac{\widehat{WTP}_U^{(m)} - \widehat{WTP}^{(m)}}{\widehat{WTP}^{(m)} - \widehat{WTP}_L^{(m)}} \tag{5.5}$$

The WTP value used here is that estimated by each method. For example, the estimated WTP value used by the Delta and Fieller methods is the ratio of the estimated parameters by the frequentist model. The shape index is the ratio of the difference between the estimated WTP value and the upper and lower bounds of the CI respectively. If the ratio is exactly equal to 1, then the CI built is symmetric. Having a shape index different from 1 indicates asymmetry in the CI built.

5.1.1 Fixed Coefficients Model

For the first simulation study, a fixed coefficient model was assumed. Using the same parameter values as those studied by Bolduc et al. (2010), a binary probit model was estimated, and CIs were built for the WTP of the second attribute (with the third attribute acting as price). The range of price values studied allows us to also analyze the effect of weak identification on the CIs built.

Table 5.1 shows the coverage of each method of building CIs. In this table, the total coverage for the Fieller method is reported, i.e. the total coverage for both bounded and unbounded intervals (excluding those intervals which are the $\Re$ line).

Table 5.1: Coverage of confidence intervals for fixed coefficient model

Sample Size	β_3	2	1	0.5	0.4	0.3	0.2	0.1	0.01	0.001	0.0001
100	Delta	0.926	0.913	0.894	0.868	0.832	0.773	0.620	0.225	0.070	0.025
	Fieller	0.960	0.959	0.960	0.961	0.954	0.957	0.954	0.957	0.956	0.955
	KR	0.957	0.954	0.972	0.982	0.970	0.963	0.929	0.035	0.000	0.000
	FQ WTP-Sp	0.934	0.917	0.784	0.710	0.586	0.458	0.298	0.041	0.002	0.000
	BayesPP	0.942	0.947	0.955	0.943	0.937	0.935	0.916	0.025	0.000	0.000
	BA WTP-Sp	0.810	0.897	0.934	0.932	0.925	0.848	0.683	0.039	0.000	0.000
250	Delta	0.940	0.940	0.917	0.904	0.886	0.855	0.740	0.284	0.089	0.030
	Fieller	0.949	0.950	0.953	0.952	0.953	0.949	0.951	0.950	0.950	0.953
	KR	0.949	0.952	0.971	0.973	0.979	0.971	0.963	0.291	0.000	0.000
	FQ WTP-Sp	0.948	0.944	0.878	0.821	0.710	0.546	0.314	0.027	0.000	0.000
	BayesPP	0.946	0.953	0.956	0.956	0.952	0.949	0.938	0.285	0.000	0.000
	BA WTP-Sp	0.801	0.910	0.900	0.861	0.819	0.623	0.205	0.000	0.000	0.000
1000	Delta	0.945	0.948	0.946	0.937	0.925	0.905	0.861	0.431	0.129	0.040
	Fieller	0.945	0.948	0.952	0.948	0.949	0.949	0.953	0.951	0.953	0.953
	KR	0.950	0.952	0.945	0.952	0.958	0.975	0.970	0.787	0.000	0.000
	FQ WTP-Sp	0.949	0.954	0.939	0.935	0.911	0.787	0.421	0.010	0.000	0.000
	BayesPP	0.947	0.952	0.957	0.957	0.961	0.956	0.955	0.785	0.000	0.000
	BA WTP-Sp	0.810	0.835	0.609	0.427	0.214	0.030	0.000	0.000	0.000	0.000
5000	Delta	0.949	0.952	0.948	0.946	0.946	0.942	0.577	0.577	0.198	0.078
	Fieller	0.949	0.953	0.948	0.949	0.947	0.954	0.948	0.951	0.951	0.950
	KR	0.951	0.953	0.948	0.954	0.948	0.954	0.973	0.915	0.000	0.000
	FQ WTP-Sp	0.949	0.950	0.954	0.958	0.946	0.884	0.488	0.006	0.000	0.000
	BayesPP	0.948	0.953	0.948	0.948	0.950	0.960	0.926	0.926	0.000	0.000
	BA WTP-Sp	0.858	0.851	0.443	0.210	0.016	0.000	0.000	0.000	0.000	0.000
10000	Delta	0.949	0.950	0.950	0.950	0.950	0.949	0.930	0.653	0.226	0.080
	Fieller	0.949	0.948	0.950	0.952	0.951	0.950	0.949	0.949	0.955	0.953
	KR	0.950	0.952	0.964	0.945	0.945	0.945	0.965	0.934	0.017	0.000
	FQ WTP-Sp	0.941	0.944	0.953	0.964	0.941	0.900	0.498	0.009	0.000	0.000
	BayesPP	0.949	0.947	0.949	0.948	0.950	0.954	0.958	0.934	0.001	0.000
	BA WTP-Sp	0.874	0.832	0.342	0.128	0.010	0.000	0.000	0.000	0.000	0.000

The patterns in the coverage values of the Delta and Fieller methods follow those reported by Bolduc et al. (2010). In general, increasing the sample size causes an improvement in the coverage values of the CIs built, while weak identification causes a decrease in the coverage of the CIs built by each method, i.e. as β_3 decreases, coverage rates also decrease.

There is a relationship between the effect of the sample size and the value of β_3, together, on the coverage value. Increasing the sample size can override the effect of weak identification on the coverage rates. For example, with a sample size of 100 individuals, the coverage rate of the Delta method is less than 0.90 for any $\beta_3 \leq 0.4$. However, with a sample size of 10000 individuals, the coverage rate of the Delta method is around the expected 0.95 even until $\beta_3 = 0.1$.

The poor coverage rates of the Delta method can be explained by the poor estimation of the WTP value itself. Since the Delta method produces symmetric intervals by construction, its accuracy is highly dependent on the accuracy of the parameter estimates. Table 5.2 gives the mean estimates of the WTP value of the second attribute as estimated by the frequentist method under preference space assumptions and WTP-space assumptions. The former values are the average midpoints of the CIs built by the Delta method. As can be seen, these estimates tend to become inaccurate as β_3 decreases, hence directly contributing to the poor coverage rates. It is interesting to note that this is in spite of the huge average widths of the CIs (see Table 5.7), which further decreases the usability of these CIs.

The coverage rates of the CIs built when the model is estimated in WTP-space are on par with or worse than that of the Delta method. Again, one can observe the same trend when comparing the estimated WTP values. Recall that in WTP-space, the estimates of the model are the WTP values directly.

Table 5.2: Mean estimates of WTP under parameter and WTP-space

Sample Size	β_3	2	1	0.5	0.4	0.3	0.2	0.1	0.01	0.001	0.0001
	True WTP	1.65	3.3	6.6	8.25	11	16.5	33	330	3300	33000
100	Frequentist	1.68	3.62	9.6	10.4	9.6	12.6	-9.0	-38.8	9.9	16.3
	FQ WTP-Sp	1.67	-4.65	-54.2	-78.7	-87.0	-66.7	-33.0	2.7	9.8	4.98
	BA WTP-Sp	1.86	5.36	15.2	18.1	23.0	27.6	35.5	38.4	40.0	40.4
250	Frequentist	1.66	3.37	7.5	10.1	10.2	22.9	13.1	120.7	-133.3	-6.2
	FQ WTP-Sp	1.66	3.37	-37.7	-54.8	72.8	-74.6	-50.3	-24.9	4.77	2.42
	BA WTP-Sp	1.70	3.56	7.33	8.88	10.6	12.3	14.6	17.2	17.1	16.9
1000	Frequentist	1.65	3.32	6.7	8.6	11.3	22.8	46.3	243.7	0.32	-46.9
	FQ WTP-Sp	1.65	3.31	6.7	5.4	-7.6	-44.4	-71.9	-24.2	-4.38	-0.65
	BA WTP-Sp	1.63	3.15	5.58	6.41	7.40	8.58	10.0	11.4	11.6	11.6
5000	Frequentist	1.65	3.30	6.6	8.3	11.1	16.9	38.4	127.8	67.7	106.3
	FQ WTP-Sp	1.65	3.30	6.6	8.3	11.2	15.1	-0.42	-8.50	2.51	-1.75
	BA WTP-Sp	1.63	3.16	5.62	6.50	7.52	8.78	10.2	11.6	11.8	11.8
10000	Frequentist	1.65	3.30	6.6	8.3	11.1	16.7	34.9	-164.2	-168.0	-160.4
	FQ WTP-Sp	1.65	3.31	6.6	8.3	11.1	16.7	23.7	16.7	-0.22	-5.43
	BA WTP-Sp	1.64	3.17	5.65	6.53	7.61	8.86	10.3	11.8	12.0	12.0

In contrast to the conclusions of Sonnier et al. (2007), the WTP estimates from the model in WTP-space do not appear any more accurate than those estimates from the model in preference space; in fact, estimating the model with Bayesian methods produces far less accurate estimates of the WTP values. Comparing first the estimates obtained from the frequentist method of estimation under parameter and WTP-space, there are several possible reasons for the disparity. Firstly, Sonnier et al. (2007) perform their simulation with 4500 total choice situations, for which both models give accurate WTP value estimates up to $\beta_3 = 0.2$. Secondly, they are not comparing CIs, but the root mean-squared errors (RMSE) and mean absolute errors (MAE) between the true and estimated WTP values. Thirdly, their parameters are not fixed. In particular, the parameter of price used is not consistently large or small, but ranges so as to allow for some weak identification. Even in this study, the average RMSE and MAE values for 5000 individuals across all β_3 values holds to the conclusions of Sonnier et al. (2007) (see Table 5.3). However, the estimated WTP values are not reported by Sonnier et al. (2007), making it impossible to compare between the two.

Table 5.3: Root mean-squared error and mean absolute error for 5000 individuals

β_3	RMSE		MAE	
	Parameter-Sp	WTP-Sp	Parameter-Sp	WTP-Sp
2	0.001	0.001	0.026	0.027
1	0.012	0.012	0.088	0.091
0.5	0.019	0.019	0.34	0.34
0.4	0.46	0.45	0.53	0.53
0.3	1.53	1.53	0.96	0.95
0.2	8.13	3452	2.16	3.93
0.1	2873	49690	12.9	46.8
0.01	8.5E7	1.7E5	904	345
0.001	2.4E8	1.1E7	3736	3296
0.0001	1.3E9	1.1E9	33278	33001
Average	1.7E8	1.1E8	3794	3670

It is clear that having lower RMSE and MAE values does not translate to building CIs that give good coverage. Note, however, that the model in WTP-space is estimated using MLE, thus assuming an asymptotic normal distribution of the WTP values. This assumption likely contributes directly to both the inaccurate WTP estimates (especially under weak identification) and the poor coverage rates. In fact the WTP values should logically follow a distribution that is truncated at zero; for example, it is normally assumed that individuals will have a positive WTP to lower travel time. Hence, the WTP values, being incorrectly specified, would thus be estimated wrongly. In addition, as with most other methods, the WTP-space method is simply unable to handle the case of weak identification.

This could also be the reason for the poor estimation by the Bayesian method. The Bayesian estimation of the model under WTP-space gives CIs with extremely poor coverage, and also very inaccurate estimates of the WTP values. By looking at the WTP estimates as the sample size increases, it appears that the Bayesian method is simply converging to the *wrong value*[1]. That, in combination with the narrow widths of the CIs produced (see Table 5.7) translates to extremely poor coverage.

The Krinsky-Robb and Bayesian post-processing methods display consistently good coverage rates for most of the β_3 values. At a certain β_3 value, however, these coverage rates drop dramatically, almost to 0. This occurs at $\beta_3 = 0.01$ for $n \in 100, 250$, and at $\beta_3 = 0.001$ for $n \in 1000, 5000, 10000$. (For $n = 1000$, the coverage rate is around 0.79 when $\beta_3 = 0.001$, which is substantially less than 0.95, but not as dramatic a drop as to 0.) A closer look at the parameter estimates gives some insight as to why this is happening (Table 5.4). As can be seen, many of the estimates for these small β_3 values are inaccurate, resulting in inaccurate WTP estimates and thus CIs.

Table 5.4: Mean Bayesian and KR estimates of β_3 and WTP

Sample Size	True Value	β_3			WTP		
		0.01	0.001	0.0001	330	3300	33000
100	Bayesian	0.0098	-0.0031	0.0015	262	-854	1772
	KR	0.0095	-0.0030	0.0096	281	908	304
250	Bayesian	0.0074	0.0016	-0.0014	320	1502	-1648
	KR	0.0073	0.0023	-0.00043	340	-1073	-5739
1000	Bayesian	0.0066	-0.0020	0.0022	347	-1154	1036
	KR	0.0073	-0.0018	0.0018	326	-1344	1291
5000	Bayesian	0.0076	-0.0012	-0.000017	301	-1972	-138431
	KR	0.0078	-0.0012	-0.000062	299	-1892	-37983
10000	Bayesian	0.0071	0.0011	-0.00041	325	2110	-5652
	KR	0.0071	0.0012	-0.00040	328	1931	-5797

The Fieller method gives excellent coverage rates of around 0.95 no matter what sample size or β_3 value is being studied. This supports the results of

[1]This convergence to an incorrect value occurred when the Bayesian method was run with a diffuse prior. Working with a tight prior around the true WTP values produced fairly accurate WTP estimates (especially under weak identification). However, coverage rates were still poor, and almost erratic.

Bolduc et al. (2010) who find the Fieller method to perform well. These coverage rates hold even when excluding the cases where the Fieller method produces the entire $\Re$ line as the CI. Observe, however, that the coverage rates displayed by only the bounded intervals mirror those of the Delta method. Table 5.5 compares the difference in coverage between considering all Fieller method CIs, and only those which are bounded (that is, reporting the conditional confidence level). It also includes the fraction of CIs which were the $\Re$ line. As can be seen, it is the unbounded intervals that are primarily contributing to the superior coverage of the Fieller method when β_3 is small. This supports the work of researchers like Gleser and Hwang (1987), Koschat (1987) and Hwang (1995), as the bounded CIs produced hardly contained the true WTP values and thus deviated greatly from the intended confidence level, while the unbounded CIs contained the true WTP values. As discussed in section 4.1, producing unbounded CIs is a natural consequence of working with ratios, and the use of a conditional confidence level (as accepting only bounded CIs is) will result in low coverage rates. Observe also how the value of β_3 at which unbounded CIs are built decreases as sample size increases. This illustrates how having a large enough sample size affords the ability to distinguish between the situations of having an improvement to an attribute (contrary to not), as discussed in section 4.1.

To gain a better understanding of the coverage rates, I analyze the shape index of the CIs. The Monte Carlo frequentist and Bayesian values reported in Table 5.6 give the average shape index of CIs built using the WTP estimates derived directly from the model estimates (note that these shape indexes are calculated with respect to the true WTP value). These serve as a point of reference for the accuracy of the other methods. Note that the shape indexes of the CIs built by the Delta method and the frequentist WTP Space method are not

Table 5.5: Comparing coverages of all Fieller method confidence intervals and bounded Fieller method confidence intervals. Note that only sample size 100 produced $\Re$ line intervals.

Sample Size	β_3	2	1	0.5	0.4	0.3	0.2	0.1	0.01	0.001	0.0001
100	All	0.960	0.960	0.954	0.960	0.954	0.957	0.954	0.957	0.956	0.955
	Bounded	0.960	0.967	0.956	0.947	0.893	0.818	0.527	0.024	0.002	0.000
	$\Re$ line	0.008	0.003	0.003	0.003	0.003	0.002	0.003	0.002	0.002	0.002
250	All	NA		0.953	0.952	0.953	0.949	0.951	0.950	0.950	0.953
	Bounded	0.949	0.950	0.974	0.970	0.958	0.912	0.740	0.094	0.010	0.000
1000	All	NA			0.948	0.949	0.949	0.953	0.951	0.953	0.953
	Bounded	0.945	0.948	0.952	0.948	0.961	0.968	0.918	0.186	0.015	0.000
5000	All	NA						0.948	0.951	0.951	0.950
	Bounded	0.949	0.953	0.948	0.949	0.947	0.954	0.971	0.391	0.045	0.000
10000	All	NA						0.949	0.949	0.955	0.953
	Bounded	0.949	0.948	0.950	0.952	0.951	0.950	0.953	0.565	0.085	0.000

reported as these, by construction, produce symmetric CIs, and thus the shape index will always be 1.

I first analyze the shape indexes for $\beta_3 \geq 0.1$. In general, the Monte Carlo intervals of both frequentist and Bayesian model estimates give shape indexes of slightly larger than 1. This shows that the CIs are asymmetric, with the upper bound being slightly further away from the true WTP value than the lower bound. This asymmetry cannot be captured by the Delta and frequentist WTP Space methods, which are symmetric by construction. The Bayesian and Krinsky-Robb methods build CIs that mostly reflect the same asymmetry of the Monte Carlo intervals, especially as sample size increases. The Fieller method tends to build bounded CIs which are very asymmetric, with the upper bound being much further away from the mean WTP value than the lower bound. However, these large shape indexes coincide directly with the existence of unbounded CIs. In fact, for any combination of sample size and β_3 that produces bounded Fieller CIs having a shape index greater than 2, that combination would also build unbounded Fieller CIs. This reiterates the point that one can-

Table 5.6: Shape index of confidence intervals. Note that the shape index reported for the Fieller method are calculated only from bounded intervals.

Sample Size	β_3	2	1	0.5	0.4	0.3	0.2	0.1	0.01	0.001	0.0001
100	MC Frequentist	1.27	1.77	5.05	1.22	1.03	0.83	0.68	-0.19	-0.87	-0.99
	MC Bayesian	1.34	2.90	1.04	1.02	0.78	0.80	0.73	-0.34	-0.90	-0.99
	Fieller	1.66	3.95	20.2	46.6	49.1	56.9	34.4	21.1	26.2	17.5
	KR	1.10	2.09	19.4	1.84	1.56	-8.49	3.07	1.11	0.72	0.64
	BayesPP	1.11	2.14	2.08	0.44	2.79	1.62	0.099	2.36	1.05	1.35
	BA WTP-Sp	1.16	1.30	1.31	1.41	1.41	1.46	1.47	1.48	1.52	1.50
250	MC Frequentist	1.09	1.35	2.18	2.94	5.83	0.83	0.75	0.054	-0.81	-0.98
	MC Bayesian	1.25	1.49	3.14	1.43	0.87	0.83	0.68	-0.13	-0.84	-0.98
	Fieller	1.31	1.82	9.10	24.4	30.5	60.0	41.5	32.7	44.8	29.1
	KR	1.22	1.57	2.20	3.19	3.02	0.94	1.27	1.22	1.01	0.93
	BayesPP	1.07	1.18	1.54	2.26	1.87	2.39	0.98	1.01	1.27	2.47
	BA WTP-Sp	1.15	1.12	0.97	0.98	0.98	1.04	1.04	1.07	1.07	1.07
1000	MC Frequentist	1.07	1.17	1.41	1.39	1.80	2.70	1.20	0.31	-0.65	-0.96
	MC Bayesian	1.05	1.26	1.59	1.89	2.90	1.14	0.99	0.14	-0.74	-0.97
	Fieller	1.14	1.31	1.79	2.18	3.63	13.8	37.3	39.3	24.0	17.5
	KR	1.10	1.23	1.53	1.72	2.10	-13.8	1.80	0.51	1.33	0.87
	BayesPP	1.04	1.08	1.18	1.24	1.40	1.66	1.30	0.88	0.95	2.11
	BA WTP-Sp	1.16	0.92	0.66	0.64	0.66	0.70	0.74	0.76	0.76	0.77
5000	MC Frequentist	1.06	1.06	1.11	1.15	1.24	1.40	2.36	0.68	-0.32	-0.87
	MC Bayesian	1.02	1.13	1.16	1.27	1.24	1.76	2.04	0.60	-0.45	-0.88
	Fieller	1.06	1.13	1.28	1.37	1.53	1.94	11.5	95.0	34.3	34.5
	KR	1.04	1.09	1.20	1.26	1.35	1.62	2.07	1.07	1.20	1.18
	BayesPP	1.02	1.04	1.07	1.08	1.11	1.21	1.66	0.84	-0.93	0.15
	BA WTP-Sp	1.38	0.53	0.45	0.49	0.54	0.59	0.66	0.71	0.71	0.70
10000	MC Frequentist	1.06	1.03	1.03	1.14	1.11	1.20	1.59	0.70	-0.19	-0.94
	MC Bayesian	1.06	1.12	1.08	1.18	1.18	1.40	2.32	0.68	-0.43	-0.88
	Fieller	1.04	1.09	1.19	1.24	1.34	1.57	3.20	52.9	53.5	52.6
	KR	1.03	1.07	1.14	1.18	1.25	1.39	2.10	1.76	2.14	1.62
	BayesPP	1.01	1.04	1.05	1.06	1.08	1.14	1.40	2.66	0.83	0.98
	BA WTP-Sp	1.48	0.43	0.41	0.45	0.51	0.58	0.64	0.71	0.71	0.70

not simply use a conditional confidence level and ignore the unbounded CIs produced by the Fieller method.

Consider next the shape indexes for $\beta_3 \leq 0.01$. Many of the Monte Carlo intervals have shape indexes of less than 0, especially for the smaller sample sizes. A shape index of less than 0 indicates that the CI does not even contain the true WTP value. This is in fact the case, and can also be conjectured from

Tables 5.2 and 5.4, where the estimated WTP values are far from accurate. Since these estimates are inaccurate, one cannot expect the shape index to be a good basis for comparison. Indeed it is not surprising that under weak identification the estimates are inaccurate, as a cost parameter that is not significantly different from zero would be difficult to estimate accurately. In addition, as mentioned in section 4.1, this would produce WTP values which are not significantly different from infinity, and thus an unbounded CI is necessary.

Aside from several small sample size cases, the CIs built by the Bayesian method in WTP-space have shape indexes opposite from those built by any other method. That is, they produce CIs which are asymmetric, but the lower bound is further away from the mean WTP value than the upper bound. This probably explains why, unlike the other methods, the Bayesian method in WTP-space builds CIs with the best general coverage when the sample consists of only 100 individuals. Since the CIs are not capturing the asymmetry of the Monte Carlo intervals, it is not surprising that their coverage rates are poor.

Finally, I analyze the average widths of the CIs built (Table 5.7). These tell a similar story to the coverage values. While all methods produce CIs of increasing width as β_3 decreases, the magnitude of those built by the Delta method is much bigger than those of any other method. It appears that the claim of the Delta method producing narrower CIs only holds when weak identification is not an issue. These large widths occur at the same β_3 values at which poor coverage occurs, signifying that the inability of the model to also accurately estimate the variance-covariance matrix of the parameters under weak identification. The Fieller, Krinsky-Robb and Bayesian post-processing methods all produce intervals of similar average widths, although the slightly narrower widths

Table 5.7: Average width of confidence intervals. Note that the average widths reported for the Fieller method contain only bounded intervals.

Sample Size	β_3	2	1	0.5	0.4	0.3	0.2	0.1	0.01	0.001	0.0001
100	Delta	1.13	20.1	691.2	3.5E4	1.3E5	1.4E6	9.2E8	2.1E6	3.1E5	7.2E5
	Fieller	1.24	12.3	66.6	167.8	167.9	224.8	150.0	166.5	190.9	161.7
	KR	1.15	7.71	93.7	136.3	201.8	261.3	306.4	321.0	328.1	322.0
	FQ WTP-Sp	1.21	6.47	68.9	97.0	99.4	104.3	171.2	145.3	162.7	138.7
	BayesPP	1.13	7.80	86.2	151.2	195.6	264.2	312.6	327.6	319.7	322.1
	BA WTP-Sp	1.06	7.71	32.6	40.9	55.5	71.8	99.0	99.6	105.1	108.0
250	Delta	0.58	2.05	25.5	283.6	1.3E4	6.6E4	3.4E5	1.7E7	2.1E7	2.7E6
	Fieller	0.61	2.37	41.7	123.5	168.6	326.9	255.7	571.8	418.8	447.6
	KR	0.61	2.41	29.0	73.6	172.3	314.7	467.2	514.9	539.3	530.6
	FQ WTP-Sp	0.58	2.06	23.3	33.5	51.2	71.9	84.3	112.7	89.6	91.3
	BayesPP	0.60	2.20	25.85	68.6	161.9	302.7	447.3	511.4	509.9	512.2
	BA WTP-Sp	0.43	1.96	7.80	10.7	14.1	17.6	22.4	27.9	27.6	27.3
1000	Delta	0.29	0.98	3.93	6.42	144.5	5826.1	2.6E5	2.6E7	8.4E5	2.4E6
	Fieller	0.29	1.01	4.36	7.85	23.6	144.6	421.6	665.2	1043.0	603.5
	KR	0.29	1.00	4.47	7.66	24.6	142.7	658.7	1048.0	1075.2	1060.6
	FQ WTP-Sp	0.29	0.98	3.91	6.65	14.6	25.7	46.4	46.3	49.1	46.5
	BayesPP	0.29	1.00	4.17	7.19	20.4	136.4	610.7	1023.3	1022.7	1020.1
	BA WTP-Sp	0.21	0.96	3.42	4.52	5.84	7.58	9.85	12.2	12.4	12.4
5000	Delta	0.13	0.43	1.67	2.61	4.69	10.9	160.7	2.1E6	8.5E6	4.1E6
	Fieller	0.13	0.44	1.70	2.69	4.93	12.5	261.8	3140.9	1509.0	1343.6
	KR	0.13	0.43	1.69	2.68	4.91	12.8	207.2	2263.2	2306.7	2381.0
	FQ WTP-Sp	0.13	0.43	1.67	2.65	5.01	13.0	18.7	19.2	18.8	22.6
	BayesPP	0.13	0.44	1.69	2.65	4.81	11.7	192.8	2249.2	2295.8	2311.9
	BA WTP-Sp	0.10	0.74	3.01	4.06	5.45	7.29	9.54	12.0	12.2	12.2
10000	Delta	0.09	0.31	1.17	1.83	3.26	7.47	34.5	1.05E7	2.41E5	1.9E6
	Fieller	0.09	0.31	1.18	1.85	3.33	7.89	59.8	2063.0	2091.0	1.1E4
	KR	0.09	0.31	1.18	1.85	3.34	7.89	41.3	3196.8	3349.2	3248.8
	FQ WTP-Sp	0.09	0.31	1.18	1.84	3.40	8.63	12.5	11.8	12.0	12.0
	BayesPP	0.09	0.31	1.18	1.84	3.29	7.67	46.9	3085.3	3285.4	3238.7
	BA WTP-Sp	0.08	0.70	2.94	4.02	5.48	7.36	9.68	12.2	12.5	12.5

of the bounded intervals built by the Fieller method under weak identification might be what contributes to the poorer coverage rates of these bounded intervals. CIs built using the WTP-space method are similar in width to the other intervals for larger β_3 values, however, they are much smaller when there is weak identification. Since the CIs built by the WTP Space method is symmetric by construction, the narrow widths and inaccurate estimates of WTP contribute to the low coverage rates.

5.1.2 Random Coefficients Model

In this simulation study, a random coefficients model is assumed. As with the fixed coefficient model, each individual n chooses from 2 alternatives $i = \{1, 2\}$, with each alternative containing three attributes. As before, the attributes consist of one constant value and two independently drawn from the standard normal distribution. The difference with the random coefficients model lies with the coefficient values. $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ consists of a constant $\beta_1 = 1$ and random β_2 and β_3 values with the following distribution:

$$\begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 3.3 \\ \hat{\beta}_3 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \tag{5.6}$$

where $\hat{\beta}_3 \in \{2,\ 0.5,\ 0.3,\ 0.1,\ 0.01,\ 0.001\}$. These mean $\hat{\beta}_3$ values were chosen based on the results of the fixed coefficient model simulation study, where poor coverage rates seem to occur around the 0.1 and 0.01 β_3 values. As such I focus on the smaller mean values to study how the CIs built will perform under weak identification. The choice model follows a mixed logit model, where

$$\begin{aligned} \mathbf{U}_{in} &= \mathbf{X}_{in}\boldsymbol{\beta} + \varepsilon_{in} \\ \varepsilon_{in} &\overset{iid}{\sim} \text{EV}(0, 1) \end{aligned} \tag{5.7}$$

As before, choice is made by maximizing utility.

Estimating a random coefficients model assumes that individuals have taste variations. Mathematically, each individual n has his own set of preferences $\boldsymbol{\beta}_n$, which he applies when making his choice over all choice situations. In order to estimate these accurately and also to more accurately simulate real surveys conducted, each individual in this simulation answers 10 choice situations. Hence, the total sample size (i.e. the total number of choice situations) $N \in \{100, 250, 1000, 5000, 10000\}$ as before. However, the total number of individuals is $N_{\text{ind}} \in \{10, 25, 100, 500, 1000\}$. Thus the population used for the simulation is generated *by individual*, i.e. the utilities for every 10 choice situations are calculated using the same $\boldsymbol{\beta_n}$, and represent the 10 choice situations faced by that same individual.

Table 5.8: Coverage of confidence intervals for random coefficients model

Sample Size	β_3	2	0.5	0.3	0.1	0.01	0.001
100	Delta (Mean)	1.000	1.000	1.000	0.999	0.966	0.878
	Delta (Median)	0.987	0.805	0.661	0.432	0.107	0.032
	KR	1.000	0.996	0.966	0.828	0.107	0.000
	BayesPP	0.999	1.000	0.999	0.828	0.000	0.000
250	Delta (Mean)	1.000	1.000	1.000	1.000	0.995	0.932
	Delta (Median)	0.980	0.530	0.309	0.092	0.011	0.007
	KR	1.000	1.000	0.997	0.925	0.022	0.000
	BayesPP	1.000	1.000	1.000	0.948	0.000	0.000
1000	Delta (Mean)	1.000	1.000	1.000	1.000	1.000	0.905
	Delta (Median)	0.967	0.034	0.005	0.000	0.000	0.000
	KR	1.000	1.000	1.000	0.964	0.000	0.000
	BayesPP	1.000	1.000	1.000	0.999	0.000	0.000
5000	Delta (Mean)	1.000	1.000	1.000	1.000	0.973	0.703
	Delta (Median)	0.686	0.000	0.000	0.000	0.000	0.000
	KR	1.000	1.000	1.000	0.998	0.000	0.000
	BayesPP	1.000	1.000	1.000	1.000	0.000	0.000
10000	Delta (Mean)	1.000	1.000	1.000	1.000	1.000	0.708
	Delta (Median)	0.953	0.000	0.000	0.000	0.000	0.000
	KR	1.000	1.000	1.000	0.991	0.000	0.000
	BayesPP	1.000	1.000	1.000	1.000	0.000	0.000

Table 5.8 reports the coverage rates of the CIs built. As in the fixed parameter model, the coverage of the Krinsky-Robb and Bayesian post-processing methods drop drastically as β_3 decreases from 0.1 to 0.01. Prior to that, however, the coverage rates of these methods are extremely high, most closer to 100% rather than the accurate 95%. As the mean estimates of the WTP are not particularly accurate (Table 5.9), it appears that the widths of the CIs being built are too large, and as a result too many of the CIs contain the true WTP value.

Table 5.9: Mean estimates of WTP

Sample Size	β_3	2	0.5	0.3	0.1	0.01	0.001
	True WTP	1.65	6.6	11	33	330	3300
100	Frequentist (Mean)	1.75	3.78	1.01	6.26	-11.12	76.40
	Frequentist (Median)	1.63	1.84	0.94	0.021	0.14	0.27
	KR	1.48	-37.44	3.86	523.0	11.95	-19.13
	Bayesian	-17.51	8.71	-1.21	7.16	-9.24	-143.1
250	Frequentist (Mean)	1.71	5.76	7.06	-7.48	10.85	1.81
	Frequentist (Median)	1.66	2.73	1.25	0.80	-0.18	-0.18
	KR	1.99	-6.55	4.46	-1.13	12.17	2.79
	Bayesian	-2.85	4.43	-4.66	-1.00	5.39	-2.11
1000	Frequentist (Mean)	1.66	7.81	-97.8	-135.6	12.26	236.6
	Frequentist (Median)	1.62	2.17	1.87	0.76	-0.10	-0.20
	KR	1.69	9.28	3.88	6.62	-0.053	1.62
	Bayesian	2.27	-20.17	1.82	-12.79	10.55	-10.98
5000	Frequentist (Mean)	1.65	6.68	11.49	12.29	-45.50	-53.14
	Frequentist (Median)	1.55	2.28	1.97	1.24	0.059	-0.13
	KR	-1.61	2.28	1.97	1.24	0.059	-0.13
	Bayesian	-1.02	63.57	-4.06	2.20	-2.21	4.72
10000	Frequentist (Mean)	1.65	6.65	11.27	42.26	21.81	-362.51
	Frequentist (Median)	1.64	2.32	2.17	1.41	-0.28	-0.53
	KR	1.67	4.43	-7.09	3.89	3.15	-12.69
	Bayesian	1.94	5.21	1.77	-35.65	-4.41	4.90

The two Delta methods give very different results. The Delta (Mean) method, which uses the mean of the WTP estimates and the mean of the calculated WTP standard errors, builds CIs with high coverage rates but also extremely large widths (as reported in Table 5.10). In contrast, the Delta (Median) method builds CIs with low widths, as suggested by Bliemer and Rose (2012), but also with very low coverage rates. The low widths actually work against this method, because the median of the WTP estimates are not accurate, and thus many of the CIs built do not contain the true WTP value.

Finally, I analyze the shape index of the CIs built (Table 5.11). The Monte Carlo CIs built under these heterogeneity assumptions are not similar in shape; in fact, most of them are asymmetric in different directions. Seeing as how the frequentist method gives better WTP estimates, I compare the other shape in-

Table 5.10: Average width of confidence intervals for random coefficients model

Sample Size	β_3	2	0.5	0.3	0.1	0.01	0.001
100	Delta (Mean)	1.5E6	1.4E9	1.2E10	1.8E9	2.0E8	3.1E8
	Delta (Median)	32.3	4800	2922	9793	2929	1517
	KR	55.4	254.1	292.1	306.9	290.2	291.1
	BayesPP	18.0	88.5	95.4	101.7	102.3	105.2
250	Delta (Mean)	1.6E9	6.3E9	2.6E8	9.2E8	2.9E9	5.3E12
	Delta (Median)	6.04	1204	4929	236.1	162.5	1616
	KR	23.0	137.2	150.6	163.2	171.0	169.3
	BayesPP	16.7	87.6	97.1	104.7	106.6	107.0
1000	Delta (Mean)	6.9E6	3.5E8	1.2E8	1.4E9	4.3E8	2.3E8
	Delta (Median)	0.84	3.82	4.17	4.37	4.25	4.45
	KR	11.7	86.8	100.8	108.9	106.0	106.6
	BayesPP	12.7	93.7	100.6	108.4	110.3	110.9
5000	Delta (Mean)	2.0E6	1.5E7	6.0E6	1.9E7	3.1E8	4.7E7
	Delta (Median)	0.27	1.51	1.66	1.81	1.81	1.81
	KR	11.7	107.9	107.8	115.6	112.7	113.0
	BayesPP	10.8	92.5	101.7	107.0	107.2	107.8
10000	Delta (Mean)	1.5E6	2.1E9	5.0E7	2.7E7	2.1E9	9.7E6
	Delta (Median)	0.21	1.06	1.20	1.30	1.31	1.32
	KR	10.04	76.2	86.0	89.4	103.4	105.3
	BayesPP	8.94	88.2	105.9	105.8	107.3	107.5

dexes to these. As the Delta method builds symmetric CIs by definition, they are unable to capture the asymmetry of the distribution of the WTP estimates, as illustrated by the shape indexes of the frequentist Monte Carlo CI. The shape indexes of both the Krinsky-Robb and Bayesian post-processing methods also do not match that of the Monte Carlo CI, even in direction of asymmetry (i.e. whether the shape index is greater than or less than 1). While this does not seem to reduce coverage, I attribute that more to the large widths rather than an accurate portrayal of the distribution of the WTP.

Table 5.11: Shape index of confidence intervals for random coefficients model

Sample Size	β_3	2	0.5	0.3	0.1	0.01	0.001
100	MC Frequentist	1.22	0.68	1.00	0.62	-0.50	-0.93
	MC Bayesian	1.62	1.37	0.81	0.60	-0.54	-0.92
	KR	3.14	1.04	1.04	1.37	1.13	0.88
	BayesPP	1.51	0.88	0.33	1.01	1.49	0.82
250	MC Frequentist	1.56	2.01	1.23	0.80	-0.27	-0.90
	MC Bayesian	1.03	0.62	0.56	0.37	-0.65	-0.96
	KR	1.48	0.89	1.28	1.02	1.01	1.20
	BayesPP	2.18	0.27	0.89	1.40	0.22	1.14
1000	MC Frequentist	1.01	2.23	4.19	0.69	0.17	-0.81
	MC Bayesian	1.23	0.70	0.43	0.62	-0.70	-0.97
	KR	2.55	1.00	0.86	0.90	1.19	0.71
	BayesPP	2.31	0.92	0.68	1.41	1.07	1.48
5000	MC Frequentist	1.35	1.15	1.52	0.86	0.39	-1.89
	MC Bayesian	1.06	0.74	0.52	0.30	-0.72	-0.97
	KR	2.94	0.92	1.04	0.86	1.11	1.01
	BayesPP	2.81	1.89	0.91	1.31	1.14	-0.24
10000	MC Frequentist	0.88	1.21	1.19	4.42	0.40	-0.34
	MC Bayesian	1.73	0.57	0.51	0.19	-0.70	-0.97
	KR	11.3	1.05	0.90	1.07	1.02	1.06
	BayesPP	3.90	0.99	1.14	0.80	0.94	0.83

5.2 Quasi-simulation

In order to illustrate the use of these methods in building CIs for specific data sets, a quasi-simulation and two case studies are performed. The quasi-simulation was performed to take advantage of existing survey data, while still having a basis with which to compare the WTP estimates and CIs built. Attribute data is taken from an unlabeled vehicle choice survey of vehicle owners in California (Train and Hudson, 2000; Sándor and Train, 2004; Train and Sonnier, 2005; Hess et al., 2006). The original data set contains 500 individuals responding to up to 15 choice experiments regarding their choice among three unspecified vehicles. This was duplicated 25 times to attain a total population of 12,500 individuals. Assuming a linear MNL model with a subset of the attributes, namely purchase price, operating cost, range, and engine type (whether gasoline, electric or hybrid), pre-determined coefficient values are then used to generate the choice indicators (Table 5.12). These coefficient values are obtained by solving a MNL model using the full original data set.

Table 5.12: Parameter values used in quasi-simulation

Attribute	Mean(β)
Purchase Price (1000$)	-0.053
Operating Cost ($/mth)	-0.027
Range (mi)	0.442
Electric	-1.436
Hybrid	0.414

Figure 5.1 outlines the flow of the quasi-simulation, which constructs the CIs of the WTP of operating cost and range over 1,000 trials. The quasi-simulation begins by initializing the individual sample size (with a choice from 10, 40, 70 or 100 individuals) and setting $nTrial = 1$. Following this, a sample from the dataset is taken, and CIs are built according to each of the six defined methods. The Bayesian model is solved using 1,000 MCMC sample iterations, while the Krinsky-Robb method uses 1,000 resamples from the multivariate normal distribution. The simulation is repeated for each $nTrial \in \{1, \ldots, 1000\}$, upon which the simulation ends. As before, I analyze the CIs by calculating their coverage, average widths and shape indexes.

Table 5.13: Coverage and average width of WTP for operating cost

	Coverage				Average Width				Shape Index			
Sample Size	10	40	70	100	10	40	70	100	10	40	70	100
MC Frequentist	0.95				1.041	0.462	0.371	0.313	1.016	1.039	1.027	1.022
MC Bayesian	0.95				1.106	0.470	0.373	0.315	1.041	1.030	1.059	1.027
Delta	0.206	0.092	0.096	0.075	0.154	0.033	0.019	0.013	1.000			
Fieller	0.946	0.960	0.949	0.943	1.211	0.514	0.381	0.315	1.569	1.215	1.156	1.127
KR	0.945	0.960	0.944	0.946	1.203	0.511	0.379	0.314	1.382	1.155	1.117	1.091
FQ WTP-Sp	0.772	0.758	0.807	0.806	0.740	0.346	0.385	0.255	1.000			
BayesPP	0.953	0.963	0.954	0.942	1.109	0.503	0.376	0.311	1.103	1.048	1.031	1.028
BA WTP-Sp	0.780	0.877	0.873	0.881	0.410	0.235	0.178	0.152	1.016	1.039	1.027	1.022

Tables 5.13 and 5.14 give the coverage rates, average widths and shape in-

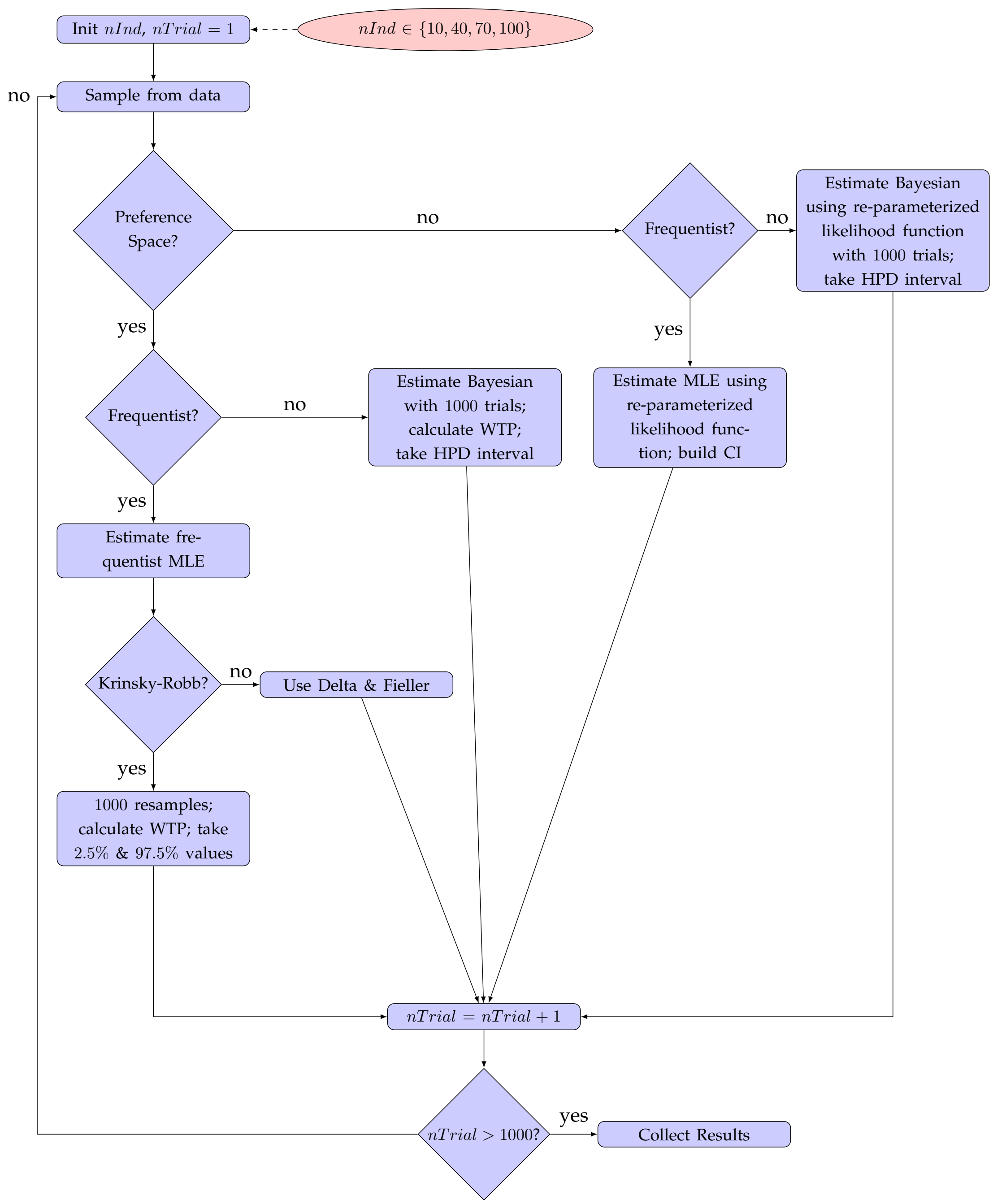


Figure 5.1: Path Flow for Quasi-simulation

Table 5.14: Coverage and average width of WTP for range

	Coverage				Average Width				Shape Index			
Sample Size	10	40	70	100	10	40	70	100	10	40	70	100
MC Frequentist	0.95				44.2	21.6	15.4	12.6	0.885	0.977	1.005	1.133
MC Bayesian	0.95				46.4	21.8	15.5	12.5	0.851	0.896	0.987	1.162
Delta	1.000	1.000	0.999	0.999	256.7	56.4	31.6	21.9	1.000			
Fieller	0.949	0.942	0.947	0.954	48.5	21.2	15.7	13.0	0.913	0.946	0.960	0.967
KR	0.947	0.940	0.946	0.950	48.3	21.1	15.7	13.0	0.931	0.959	0.972	0.975
FQ WTP-Sp	0.732	0.707	0.841	0.776	34.4	16.6	18.0	11.8	1.000			
BayesPP	0.950	0.945	0.949	0.954	46.0	20.8	15.5	12.9	0.982	0.959	0.972	0.975
BA WTP-Sp	0.455	0.433	0.416	0.414	0.593	0.454	0.369	0.325	1.041	1.030	1.059	1.027

dexes of the CIs built by each of the methods for the WTP of operating cost and range respectively. These are also compared to the results from the CIs built from the WTP estimates derived directly from the frequentist and Bayesian model estimates (i.e. the Monte Carlo CIs). Note that the coverage of the Monte Carlo CIs are 0.95 by construction. As can be seen, the Fieller, Krinsky-Robb and Bayesian post-processing methods all perform well, and give comparable results. They also match the results from the Monte Carlo CIs very closely. Recall that in the fixed coefficient model simulation, the coverage rates of the CIs built by the Krinsky-Robb method and the Bayesian method dropped dramatically when the magnitude of price was 0.01, and that a sample size of 100 gave Fieller CIs spanning the entire $\Re$ line. In the quasi-simulation, the smallest sample size is around 150 choice situations, while the magnitude of the price coefficient is around 0.05. This appears to be just the right combination to garner good coverage results from all three methods, without producing unbounded Fieller CIs.

The coverage rate of the CIs produced by the Delta method for the WTP of operating cost is extremely low. Unlike in the fixed coefficient model simulation conducted, the mean frequentist WTP estimates are extremely accurate. However, observe how the average width of the CIs produced by the Delta method are extremely small, much more so than that of the CIs built by the other methods. The narrowness of the CI built is reflective of the fact that the Delta Method can produce CIs with narrower widths than other methods; however, it is also contributing directly to the low coverage, as argued by Bernard et al. (2007). The problem is compounded by the fact that the Delta method produces symmetric CIs by construction. The shape index of the Monte Carlo CIs in Table 5.13 show that the distribution of the WTP is asymmetric, with shape indexes slightly more than 1. Hence, the symmetric CI in addition to the small widths results in the low coverage rates.

In contrast, the coverage rate of the CIs produced by the Delta method for the WTP of range is approximately 1 across all sample sizes. This is also not accurate, especially when the average widths are also taken into account. Here, the average widths of the CIs are much larger than those produced by the other methods; this results in the high coverage rate where every CI contains the true WTP value. Hence the CI being built is not specific enough to be of any use to the practitioner.

Finally, the CIs built by re-parameterizing the utilities into WTP-space again have lower coverage rates. The CIs for the WTP for operating cost have similar coverage rates, whether they were estimated by frequentist or Bayesian methods. However, the coverage rate of the CIs for the WTP for range built using the Bayesian model estimates are much lower than those built using the frequentist model estimates. Note also that while the CIs from both methods have average widths narrower than those built in preference space, the Bayesian WTP-space CIs for the WTP for range is extremely narrow, even up to a tenth of the widths of the other methods. This suggests that the method of building CIs from the

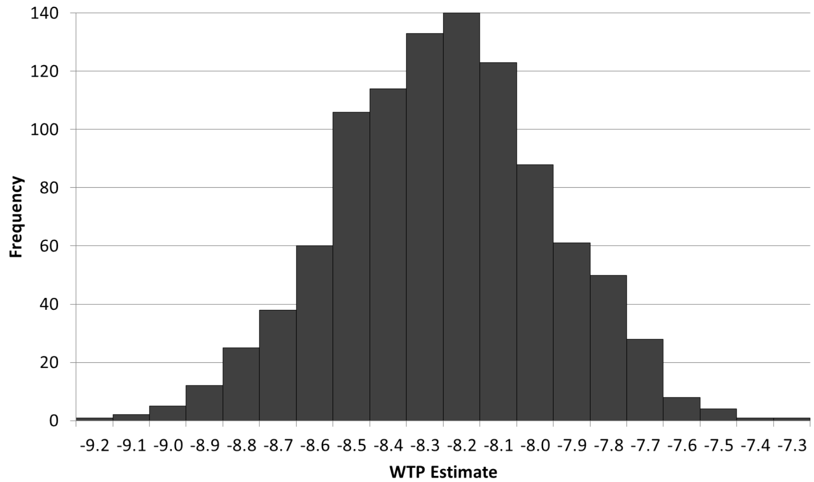


Figure 5.2: Mean WTP estimates in Bayesian WTP-space for 1000 individuals

Bayesian WTP-space model estimates need to be rethought. As can be seen, since the mean WTP for range is much larger (in magnitude) than that for operating cost, the widths of the CIs naturally increased as well. However, the width of the Bayesian WTP-space CIs for the WTP for range remained small. This narrowness combined with the large range of estimated WTP values (Figure 5.2) resulted in many of these CIs not containing the true WTP value, despite the fact that the mean WTP estimate is close to the true WTP value. (Note that this low coverage rate occurred even if the percentile CI was used rather than the HPD CI.)

5.3 Case Studies

5.3.1 Travel Mode Choice

The revealed-preference data used for this case study contains 600 individuals choosing between two different modes of travel. Each alternative contains two

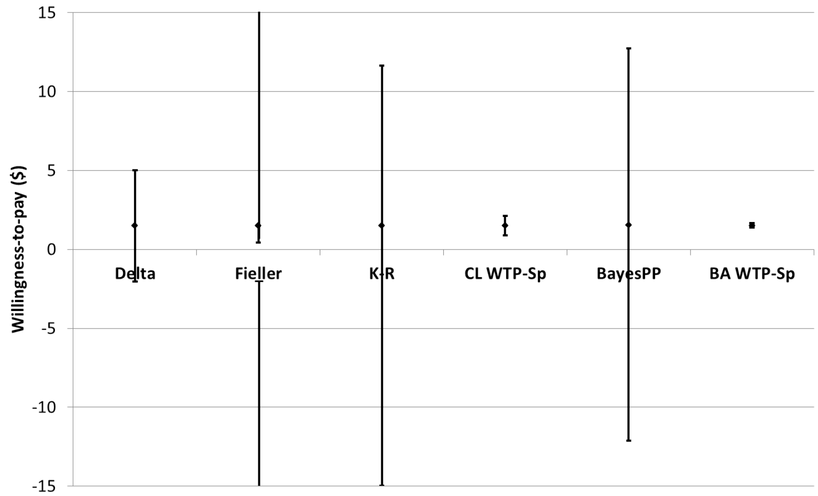


Figure 5.3: Confidence intervals for willingness to pay for time under travel mode choice

attributes, namely travel time and cost. The data set is of interest because the cost parameter is not significantly different from 0, when an alternative specific constant is included. As a result, weak identification is present, and hence we can observe the types of CIs built using the six different methods.

Figure 5.3 show that the intervals produced for this data set follow the patterns discovered in the simulation conducted in section 5.1. The WTP-space methods produce intervals that gives the shortest widths, followed by that built using the Delta Method. The Fieller method gives an unbounded interval. It is interesting to note that despite the interval being unbounded, it still contains the calculated value of time.

Note also that the intervals produced by the Delta, Krinsky-Robb and Bayesian post-processing methods contain 0. These three methods conclude that the consumer's value of time cannot be statistically distinguished from 0; a conclusion which should not logically be the case if the consumer is insensitive to cost. Hence, it is the Fieller and WTP-space methods that produce sensible

CIs. As the coverage rate of the intervals produced in WTP-space do not meet the required confidence level, it seems reasonable to conclude that the Fieller method produces the best CIs when weak identification is present.

5.3.2 Itinerary Choice

The stated preference data (Garrow et al., 2007) used for this case study contains 2907 responses from customers who were searching flight itineraries from an internet-based airline ticket booking service. The choice question was tailored to fit the origin and destinations that they had just searched for, and only customers who were looking for flights within the continental USA were recruited. Customers were asked to rank three different itineraries based on choice of airline, price, fare level of service (i.e., non-stop, single connection on the same airline, and single connection on a different airline), arrival and departure times, airplane type, and amount of legroom. Only the itinerary of the outbound flight was provided (no return or inbound itinerary was given). In addition, customers were asked whether they would continue to fly, take a different transportation mode (e.g. car, bus, train), or not make the trip at all, if they were only given these three itinerary options. Table 5.15 gives a description of the attributes that were used to generate the choice survey.

In addition to the itinerary attributes used in the choice survey, other socioeconomic, demographic, and trip context information was asked. Socioeconomic and demographic information included annual household income, gender, education, occupation and prior air travel experience, while trip context information included trip purpose, the number of people travelling together,

Table 5.15: Description of attributes in survey

Variable	Description
Departure time	This is the departure time of the first flight of the itinerary. Eight different levels are used and are applied based on the flight length and direction of travel to ensure realistic arrival times at connecting airports and destinations.
Flight time	The flight time of a connecting itinerary is assumed to be 30 minutes more than the flight time of a non-stop. These times were determined *a priori* based on the average non-stop flight time in a origin-destination pair (hence this does not vary between pairs).
Stop penalty	Increases the total travel time on connecting itineraries by 60, 90, 120 or 150 minutes (four levels).
Arrival time	Sum of departure time, stop penalty time, and non-stop flight time, adjusted for time zone changes.
Airline	Dummy variable for one of 11 airlines: American, Continental, Delta, Northwest, United, US Airways, Southwest, Alaska, AirTran, America West, and Frontier.
Airplane	Dummy variable for one of eight airplane types: regional jet, DC9, MD80, Airbus 320, and Boeing 717, 737, 757, 767.
Legroom	Dummy variable for one of four different legrooms: 2 inches less than typical, typical, 2 inches more than typical, or 4 inches more than typical.
Base fare	Equal to the average round-trip fare in the origin-destination pair multiplied by 0.75, 1.0, 1.25 or 1.5 (four levels); this is equal for all itineraries in the same choice set.
Fare premium	One of eight levels: 0.85, 0.9, 0.95, 1.0, 1.05, 1.10, 1.15 or 1.20. This is multiplied by the base fare to create the different fares for the itineraries in the survey.

and whether the trip was being paid for by a third party. The trip purpose allowed for the study and comparison of price sensitivities between passengers travelling for business as compared to leisure.

The passenger's decision to fly and their itinerary choice was modeled under both fixed and random coefficient model assumptions. For the fixed coefficient model, both logit and probit error assumptions are applied to achieve a comprehensive picture of the CIs that would be obtained. For the random coefficient models, two distributions are assumed for the coefficients of time and cost, namely the normal and lognormal distributions. However, only a logit error is assumed. CIs were built for the WTP of business and leisure passengers separately for travel time and legroom.

Fixed Coefficient Model

Table 5.16: Parameter estimates of fixed coefficient model (standard errors reported in brackets)

Variables	MNL		MNP	
	Frequentist	Bayesian	Frequentist	Bayesian
Constants (reference: no air travel)				
Nonstop	0.87 (0.13)	0.88 (0.13)	0.63 (0.08)	0.58 (0.07)
Same airline connection	-0.01 (0.20)	-0.01 (0.20)	0.53 (0.09)	0.43 (0.09)
Diff airline connection	-0.16 (0.20)	-0.16 (0.20)	0.50 (0.09)	0.38 (0.09)
Price (hundreds of dollars)				
Leisure Avg	-0.47 (0.03)	-0.47 (0.03)	-0.24 (0.02)	-0.23 (0.02)
Leisure Deviation	-1.69 (0.09)	-1.69 (0.09)	-0.29 (0.08)	-0.43 (0.07)
Business Avg	-0.25 (0.04)	-0.25 (0.04)	-0.13 (0.02)	-0.13 (0.02)
Business Deviation	-0.91 (0.17)	-0.92 (0.17)	-0.15 (0.05)	-0.23 (0.05)
Incremental flight time (hours)				
Leisure	-0.42 (0.08)	-0.42 (0.09)	-0.05 (0.02)	-0.08 (0.02)
Business	-0.70 (0.11)	-0.70 (0.11)	-0.10 (0.03)	-0.15 (0.03)
Deviation from preferred departure time				
Leisure: Departs 4 or more hours before	-0.47 (0.15)	-0.48 (0.15)	-0.16 (0.06)	-0.16 (0.06)
Leisure: Departs 4 hours before - 2 hours after	0.12 (0.11)	0.12 (0.11)	-0.07 (0.06)	-0.03 (0.05)
Leisure: Departs 2 - 4 hours after	-0.24 (0.17)	-0.24 (0.17)	-0.14 (0.06)	-0.12 (0.06)
Leisure: Departs 4 or more hours after	-0.77 (0.13)	-0.77 (0.13)	-0.25 (0.06)	-0.29 (0.06)
Business: Departs 4 or more hours before	-0.65 (0.24)	-0.67 (0.24)	-0.10 (0.09)	-0.16 (0.10)
Business: Departs 4 hours before - 2 hours after	0.73 (0.18)	0.73 (0.19)	0.12 (0.08)	0.18 (0.08)
Business: Departs 2 - 4 hours after	-0.22 (0.29)	-0.23 (0.29)	-0.01 (0.08)	-0.02 (0.10)
Business: Departs 4 or more hours after	-0.42 (0.21)	-0.42 (0.21)	-0.09 (0.09)	-0.13 (0.09)
Others				
Distance (hundreds of miles)	0.06 (0.01)	0.06 (0.01)	0.03 (0.004)	0.03 (0.004)
Income greater than $100K	0.26 (0.09)	0.26 (0.09)	0.13 (0.05)	0.13 (0.05)
Booking within 7 days of departure	0.26 (0.13)	0.26 (0.13)	0.12 (0.08)	0.11 (0.07)
Booking more than 30 days from departure	0.23 (0.09)	0.23 (0.09)	0.10 (0.05)	0.09 (0.05)
Staying more than five nights	0.16 (0.09)	0.16 (0.09)	0.10 (0.05)	0.09 (0.05)
Leg Room	0.07 (0.01)	0.07 (0.01)	0.01 (0.004)	0.02 (0.004)

Table 5.16 show the mean and standard errors of the coefficients as estimated by frequentist and Bayesian methods under both logit and probit error assumptions in preference space. Observe how, in general, both frequentist and Bayesian methods give the same coefficient estimates. The coefficients of price and time have the expected negative signs, while that of leg room has the expected positive sign. There are slight disparities between the estimates under

logit and probit error assumptions, a difference that will also be reflected in the WTP estimates.

Table 5.17 report the WTP estimates produced by the frequentist and Bayesian estimations under both parameter and WTP-space for both leisure and business passengers. As can be seen, the estimates are all very similar to each other, whether this be for the WTP for shorter travelling time or extra leg room. The probit model tends to estimate lower WTP values. For leisure passengers, the estimates of the WTP tend to be slightly higher when estimated by Bayesian methods in WTP-space. However, these differences are slight, and as will be seen, do not result in much differences in the CIs built.

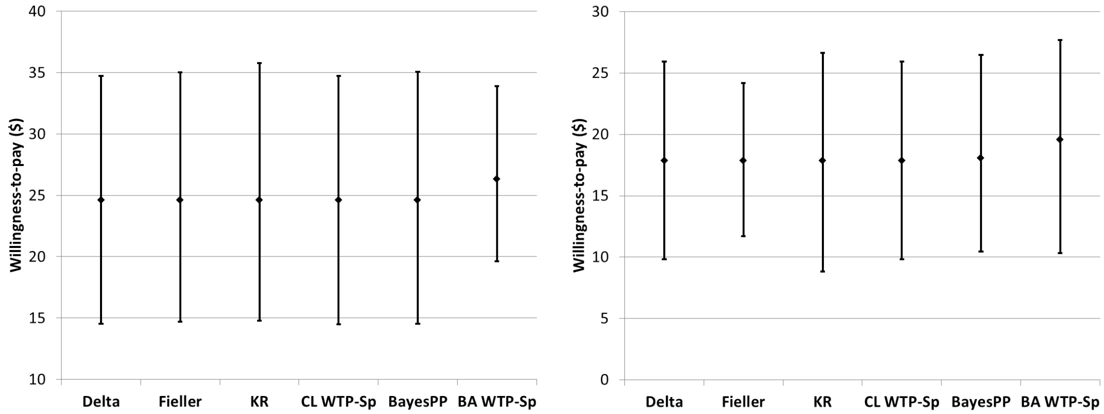
Table 5.17: Willingness-to-pay estimates

	Leisure Passengers				Business Passengers			
	Travel Time ($)		Leg Room ($)		Travel Time ($)		Leg Room ($)	
	MNL	MNP	MNL	MNP	MNL	MNP	MNL	MNP
Frequentist	24.62	17.88	4.13	4.02	76.78	62.24	7.67	7.43
FQ WTP-Sp	24.62	17.88	4.13	4.02	70.85	63.10	7.27	7.45
Bayesian	24.61	18.08	4.12	3.99	76.90	63.68	7.64	7.26
BA WTP-Sp	26.34	19.56	4.09	4.32	67.69	66.18	7.94	7.82

It is interesting to pause and observe the differences in WTP between business and leisure passengers. As can be expected, business passengers are willing to pay much more in order to cut short their travel time by an hour. They are also willing to pay more for extra leg room. There are several reasons for this. Probably most importantly, business passengers often do not pay for their own tickets. Instead, their company covers the amount, and as such these passengers would be freer to purchase a more expensive ticket since they are less sensitive to the price. As shorter travel time and more leg room raises utility (both parameters were statistically different from 0), they would not only be willing to pay for these advantages, but also be willing to pay more than leisure passengers, who are more sensitive to price. In addition, business passengers are often on a tight schedule; they have to arrive by a certain time for meetings or predetermined appointments. Hence they would value time more highly than leisure passengers, who are less likely to have to meet a fixed schedule or be willing to pay extra simply to shorten their travel. Business passengers are also willing to pay more for extra leg room; this is a factor perhaps already accounted for by airlines in that one of the perks of business class is extra space, at a higher cost.

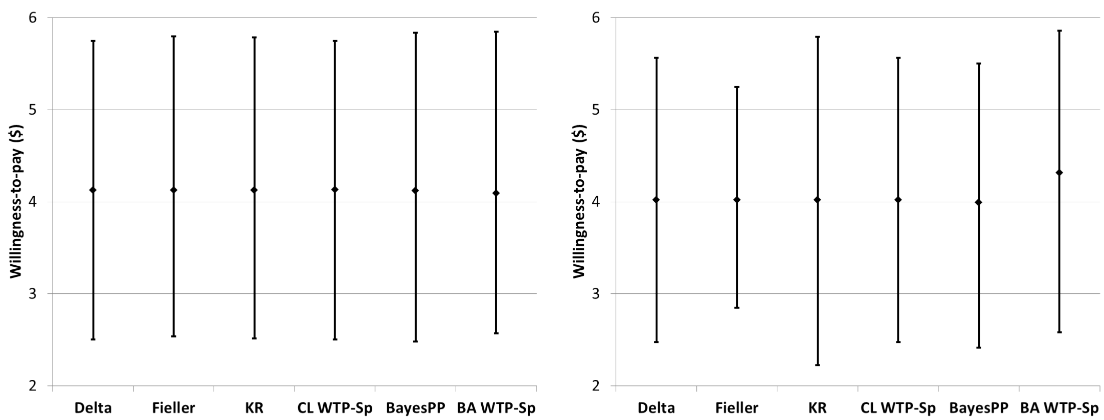
With a large data set and a cost coefficient that is statistically significant from zero, I expect that the CIs produced by the different methods will contain little disparity. Figures 5.4, 5.5, 5.6 and 5.7 show, respectively, the CIs for the WTP for shorter travel time and extra leg room for leisure passengers, followed by business passengers. Each figure compares the CIs produced under logit and probit error assumptions. As can be seen, the CIs built by all methods under most model assumptions are the same. In addition, it is interesting to observe that the WTP estimated under each error assumption (logit and probit) is contained within the CI built under the other error assumption. For example, the CI built under logit error assumptions for leisure passengers' WTP for travel time all contain the WTP value estimated under probit error assumptions. This demonstrates the viability of either model estimate; it also demonstrates the importance of CIs. Purely taking the mean WTP estimates from either model accounts to a difference (for travel time) of \$6 - \$7 per passenger per hour; an amount that builds up quickly over time and passengers. However, the CIs show that both WTP values estimated are plausible candidates for the true WTP value under 95% confidence.

The good data and parameter conditions here even result in a comparable



(a) MNL (b) MNP

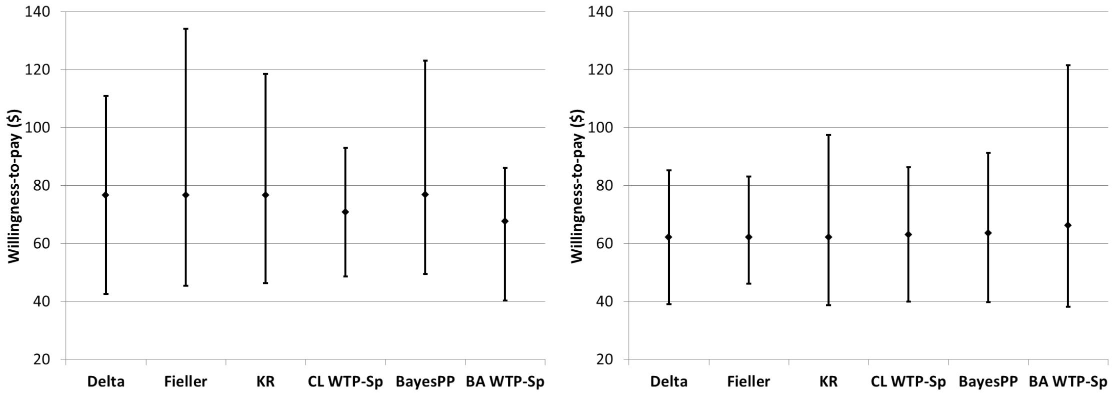
Figure 5.4: Confidence intervals for leisure passengers' willingness-to-pay for travel time



(a) MNL (b) MNP

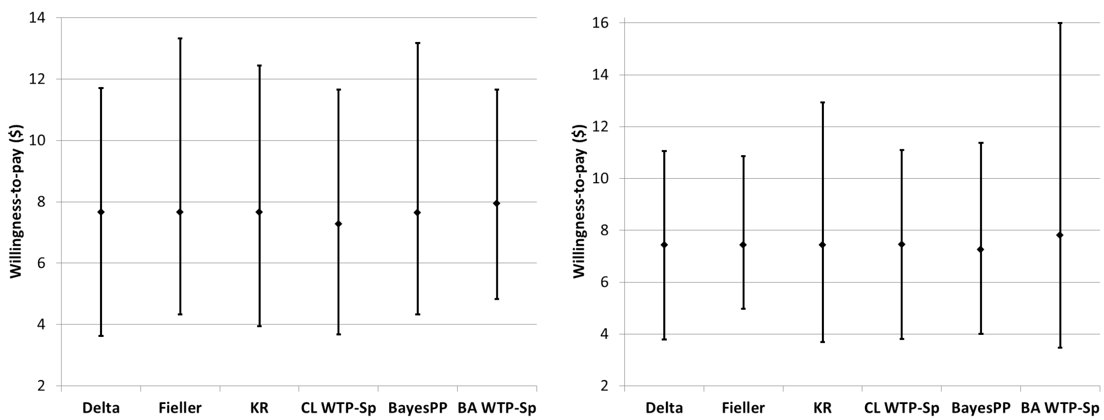
Figure 5.5: Confidence intervals for leisure passengers' willingness-to-pay for leg room

CI being built by the Bayesian model estimates under WTP-space. Based on the simulations conducted, this method does not seem to produce reliable CIs. Even under the conditions of this data set, the Bayesian WTP-space CIs differ the most from the other CIs built, especially for the larger WTP values (i.e. business passengers' WTP for travel time). This concurs with the results from the quasi-simulation in section 5.2. Interestingly, the Bayesian WTP-space CIs built under probit error assumptions for business passengers' WTP have larger widths than



(a) MNL (b) MNP

Figure 5.6: Confidence intervals for business passengers' willingness-to-pay for travel time



(a) MNL (b) MNP

Figure 5.7: Confidence intervals for business passengers' willingness-to-pay for leg room

those built by the other methods, which is counter to what has been seen in the simulations. This might simply be a result of the MCMC sample used to built the CIs. Despite all these differences, however, the CIs built still contain the mean WTP estimates obtained by the other methods of estimation, and illustrate that under these well behaved conditions, any method can be used to build the CIs.

Random Coefficient Model

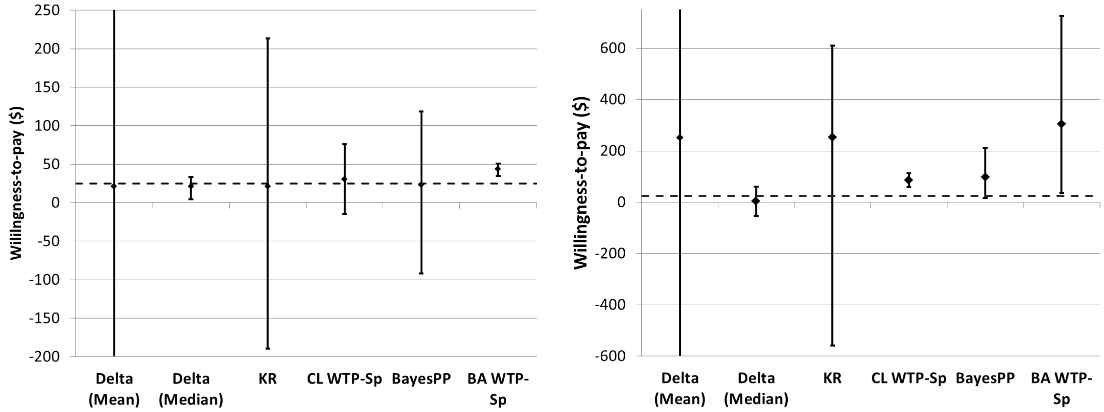
In this section, two random coefficient models are estimated via frequentist and Bayesian estimation methods. Both assume that the coefficients for price, time and leg room are random; however, the first model assumes that these variables all assume normal distributions, while the second model assumes that the variables of price and time are lognormally distributed, while that of leg room remains normally distributed. Often random variables are assumed to take on normal distributions due to ease of model estimation. However, because the range of a normal distribution is the entire $\Re$ line, such an assumption implies that there exist some passengers who have positive marginal utilities for price and time, i.e. a higher cost or longer travel time increases their utility. As this is not often a sensible assumption, lognormal distributions can be assumed for these variables. The range of the lognormal distribution is the positive $\Re$ line, thus forcing all passengers to have the same sign for their marginal utility.

Table 5.18 gives the coefficient estimates of the mean and standard deviations of the ML models estimated in parameter space. Note that the mean and standard error of the underlying normal distribution is reported in the table. The primary values of interest are the standard deviation values. These values being statistically significant (different from 0) indicates that taste variation does exist among the passengers. In general, both frequentist and Bayesian methods of estimation show that taste variation does exist among the passengers regarding their preferences for time, price and leg room.

The CIs produced by each of the methods show a great disparity. As with the simulation, the Delta (mean) method produces extremely wide CIs which are unpractical for use (the figures do not show the bounds of the Delta (mean)

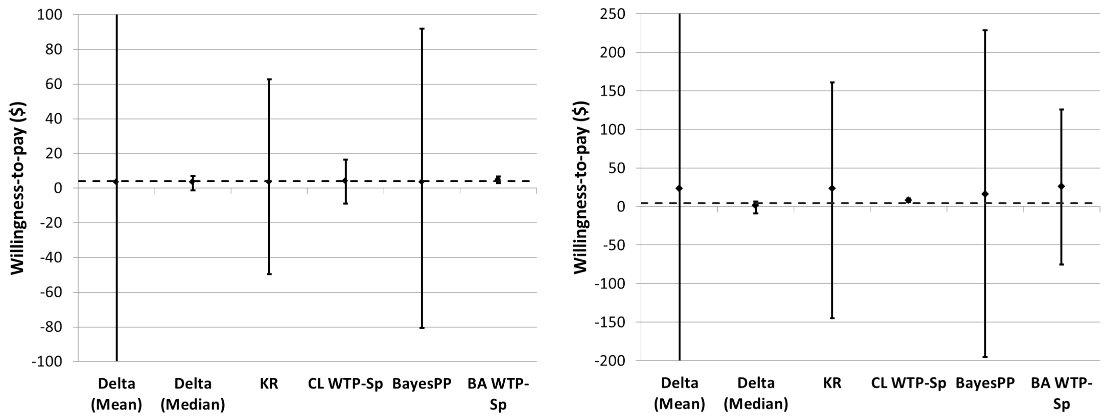
Table 5.18: Parameter estimates of random coefficient model (standard errors reported in brackets)

Variables	Normally distributed		Lognormally distributed	
	Frequentist	Bayesian	Frequentist	Bayesian
Constants (reference: no air travel)				
Nonstop	1.35 (0.24)	2.17 (0.32)	0.27 (0.14)	0.61 (0.26)
Same airline connection	0.37 (0.31)	1.02 (0.40)	-0.57 (0.30)	-1.93 (0.14)
Diff airline connection	0.16 (0.31)	0.68 (0.40)	-0.79 (0.30)	-2.02 (0.09)
Price (hundreds of dollars)				
Leisure Avg	-0.84 (0.08)	-1.89 (0.29)	-3.35 (0.89)	-4.96 (0.38)
Leisure Deviation	-2.72 (0.20)	-4.27 (0.56)	0.14 (0.14)	-0.88 (0.03)
Business Avg	-0.15 (0.20)	0.47 (0.61)	-3.24 (1.78)	-1.84 (0.38)
Business Deviation	-1.56 (0.32)	-1.97 (0.34)	0.06 (0.20)	-0.17 (0.04)
Incremental flight time (hours)				
Leisure	-0.58 (0.12)	-0.98 (0.27)	-0.34 (0.30)	-1.07 (0.08)
Business	-1.19 (0.32)	-1.44 (0.43)	-0.21 (0.49)	-1.13 (0.07)
Deviation from preferred departure time				
Leisure: Departs 4 or more hours before	-0.63 (0.23)	-1.56 (0.39)	-0.72 (0.18)	-0.41 (0.10)
Leisure: Departs 4 hours before - 2 hours after	0.21 (0.18)	-0.27 (0.26)	0.01 (0.13)	0.38 (0.09)
Leisure: Departs 2 - 4 hours after	-0.31 (0.25)	-1.12 (0.26)	-0.39 (0.20)	-0.18 (0.10)
Leisure: Departs 4 or more hours after	-1.11 (0.22)	-2.39 (0.39)	-1.10 (0.15)	-0.69 (0.04)
Business: Departs 4 or more hours before	-1.08 (0.41)	-1.76 (0.50)	-0.34 (0.28)	-0.42 (0.09)
Business: Departs 4 hours before - 2 hours after	1.01 (0.29)	1.03 (0.34)	1.05 (0.19)	0.93 (0.18)
Business: Departs 2 - 4 hours after	-0.21 (0.40)	-1.07 (0.63)	0.05 (0.34)	-0.15 (0.13)
Business: Departs 4 or more hours after	-0.80 (0.35)	-1.93 (0.29)	-0.18 (0.22)	-0.04 (0.07)
Others				
Distance (hundreds of miles)	0.12 (0.02)	0.41 (0.09)	0.003 (0.007)	-0.03 (0.01)
Income greater than $100K	0.39 (0.16)	0.88 (0.50)	0.33 (0.10)	0.01 (0.09)
Booking within 7 days of departure	0.40 (0.24)	1.69 (0.36)	0.27 (0.15)	-0.33 (0.12)
Booking more than 30 days from departure	0.46 (0.17)	1.68 (0.36)	0.27 (0.10)	-0.05 (0.12)
Staying more than five nights	0.30 (0.17)	1.23 (0.29)	0.07 (0.10)	0.08 (0.11)
Leg Room	0.09 (0.02)	0.14 (0.03)	0.08 (0.02)	0.06 (0.19)
Standard Deviations				
Leisure Avg Price	0.69 (0.12)	2.82 (0.51)	1.14 (0.60)	2.36 (0.25)
Leisure Deviation Price	2.07 (0.30)	3.37 (0.58)	1.91 (0.08)	0.33 (0.02)
Business Avg Price	0.88 (0.28)	3.67 (0.91)	0.87 (1.77)	2.28 (0.18)
Business Deviation Price	1.42 (0.72)	1.86 (0.39)	0.04 (1.10)	0.40 (0.03)
Leisure Incremental Time	0.11 (0.19)	0.78 (0.25)	1.13 (0.28)	0.50 (0.04)
Business Incremental Time	0.82 (0.32)	1.08 (0.39)	1.89 (1.26)	0.57 (0.04)
Leg Room	0.17 (0.06)	0.14 (0.03)	0.23 (0.06)	0.36 (0.03)



(a) Normal Distribution (b) Lognormal Distribution

Figure 5.8: Confidence intervals for leisure passengers' willingness-to-pay for travel time

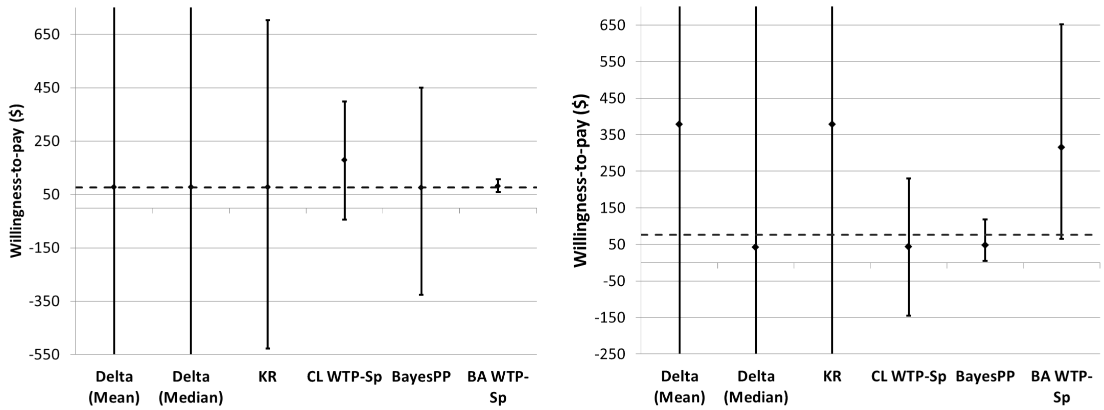


(a) Normal Distribution (b) Lognormal Distribution

Figure 5.9: Confidence intervals for leisure passengers' willingness-to-pay for leg room

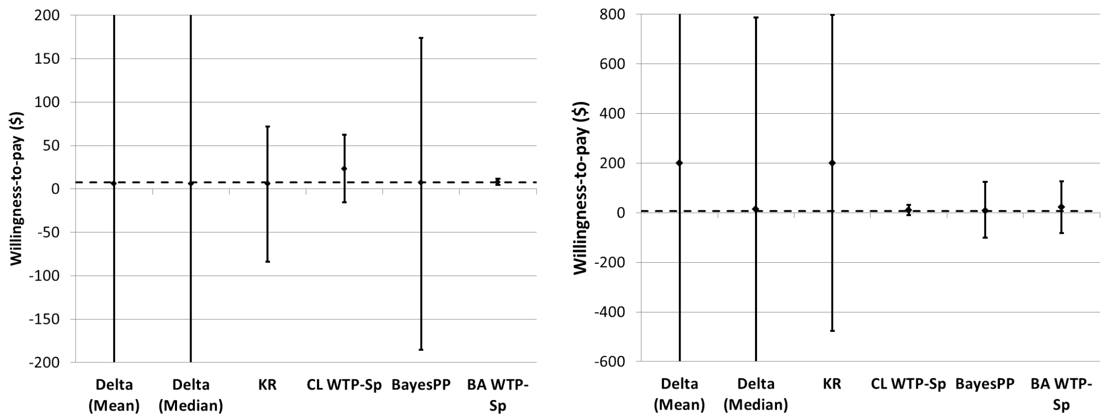
intervals due to its large width compared to the other intervals). The Delta (median) method builds much narrower intervals in comparison. Interestingly, the intervals built for leisure passengers are much narrower than those built for business passengers. Again, as before, the CIs built under WTP space are generally narrower than those built under preference space.

In general, these CIs are much wider than those built when a fixed coefficient model is estimated. This is unsurprising given that by construction, the random



(a) Normal Distribution (b) Lognormal Distribution

Figure 5.10: Confidence intervals for business passengers' willingness-to-pay for travel time



(a) Normal Distribution (b) Lognormal Distribution

Figure 5.11: Confidence intervals for business passengers' willingness-to-pay for leg room

coefficient model would result in problems of weak identification. It is useful to note that most of the CIs built under the random coefficient model assumptions do cover the fixed coefficient model WTP estimates (as indicated by the dashed line in the figures).

In addition, most of the CIs contain 0, indicating that we cannot reject the hypothesis that the WTP is equal to 0. This is again not surprising due to the fact that we still assume that the WTP follows a normal distribution despite the coefficients assuming lognormal distributions. It is here that the Bayesian method has an advantage over the others; because the Bayesian estimator uses MCMC samples from the posterior, it avoids the assumption of normality. Hence, as can be seen, the Bayesian CIs for WTP for less travel time do not contain 0. Those CIs for WTP for extra leg room still contain 0 due to the assumption of normality for the leg room coefficient.

CHAPTER 6

BEYOND WILLINGNESS-TO-PAY: BUILDING CONFIDENCE INTERVALS FOR INDIVIDUAL LIKELIHOODS

CIs are not just built for WTP measures or other ratio measures of parameter estimates. Rather, CIs can (and should) be built for all variables that are random. I illustrate this with an example in which CIs should be reported, and has not (so far) been often done. This is in the airline industry, where revenue management is used to determine the number and price of seats to make available to customers. Although every aircraft has a maximum capacity, airlines often sell more seats than are available as some customers will either cancel their tickets prior to departure or fail to show up on the day of departure. Overbooking allows airlines to minimize the number of empty seats on a departing flight. Cancellation and no-show models are used to determine these overbooking levels. These models allow airlines to not only maintain efficiency on each flight, but also prevent the excessive sale of tickets, which would result in the need to compensate passengers and thus reduce revenue.

In order to understand a passenger's reasons for changing his flight, it is necessary to analyze his behavior. As early as the 1980s, attempts have been made to integrate discrete choice models of passenger behavior into revenue management (Garrow, 2010). Examples include Belobaba (1989) and Brumelle and McGill (1993), who model airline seat allocation with a probabilistic decision model and multiple nested fare classes respectively. However, with few exceptions, these attempts gave way to the use of more simplistic probability models or time-series methodologies that, while easier to implement, make strong independence assumptions. For example, most models assumed that the demand for a particular booking class for a flight was independent from that of all other booking classes on that (and other) flights. In recent years, this and other revenue management algorithm assumptions have been questioned (Boyd, 2004; Boyd and Kallesen, 2004; Dunleavy and Westermann, 2005; Hornick, 2004; Lieberman, 2004; Oliveira, 2003; Ratliff, 1998; Talluri and van Ryzin, 2004), resulting in a re-examination of how discrete choice models could be used to model the individual passenger's behavior. In particular, studies have attempted to apply discrete choice methods to cancellation models. For example, Talluri and van Ryzin (2004) model survivals using a binomial distribution; however, this distribution contains inherent assumptions that are violated when observing passenger behavior (Westerhof, 1997; Chatterjee, 2001). In particular, the binomial distribution assumes that customers cancel independently of each other, that each customer has the same probability of cancellation, and that cancellation probabilities are "memoryless", i.e., they depend only on the time to flight departure and not on when the ticket was first booked.

In recent years, there has been more work done in integrating passenger behavior models with revenue management models, and particularly with cancellation models (Iliescu et al., 2008; Graham et al., 2010). However, little work has been done in building CIs for these cancellation probabilities. As with all random variables, cancellation probabilities are not fixed values which can be used to precisely predict the number of people who will cancel their air tickets. Rather, CIs should be reported in order to give a better insight into the reliability and usefulness of the probability estimates.

In addition, the nature of airline ticketing data is such that the passengers who bought tickets are identifiable. This allows for two further refinements of the model assumptions. Firstly, it allows for an estimate of parameters by individual (i.e., using a model with random taste variations), which would result not only in more accurate estimates of cancellation probabilities, but also the ability to determine such probabilities on an individual level. Secondly, this makes it possible to trace an individual passenger's historical cancellation behavior. This historical behavior can be applied as a prior to determine the passenger's current cancellation probabilities. These two aspects can be incorporated to further refine the estimates of the cancellation probabilities and the CIs built. In addition, CIs can now be built for the individual cancellation probability rather than simply the aggregate.

My methodology for this chapter is as follows. I use a random coefficient model to study a set of airline ticketing data containing ticket characteristics, as well as identifiable passengers. CIs are then built for the individual likelihood of a passenger having performed a certain event, whether this be cancelling or using his ticket. These CIs are built using the Bayesian post-processing method. As seen in chapter 5, the Bayesian post-processing method is a viable method for building CIs, especially with large sample sizes and without the problems of weak identification. There are other advantages to using the Bayesian estimation process, from which the Bayesian post-processing CIs are built. Firstly, using a hierarchical Bayes model can generate individual-specific parameters, and thus individual-specific probabilities. Current probabilities of passenger cancellations being reported are aggregate measures, but it would be of greater use to have individual passenger probabilities based on the ticket characteristics and passenger cancellation history. Furthermore, since the Bayesian methods take draws from the posterior distribution of the parameters, post processing methods can be used to obtain individual-specific Bayesian confidence intervals for these probabilities.

6.1 Literature Review: Discrete Time Proportional Odds Model

Much of this study builds off the research of Iliescu et al. (2008) and Graham et al. (2010). To study cancellation behavior and incorporate passenger behavior into the cancellation model, the authors make use of the discrete time proportional odds (DTPO) model (Iliescu et al., 2008). As a survival analysis method, this model was applied to calculate the hazard probability, i.e. the conditional probability that a ticket would be cancelled on a certain day given that it had survived until that day. By using this model, the authors were not only able to analyze how the cancellation probabilities changed over time and how they were affected by ticket characteristics, they were also able to find empirical evidence for the violation of the assumptions inherent in the use of the binomial distribution by Talluri and van Ryzin (2004).

The DTPO model partitions the time-to-event (where the event is either the ticket being cancelled, the passenger not showing up for the flight, or the passenger using the ticket and departing) of the i^{th} ticket (T_i) into k disjoint time intervals $(t_0, t_1], (t_1, t_2], \ldots, (t_{k-1}, t_k]$, where $(t_0, t_1, \ldots, t_k)$ identify the days from issue of the ticket, and t_0 and t_k are the issue date and the time of event respectively. Additionally, the discrete hazard probability (that is, the probability of the cancellation event) is given by $h_{ij} = P(T_i = j | T_i \geq j)$. By using conditional probability theory, the likelihood function of the entire sample can be written out explicitly. In particular, suppose we have the choice variable y_{ij} where

$$y_{ij} = \begin{cases} 1, & \text{if ticket } i \text{ is cancelled } j \text{ days from issue} \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

Then the likelihood function of the entire sample becomes

$$L = \prod_{i=1}^{n} \prod_{j=1}^{k} h_{ij}^{y_{ij}} (1 - h_{ij})^{(1-y_{ij})} \tag{6.2}$$

The DTPO model incorporates the ticket characteristics (known as covariates) into the model through the following formulation. For a set of covariates X_i, the hazard probability h_{ij} is

$$\log\left(\frac{h_{ij}}{1-h_{ij}}\right) = \Psi_{ij} + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_l X_{ijl}, \tag{6.3}$$

where Ψ_{ij} is the baseline hazard function, $j = 1, 2, \ldots, k$ time intervals, $i = 1, 2, \ldots, n$ observations and l is the number of covariates. Using this formulation, the DTPO model can be solved for the parameters β.

However, since the exact time at which the ticket is cancelled is known and can be written using a binary variable (y_{ij} above), the likelihood function is equivalent to that of a binary logistic regression model. Recall that the logistic regression model is described as follows:

$$\begin{aligned} y_i^* &= X_i'\beta + \varepsilon, \text{ where } \varepsilon \sim \text{Logistic}(0,1) \\ y_i &= \begin{cases} 1, & \text{if } y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{6.4}$$

Then the probability that $y_i = 1$ given covariates X_i, $p_i = P(y_i = 1|X_i) = \frac{e^{X_i'\beta}}{1+e^{X_i'\beta}}$. Again using conditional probability theory, the likelihood function of the entire sample becomes

$$L = \prod_{i=1}^{n} \left(\frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}} \right)^{y_i} \left(\frac{1}{1 + e^{X_i' \beta}} \right)^{(1-y_i)} \tag{6.5}$$

Note that equations (6.2) and (6.5) are equivalent if the baseline hazard function Ψ_{ij} is added as a covariate. Hence the DTPO model is equivalent to that of a binary logistic regression model, which has known numerical solutions and thus makes the DTPO model analytically attractive.

Iliescu et al. (2008)'s use of the DTPO model builds on prior research in four ways. First, while most survival models consider a single time dimension, this model accounts for multiple time dimensions by accommodating time-varying covariates. Second, it allows for the analysis of the effect of ticket heterogeneity on cancellation probability. Third, it assumes that the ticket heterogeneity is fully contained within the covariates used, and that its effect on cancellation probability is separate from that of time. Finally, due to the discrete nature of the time-scale used, the model is able to test for different distributional structures of the baseline cancellation rate.

The study found that cancellation probabilities depended on multiple time dimensions; in fact, recently purchased tickets and those with nearer departure dates tended to have higher cancellation probabilities. In addition, the study successfully incorporated customer behavior into cancellation models, finding that certain covariates were associated with lower cancellation rates.

Following this study, Graham et al. (2010) used the DTPO model to examine business travelers' cancellation behavior. The nature of the data allowed the authors to track individual passengers over time, and they found that frequent travelers were 1.4 times more likely to cancel their tickets than non-frequent travelers. In addition, the authors analyzed the effect of specific covariates, such as whether the ticket was discounted. This study further demonstrated how passenger behavior plays an important role in cancellation rates and revenue management.

This chapter aims to build on these previous works through the application of a random parameter model and the use of historical data. In particular, we are able to sort ticket purchases according to who made the transaction, and hence know how many tickets an individual purchased, and how many of these were cancelled. Graham et al. (2010) found that frequent travelers are more likely to cancel their tickets than infrequent travelers. However, they did not use the individual passenger's past cancellation history, but divided travelers into two groups (frequent versus infrequent) and determined the percentage of tickets that were cancelled. In addition, it seems logical that a passenger's history of ticket purchase and cancellation behavior should be used to better estimate his current probability of cancelling a ticket. All these refinements to the model aim to give accurate estimates of the cancellation probabilities, and also accurate constructions of the confidence interval of the individual likelihood of a passenger having performed a certain event, whether this is cancelling or using his ticket.

6.2 Methodology

6.2.1 Data

The dataset used contains information of 3542 tickets purchased by a dedicated ticket agency on behalf of Georgia Institute of Technology (Georgia Tech). As a state institution, individuals travelling on official business on behalf of the university are eligible to receive discounts on air travel from the State of Georgia. Individuals do so by purchasing their ticket through this ticket agency, who issues all discounted air tickets on behalf of the university (with few exceptions). In addition, they issue a vast majority of the non-discounted air tickets, with an estimated 70-80% of all Georgia Tech airline travel being handled by this agency.

The dataset contains primarily information about the ticket that was purchased. In particular, only the passenger's name and a university issued ID number identify the passenger using that ticket. The remaining variables all describe the ticket that was purchased, as detailed in Table 6.1. Of particular interest are the variables showing the time from departure that the ticket was booked, and the time from issue that an event occurred (either the ticket was used or cancelled). These were shown by Iliescu et al. (2008) to have a significant effect on ticket cancellation rates.

Out of 3542 tickets, 525, or 14.8%, were cancelled prior to departure. In addition, there are a total of 1870 passengers, 639 (34.2%) of whom bought more than one ticket (for a total of 2311, or 65.2%, of the tickets). This gives us a large dataset on which to estimate the cancellation model. Further details of the dataset can be found in Graham et al. (2010).

Table 6.1: Attributes in dataset describing ticket characteristics

Variable	Distributional Characteristics
Days before departure that ticket is booked	Ranges from 1 - 60; majority is from 1-7 days
Days after issue that event occurs	Ranges from 0 - 60; majority is from 1-8 days
Departure Day of Week	Sun 16%; Mon 18%; Tues 17%; Wed 18%; Thurs 15%; Fri 9%; Sat 7%
Departure Month	Jan 12%; Feb 10%; Mar 2%; Jul 7%; Aug 10%; Sep 14%; Oct 18%; Nov 18%; Dec 9%
Carrier	Delta 83%; AirTran 17%
Trip origin in Atlanta	87% originate in Atlanta
State Discount	91% receive state discount
Length of Stay (number of nights)	Same Day 9%; One 27%; Two 24%; Three 17%; Four 10%; Five 5%; Six 3%; Seven or more 5%
Departure Hour	89% depart between 8:00 AM - 9:59 PM, and are distributed fairly evenly
Arrival Hour	94% arrive between 8:00 AM - 10:59 PM; most popular times are between 4:00 PM - 7:59 PM

Because the data set contains only information from business travelers, certain otherwise important variables will be statistically insignificant when the cancellation model is estimated using this data. For example, Graham et al. (2010) found that only the variables of days of departure, days from issue state rate and carrier were statistically significant, while other variables such as departure day of week and length of stay were not. Such variables are generally useful for distinguishing between leisure and business passengers, hence accounting for their statistical insignificance. While these results are not representative of a general population and should not be taken as such, they still provide useful insights into the business travelling community.

6.2.2 Reformulating the DTPO model and Data Manipulation

In order to easily use Bayesian methods to solve the DTPO model, the model was rewritten into that of a discrete choice model formulation, while maintaining the same assumptions as the DTPO model. The reason for doing this is twofold. Firstly, although the DTPO model formulation is fairly complicated, it is equivalent to that of a logistic regression model (Iliescu et al., 2008), which is a standard model that is easy to solve. Secondly, pre-written functions exist (such as *mlogit* in the R programming language) which solve discrete choice models. By rewriting the DTPO model equations, the existing computer functions can be used to estimate the DTPO model.

To use these existing computer functions for solving discrete choice models via Bayesian methods, the model must be reformulated into one with two latent variables. In particular, the cancellation data can be modeled as an individual making a choice, in each time period, whether to cancel his ticket (alternative 1), or not (alternative 2). This reformulation of the logistic regression model into a sequence of binary choice situations provides an intuitive representation of the cancellation behavior of a passenger. Specifically, as the days pass from the purchase of a ticket, a passenger can decide whether or not to cancel his ticket. This is clearly modelled in the sequence of binary choices, where for every time period (day), the individual chooses between two alternatives - whether to cancel his ticket or not. In addition, since some of the covariates (in particular, those associated with the number of days before departure and the number of days after issue) are time dependent, this model represent a dynamic choice problem. Hence this model, unlike other survival models, is also able to represent multiple time dimensions due to its inclusion of time-dependent covariates.

For each binary choice situation, the individual has a utility associated with each choice:

$$\begin{aligned} U_1 &= X_1'\beta + \varepsilon_1 \\ U_2 &= X_2'\beta + \varepsilon_2 \end{aligned} \quad \text{, where } \varepsilon_1, \varepsilon_2 \sim \text{EV1} \tag{6.6}$$

As is standard in random utility maximization models, the only concern is with the utility in differences, i.e., the difference in the utilities obtained from choosing alternative 1 or 2. Hence the following model is achieved:

$$U = U_1 - U_2 = (X_1 - X_2)'\beta + (\varepsilon_1 - \varepsilon_2), \text{ where } \varepsilon_1 - \varepsilon_2 \sim \text{Logistic} \qquad (6.7)$$

Setting all X_2 variables to 0 makes this discrete choice model equivalent to the logistic regression model, and thus the original DTPO model.

Using the discrete choice model involves expanding the original data set to include the attributes of the second alternative. Table 6.2 shows how the dataset is created, with Table 6.2(a) being the original dataset used when estimating the DTPO model, and Table 6.2(b) being the dataset used for estimating the equivalent discrete choice model.

Table 6.2: (a) Original dataset used for DTPO model in Graham et al. (2010). Table shows customer ID, their choice (to cancel or not), and variables of ticket characteristics; (b) Dataset used for discrete choice model, where the variables of ticket characteristics have been expanded to include those of alternative 2 (not cancelling)

Customer	**choice**	**DFI**	**DFI_0_3**	**DFI_4_7**	**DFD**	**...**	**A22**
52463	2	0	1	0	6		1
52463	2	1	1	0	5		1
52463	2	2	1	0	4		1
52463	2	3	1	0	3		1
52463	2	4	0	1	2		1
52463	2	5	0	1	1		1
52463	1	6	0	1	0		1

(a)

Customer	**choice**	**DFI.1**	**DFI.2**	**DFI_0_3.1**	**DFI_0_3.2**	**DFI_4_7.1**	**DFI_4_7.2**	**DFD.1**	**DFD.2**	**...**	**A22.1**	**A22.2**
52463	2	0	0	1	0	0	0	6	0		1	0
52463	2	1	0	1	0	0	0	5	0		1	0
52463	2	2	0	1	0	0	0	4	0		1	0
52463	2	3	0	1	0	0	0	3	0		1	0
52463	2	4	0	0	0	1	0	2	0		1	0
52463	2	5	0	0	0	1	0	1	0		1	0
52463	1	6	0	0	0	1	0	0	0		1	0

(b)

To my knowledge, setting all X_2 variables to 0 in this way to model the logistic regression (and thus the DTPO) model has not been done before. However, much of the literature which involves an opt-out alternative uses this method of setting all variables for that alternative to 0. An opt-out alternative is where the individual taking a survey is given the option to choose either one of the scenarios portrayed or to choose none of the scenarios portrayed. Hence, the individual ”opts out” of choosing an available scenario or alternative. Many surveys include the opt-out alternative as a means of better simulating a real life situation. Examples include Garrow et al. (2007) and Veldwijk et al. (2014).

6.3 Results

Although the focus of this chapter are the results from the random coefficient model, I first estimated the data using a MNL model. This being the simplest model to estimate, I also wanted to compare my results to that reported in Graham et al. (2010), who do not study random coefficients. In addition, the data obtained for this study was updated from that used by Graham et al. (2010), and hence I felt it necessary to have some basis for comparison also with the random coefficient model.

6.3.1 Multinomial Logit Model

In estimating the MNL model, the fact that certain individuals purchased multiple tickets was not taken into account. Hence, each ticket purchase is regarded as having been bought by a separate individual. Results from both the frequentist and Bayesian model estimates are reported in Table 6.3.

As can be seen most of the results obtained from the two methods are fairly

Table 6.3: Parameter estimates of multinomial logit model estimated using both frequentist and Bayesian methods. DFI = number of days from issuing the ticket, and DFI_0_3, DFI_4_7 = binary variables for 0-3 and 4-7 days from issue respectively. DFD = number of days from departure. FL = carrier binary variable (Delta airlines as base). ATLDST = Atlanta origin binary variable. STATE = state rate binary variable. Significant codes: *** = 99.9%; ** = 99%; * = 95%

Parameters	Frequentist	Bayesian		Parameters	Frequentist	Bayesian	
CONST	-8.47	-8.80	***	D9	-0.57	-0.48	**
DFI	0.0083	0.0066		D10	-0.68	-0.55	**
DFI_0_3	0.99	0.96	***	D11	-0.94	-1.15	***
DFI_4_7	0.80	0.80	***	D12	-0.87	-0.76	**
DFD	-0.031	-0.025	***	D13	-0.40	-0.30	*
SUN	-0.18	-0.12		D14	-0.86	-0.90	***
MON	-0.26	-0.12		D15	-0.53	-0.50	*
TUES	-0.01	0.14		D16	-0.71	-0.79	***
WED	0.14	0.31		D17	-0.61	-0.53	**
THURS	-0.13	-0.03		D18	-0.62	-0.51	*
FRI	0.09	0.31		D19	-0.71	-0.50	**
JAN	-0.20	0.01		D20	-0.36	-0.06	
FEB	-1.10	-0.96	***	D21	0.54	0.55	
MAR	-1.62	-1.62	***	A8	-0.42	-0.58	
JUL	-0.92	-0.66	**	A9	0.71	0.52	**
AUG	-0.47	-0.32	*	A10	0.32	0.06	
SEP	-0.11	0.10		A11	0.51	0.14	
OCT	-0.31	-0.18		A12	0.26	0.12	
NOV	-0.19	0.10		A13	-0.37	-0.47	
FL	-0.74	-0.69	***	A14	-0.26	-0.28	
ATLDST	-0.11	-0.13		A15	-0.52	-0.74	
STATE	0.70	0.58	**	A16	0.11	-0.18	
LOS1	4.17	4.43	***	A17	-0.30	-0.51	
LOS2	3.95	4.08	***	A18	-0.41	-0.60	
LOS3	3.96	4.14	***	A19	-0.11	-0.30	
LOS4	4.07	4.28	***	A20	1.34	1.03	***
LOS5	4.18	4.13	***	A21	0.04	0.25	
LOS6	3.84	4.31	***	A22	-0.19	-0.30	
LOS7P	4.97	5.08	***				
D8	-0.45	-0.26	*	LLH	-2713.6	-2752.7	

similar, and differences tend to exist in those variables which are not statistically significant with at least 95% confidence. Log-likelihood values have also been reported for the two models, but note that the Bayes estimator does not seek to maximize the likelihood function, and thus these should not be used as a basis for comparison.

There are many variables which affect the cancellation rate. In particular, binary variables indicating number of days from ticket issue being 0-3 and 4-7 (DFI_0_3 and DFI_4_7 respectively), number of days from departure (DFD), whether the plane was Delta or AirTran (FL), and whether the ticket was state discounted (STATE) were all significant with at least 95% confidence, in agreement with Graham et al. (2010). There are also many other variables that are significant, in contrast to the results of Graham et al. (2010). However, as the dataset in this paper is slightly different from that used by Graham et al. (2010), it is to be expected that there will be some differences.

Table 6.4 shows the odds ratio for those variables listed above that were also found to be significant by Graham et al. (2010). Cancellations are about three times more likely to occur 3 days or less from issue date, and about two times more likely to occur 4-7 days from issue date (as opposed to 8-60 days from issue date). Similarly, tickets are more likely to be cancelled as the date of departure approaches. There is a clear difference between travelling on a legacy carrier (Delta) as opposed to a budget carrier (Air Tran), with tickets purchased on Air Tran being approximately half as likely to be cancelled (as illustrated by the variable FL). In addition, tickets receiving a state discount are about two times more likely to be cancelled than those not receiving the state discount.

Table 6.4: Odds ratios of significant variables

Variables	Frequentist	Bayesian
DFI_0_3	2.70	2.64
DFI_4_7	2.22	2.22
DFD	0.97	0.97
FL	0.48	0.50
STATE	2.02	1.79

6.3.2 Mixed Logit Model

By identifying passengers who have bought more than one ticket with this ticket agency, I was able to estimate a ML model in order to test for random taste variation among passengers. Based on the results of estimating the MNL model and Graham et al. (2010), only the variables concerning date from issue, date from departure, carrier type, and state were assumed to be randomly normally distributed.

The ML model was estimated with the Bayesian method using only the ticket data from passengers who purchased multiple tickets. This resulted in a data set containing 639 passengers having purchased 2311 tickets. Passengers purchased between 2 to 23 tickets, with the majority (293 passengers, or 45.8%) having purchased two tickets.

Table 6.5: Mean and standard errors of the estimated standard deviation values of normally distributed random variables. Significant codes: ***=99.9%; **=99%

Variable	Mean	Std Err	
DFI_0_3	0.59	0.18	**
DFI_4_7	0.49	0.09	***
DFD	0.16	0.0032	***
FL	0.65	0.25	**
STATE	0.54	0.14	***
Log-likelihood	-1824.06		

The assumption that the variables are normally distributed is proven accurate if the estimated standard deviations are found to be significant. As can be seen in Table 6.5, these standard deviations were found to be statistically significant with at least 99% confidence, hence affirming that random taste variation exists amongst passengers for these variables.

As stated at the beginning of this chapter, one of the reasons for using the Bayesian method of estimation is so that individual-specific probabilities can be estimated. I illustrate how these can be used to create individual probabilities for certain events happening, as well as CIs for these events. I create a hypothetical situation in which a passenger books a ticket five days before the scheduled departure date. The flight is booked with Delta airlines, and originates from Atlanta. It is scheduled to depart at 5:15 PM and arrive at its destination at 7:45 PM. Finally, the passenger is staying away from home for two nights, and receives a state rate when purchasing the ticket.

With this hypothetical ticket, I investigate the probability that an individual will cancel his ticket on the day before departure. This probability and accompanying CI are calculated for all the individuals who were included in the taste variations model. Figure 6.1 gives the means and CIs for 100 sampled individuals' probabilities of cancelling the ticket described one day before departure.

It is also possible to calculate the probabilities and CIs of a particular individual's decision process on a daily basis. As an example, I again use the hypothetical ticket as described above. For any given individual, knowing his taste preferences, I can build the CIs (as in Figure 6.2) for the probability that he cancels on any given day from the time he purchased his flight to the day he flies (technically a no-show), and also the probability that he does use the ticket (this was not included in the figure due to the large difference in magni-

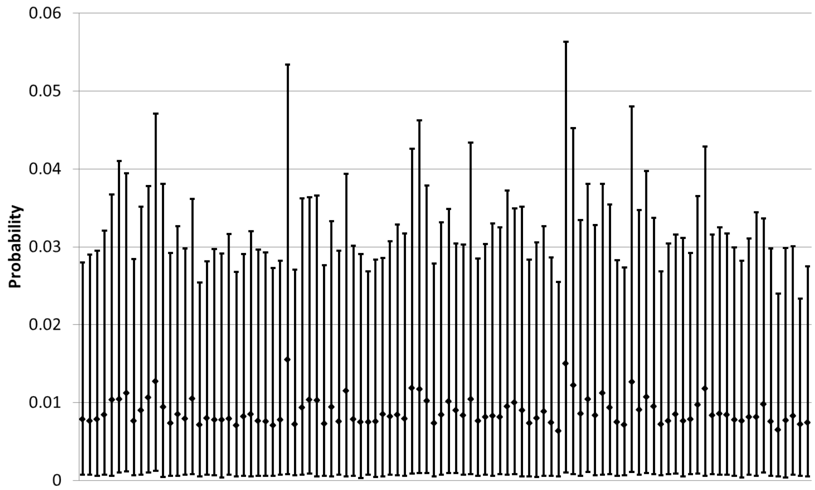


Figure 6.1: Individual means and confidence intervals for probability of cancellation

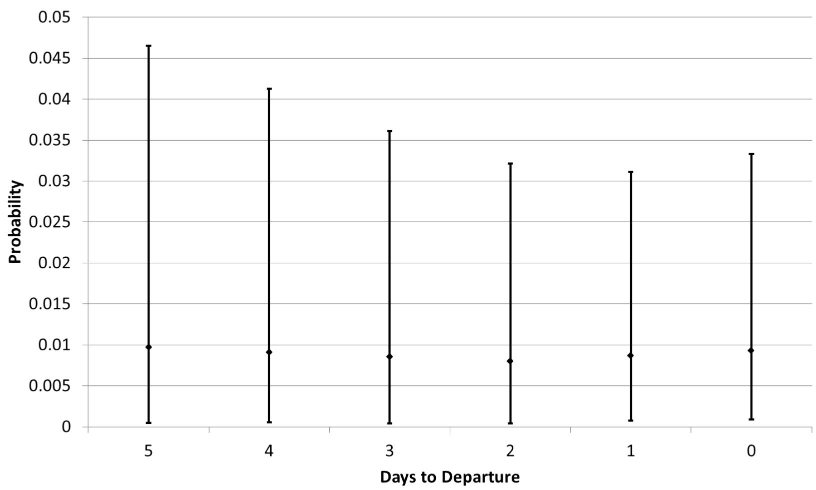


Figure 6.2: Means and confidence intervals for probability of cancellation by day

tudes, which would have made it impossible to see the CIs of the probabilities of cancellation).

CHAPTER 7

CONCLUSION

The primary goal of this research was to investigate and develop methods for the building of CIs, with a large focus on building CIs for WTP measures. Building and reporting CIs are important endeavors due to the importance and practical uses of CIs. CIs are important as a statistical measures of uncertainty for random variables, and while some may claim that hypothesis testing serves the same purpose, CIs are superior on many counts, including that they account for the imprecision of the statistic in question that arises both from the variability of the parameter and the limited sample size. This last is an important push for using CIs within the social sciences and transportation science, which often makes use of data from surveys that have been conducted with a limited number of people.

The reporting of CIs for unknown coefficients is important, but in particular, CIs should be reported for WTP measures. In the context of a linear-in-attributes discrete choice model, the WTP for a certain attribute is calculated as the ratio of the coefficients of the attribute in question and cost. Since WTP is associated with the model coefficients, it is easy to recognize the necessity of CIs when using a random coefficient model. However, many people do not remember or realize that even the coefficients of a supposedly "fixed" coefficient model is asymptotically distributed normal. Hence, the WTP is not a fixed variable, but random, following an *a priori* unknown distribution, and as a result, CIs should be reported along with the mean value of the WTP.

Since CIs are necessary, especially in applied work, the accurate building of CIs is also of great importance. This is not a trivial problem to solve. Particularly when building CIs for WTP (or in general, any ratio measure), one runs into two main problems: the unknown probability distribution of WTP, and the possible discontinuity as the coefficient of cost approaches zero. These causes difficulties when one is trying to estimate both the mean and standard error of the WTP measure.

There are a host of different methods that have been proposed by which CIs are built for WTP measures. These include the Delta method, Fieller method and Krinsky-Robb (or bootstrap) method, among other less commonly known methods. With so many methods existing, there has been no consensus in the literature over the best method by which to build CIs. In fact, many papers contradict each other, claiming one method to be better and providing reasons why the other methods fail. However no paper (other than Gatta et al. (2013), who still focus primarily on bootstrap methods) gives a comprehensive analysis of all the popular methods used. In addition, there is a distinct lack of attention to the Bayesian methods of model estimation, by which CIs can also be built. Hence this research aims to fill this gap by giving a comprehensive analysis of the most commonly used methods of building CIs in the transportation science and economics literature. In addition, this work proposes three further methods of building CIs: estimation of WTP measures in WTP-space using both frequentist and Bayesian methods, and Bayesian post-processing. Attention is given to the conditions under which each method works or does not work. Doing so would give a better understanding towards the reasons for such disparity within the literature, and help to consolidate all the different views. In addition, this research also aims to investigate methods by which CIs can be built when taste variations are assumed, i.e. when a random coefficient model is used. There is currently little discussion in the literature on the building of CIs for WTP under these model assumptions, even though it is extremely important. Not only is it logical to assume that individuals will have different taste preferences (and thus taste variations would exist in a model), the mixed logit model is also growing in popularity within the discrete choice model literature. Especially if used in applied work, CIs should be reported and thus there needs to be a clear method by which to build them.

In order to do this, simulations were run for both fixed and random coefficient models. Simulations are useful because they contain predetermined coefficient values, and thus the end goal is not only known, it is possible to calculate such useful statistics as coverage rates of the CIs built in order to see if one is meeting the stated confidence level. Under both model assumptions, CIs were built for varying sample sizes and cost coefficient values in order to study the effect of both sample size and weak identification.

The following main points can be concluded from the simulations. With a fixed coefficient model, all the methods work equally well under standard conditions (i.e. in the absence of weak identification). This concurs with the results of papers like Hole (2007). In the presence of weak identification, the Fieller method maintains a high coverage rate due to the formation of unbounded CIs. In contrast, all other methods perform poorly, with the Delta method performing especially badly. Since the Delta method builds symmetric CIs by construction, they are unable to capture the asymmetry of the WTP distribution as demonstrated by the Monte Carlo CIs obtained. The Bayesian post-processing and Krinsky-Robb methods give similar results, and hence under fairly standard conditions (and even allowing for fairly small cost coefficients if the sample size is large enough), the Bayesian post-processing is a viable method for building CIs.

In contrast, working in WTP-space builds incorrect CIs, with the Bayesian estimation faring worse than the frequentist method of estimation. The widths of the CIs built are much narrower than those of the other methods, indicating that the estimated standard error is smaller. However, coverage rates are extremely poor. Although not a definitive explanation, I suggest that working in WTP-space fares poorly due to the inherent assumptions involved. Under the frequentist method of estimation, we assume that the coefficients (here the WTP) being estimated are asymptotically normally distributed. However, this is an incorrect assumption, and thus would lead to inaccuracy in estimation of both the mean and standard error of the coefficient, by which the CIs are built. Under the Bayesian method of estimation, a closer look at the convergence of the estimators shows erratic convergence. Often times the Bayesian results do not converge to a single value, or seemingly converges to the wrong value, thus leading to inaccuracy in the credible intervals built.

With a random coefficient model, weak identification again causes inaccurate CIs to be built. However, the random coefficients result in CIs with much wider widths being built by the Bayesian post-processing and Krinsky-Robb methods, bringing about coverage rates that are too high. In contrast, the Delta method produces CIs which are either extremely wide or extremely narrow, depending on whether the mean or median of the WTP estimates and standard errors are used. Hence while it is true that using the median of WTP standard errors will produce narrower CIs (as concluded by Bliemer and Rose (2012)), it still does not result in accurate CIs due to the inaccurate median WTP estimates used.

To obtain a sense of how these methods can be used, several data sets were employed on which CIs were built for various WTP measures. The first quasi-simulation utilized vehicle choice data, where individuals had a choice between three different vehicles, whether standard gasoline, hybrid, or electric. The second used a data set where individuals chose between two travel modes that differed in travel time and cost. Finally, a third study used a data set where individuals ranked three different flight itineraries in order of their preference. The CIs built in each of these studies demonstrated the results of the simulation, and a comparison between them especially highlights the effect of weak identification on the building of CIs. In particular, the travel mode choice case study demonstrates the CIs that are built under weak identification, with the Fieller method even producing an unbounded CI. In contrast, the itinerary choice case study resulted in CIs which were all comparable, demonstrating that under good conditions, any method can be used to build the CI.

Based on the results of the simulations and case studies, I find that in the absence of weak identification and with a large sample size, any of the methods that build CIs in preference space work equally well, including the Bayesian post-processing method. In general, however, the Fieller method is the most versatile method as it also performs well in the presence of weak identification. However one must consider whether the unbounded CI is useful in applied work, or whether it should just be used as an indication of an insignificant marginal utility of cost.

The remainder of this research delves into building CIs for other functions of parameters; in particular, it looks at the example of building CIs for the probability that an airline passenger will cancel his ticket. These estimates are of value because the probability of cancellation is used in airline cancellation models, in which an airline estimates the number of tickets by which to overbook a flight so as to maximize their revenue. Building a CI would allow the airline to better understand not only the amount of trust they should put in the probability estimate, but also be able to incorporate these interval bounds into their revenue management models. In addition, I incorporate taste variation into the model, through which it is possible to build CIs of the individual cancellation probabilities. This allows for a greater understanding of the actions and decisions of individual airline customers.

Using a revealed preference data set containing information regarding passengers who purchased tickets and then subsequently used or cancelled them, I estimated both a MNL and a ML model. The taste variations incorporated into the ML model were statistically significant, indicating the importance of incorporating these taste variations into cancellation models. In addition, I demonstrated how CIs can be built not only for the individual's probability of cancelling his ticket on one particular day, but also every day from when he first purchased his ticket to the day of departure. The results of this work not only demonstrates how the method works, but also promotes the use of the Bayesian post-processing method due to the ease by which the individual CIs can be constructed.

This research takes a step forward in consolidating the many methods for building CIs for WTP measures, and also in promoting the use of Bayesian post-processing methods for building CIs, especially under heterogeneity conditions. It also identifies areas in which improvements can be made. In particular, I believe that more thought should be put into the use of WTP-space and its effectiveness. In addition, seeing the performance of the Fieller method in the fixed parameter model causes one to wonder if a similar method can be developed for the random parameter model.

Finally, more work can also be done to further improve on the CIs built for the cancellation probabilities of airline passengers. In particular, since the Bayesian method of estimation is being used, it is possible to more intuitively incorporate historical ticketing data. Taking advantage of the fact that certain individuals have purchased more than one ticket should further increase the accuracy of the Bayesian model estimates.

BIBLIOGRAPHY

American Psychological Association, 2001. Publication manual of the American Psychological Association (5th ed). American Psychological Association, Washington, D.C.

Armstrong, P., Garrido, R., Ortúzar, J.D., 2001. Confidence intervals to bound the value of time. Transportation Research Part E 37, 143-161.

Arnold, B., Brockett, P., 1992. On distributions whose component ratios are Cauchy. The American Statistician 46, 25-26.

Belobaba, P.P., 1989. Application of a probabilistic decision model to airline seat inventory control. Operations Research 37(2), 183-197.

Bernard, J.-T., Idoudi, N., Khalaf, L., Yélou, C., 2007. Finite sample inference methods for dynamic energy demand models. Journal of Applied Econometrics 22, 1211-1226.

Beyene, J., Moineddin, R., 2005. Methods for confidence interval estimation of a random parameter with application to location quotients. BMC Medical Research Methodology 5, 32.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit based unordered response choice models. Transportation Research Part B 45(7), 923-939.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. Transportation Research Part B 46(7), 817-833.

Bliemer, M.C.J., Rose, J.M., 2012. Confidence intervals of willingness-to-pay for random coefficient logit models. Presented at the 13th International Conference on Travel Behaviour Research, Toronto, Canada.

Bolduc, D., Khalaf, L., Yélou, C., 2010. Identification robust confidence set methods for inference on parameter ratios with application to discrete choice models. Journal of Econometrics 157, 317-327.

Borenstein, M., 1994. The case for confidence intervals in controlled clinical trials. Controlled Clinical Trials 15, 411-428.

Box, G.E.P., Tiao, G.C., 1992. Bayesian inference in statistical analysis. Wiley, New York, NY.

Boyd, E.A., 2004. Dramatic changes in distribution will require renewed focus on pricing and revenue management models. Journal of Revenue & Pricing Management 3(1), 100-103.

Boyd, E.A., Kallesen, R., 2004. The science of revenue management when passengers purchase the lowest available fare. Journal of Revenue & Pricing Management 3(2), 171-177.

Brumelle, S.L., McGill, J.I., 1993. Airline seat allocation with multiple nested fare classes. Operations Research 41(1), 127-137.

Buonaccorsi, J.P., Iyer, H.K., 1984. A comparison of confidence regions and designs in estimation of a ratio. Communications in Statistics - Simulations and Computations 13, 723-741.

Chatterjee, H., 2001. Forecasting for cancellations. Presented at AGIFORS Reservations and Yield Management Study Group, Bangkok, Thailand.

Chen, Y., Ordóñez, F., Palmer, K., 2006. Confidence intervals for OD demand estimation. Working Paper #2006-01.

Clark, S.D., McKimm, J., 2005. Estimating confidence intervals for transport mode share. Journal of Transportation and Statistics 8(2), 894-911.

Coladarci, T., 2003. Gallup goes to school: the importance of confidence intervals for evaluating ”adequate yearly progress” in small schools. Policy Brief. Rural School and Community Trust, Washington, D.C.

Cramer, J.S., 1986. Econometric applications of maximum likelihood methods. Cambridge University Press, Cambridge.

Daly, A., Hess, S., de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. Transportation Research Part B 46, 333-341.

Daziano, R.A., Achtnicht, M., 2014. Accounting for uncertainty in willingness to pay for environmental benefits. Energy Economics 44, 166-177.

De Bruijn, N.G., 1970. Asymptotic methods in analysis (3rd ed). North-Holland Publishing Co., Amsterdam, The Netherlands.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365-1389.

Dunleavy, H., Westermann, D., 2005. Future of airline revenue management. Journal of Revenue & Pricing Management 3(4), 380-383.

Fieller, E.C., 1932. The distribution of the index in a normal bivariate population. Biometrika 24, 428-440.

Fieller, E.C., 1944. A fundamental formula in the statistics of biological assay, and some applications. Quarterly Journal of Pharmacy and Pharmacology 17:117-123.

Fieller, E.C., 1954. Some problems in interval estimation. Journal of the Royal Statistical Society, Series B 16(2), 175-185.

Franz, V.H., 2003. Planning versus online control: Dynamic illusion effects in grasping? Spatial Vision, 16(3-4), 211-223.

Franz, V.H., 2007. Ratios: A short guide to confidence limits and proper use. Submitted for publication. Preprint available at <http://arxiv.org/abs/0710.2024>.

Gardner, M.J., Altman, D.G., 1986. Confidence intervals rather than p values: estimation rather than hypothesis testing. British Medical Journal 292, 746-750.

Garrow, L.A., 2010. Discrete choice modelling and air travel demand: theory and applications. Ashgate Publishing Limited, England.

Garrow, L.A., Jones, S.P., Parker, R.A., 2007. How much airline customers are willing to pay: An analysis of price sensitivity in online distribution channels. Journal of Revenue and Pricing Management 5(4), 271-290.

Gatta, V., Marcucci, E., Scaccia L., 2013. Willingness to pay/accept confidence interval estimation methods: a comparison. Working Paper.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317-1339.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In E. M. Keramidas, ed., Computer science and statistics: proceedings of the twenty-third symposium on the interface, Interface Foundation of North America, Inc., Fairfax, VA, 571-578.

Gleser, L.J., Hwang, J.T., 1987. The nonexistence of 100(1-alpha)-percent confidence sets of finite expected diameter in errors-in-variables and related models. Annals of Statistics 15, 1351-1362.

Graham, R.J., Garrow, L.A., Leonard, J.D., 2010. Business travelers' ticketing, refund, and exchange behavior. Journal of Air Transport Management 16(4), 196-201.

Greene, W.H., 2003. Econometric analysis (5th ed). Prentice-Hall, Englewood Cliffs.

Guerron-Quintana, P., Inoue, A., Kilian, L., 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. Quantitative Economics 4, 197-229.

Hajivassiliou, V., McFadden, D., 1998. The method of simulated scores for the estimation of LDV models. Econometrica 66, 863-896.

Hess, S., Train, K., Polak, J.W., 2006. On the use of a modified latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. Transportation Research Part B 40, 147-163.

Hinkley, D.V., 1969. On the ratio of two correlated normal variables. Biometrika 56, 635-639.

Hole, A.R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Economics 16, 827-840.

Hornick, S., 2004. Fare play. Airline Business 20, 72-75.

Hwang, J.T., 1995. Fieller's problems and resampling techniques. Statistica Sinica 5, 161-171.

Iliescu, D.C., Garrow, L.A., Parker, R.A., 2008. A hazard model of US airline passengers' refund and exchange behavior. Transportation Research Part B 42(3), 229-242.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Mathematical Proceedings of the Cambridge Philosophical Society 31(2), 203-222.

Theory of probability (3rd ed). Oxford University Press, Oxford, U.K.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90(430), 773-795.

Keane, M., 1990. Four essays in empirical macro and labor economics. Ph.D. Thesis, Brown University.

Keane, M., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95-116.

Koschat, M.A., 1987. A characterization of the Fieller solution. The Annals of Statistics 15(1), 462-468.

Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Review of Economics and Statistics 68, 715-719.

Krinsky, I., Robb, A.L., 1990. On approximating the statistical properties of elasticities: a correction. Review of Economics and Statistics 72, 189-190.

Lieberman, W.H., 2004. Revenue management trends and opportunities. Journal of Revenue & Pricing Management 3(1), 91-99.

Luce, D.R., 1959. Individual choice behavior. Wiley, New York, NY.

Marschak, J., 1960. Binary choice constraints on random utility indicators. In K. Arrow, ed., Stanford symposium on mathematical methods in the social sciences, Stanford University Press, Stanford, CA, 312-329.

Masson, M.E.J., Loftus, G.R., 2003. Using confidence intervals for graphically based data interpretation. Canadian Journal of Experimental Psychology 57(3), 203-220.

Mau, P., Eyzaguirre, J., Jaccard, M., Collins-Dodd, C., Tiedemann, K., 2005. The 'neighbor effect': simulating dynamics in consumer preferences for new vehicle technologies. Ecological Economics 68, 504-516.

McFadden, D., 2001. Economic choices. American Economic Review 91, 351-370.

McFadden, D., Train, K., 2000. Mixed MNL models of discrete response. Journal of Applied Econometrics 15, 447-470.

Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. Journal of the Royal Statistical Society, Series B 56(1), 3-48.

Neyman, J., 1937. Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences 236(767), 333-380.

Oliveira, V.M., 2003. Simulating revenue management in an airline market with demand segmentation and strategic interaction. Journal of Revenue & Pricing Management 1(4), 301-318.

Park, T., Loomis, J.B., Creel, M., 1991. Confidence intervals for evaluating benefits estimates for dichotomous choice contingent valuation studies. Land Economics 67(1), 64-73.

Rao, C.R., 1973. Linear statistical inference and its applications, 2nd ed. John Wiley & Sons, New York, NY.

Ratliff, R., 1998. Ideas on overbooking. Presented at AGIFORS Reservations and Yield Management Study Group, Melbourne, Australia.

Sándor, Z., Train, K., 2004. Quasi-random simulation of discrete choice models. Transportation Research Part B 38(4), 313-327.

Siani, C., de Peretti, C., 2003. Is Fieller's method applicable in all situations? GREQAM working paper, 03A33.

Sonnier, G., Ainslie, A., Otter, T., 2007. Heterogeneity distributions of willingness-to-pay in choice models. Quantitative Marketing and Economics 5(3), 313-331.

Talluri, K.T., van Ryzin, G.J., 2004. Revenue management under a general discrete choice model of consumer behavior. Management Science 50(1), 15-33.

Thompson, B., 2002. What future quantitative social science research could look like: confidence intervals for effect sizes. Educational Research 31(3), 25-32.

Thurstone, L.L., 1927. A law of comparative judgment. Psychological Review 34, 273-286.

Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments and marginal densities 81, 82-86.

Tilahun, N.Y., Levinson, D.M., Krizek, K.J., 2007. Trails, lanes, or traffic: valuing bicycle facilities with an adapted stated preference survey. Transportation Research Part A 41, 287-301.

Train, K.E., 2009. Discrete choice methods with simulation, 2nd ed. Cambridge University Press, New York, NY.

Train, K., Hudson, K., 2000. The impact of information in vehicle choice and the demand for electric vehicles in California. Project report, National Economic Research Associates.

Train, K., Sonnier, G., 2005. Mixed logit with bounded distributions of correlated parthworths. In R. Scarpa and A. Alberini, eds., Applications of Simulation Methods in Environmental and Resource Economics, Kluwer Academic Publishers, Boston, MA.

Tsao, C.A., Hwang, J.T.G., 1998. Improved confidence estimators for Fieller's confidence sets. Canadian Journal of Statistics - Revue Canadienne de Statistique 26, 299-310.

U.S. Department of Energy, 2014. All-Electric Vehicles. (online) available at <http://www.fueleconomy.gov/feg/evtech.shtml>. Last accessed: September 27, 2014.

Veldwijk, J., Lambooij, M.S., de Bekker-Grob, E.W., Smit, H.A., de Wit, G.A., 2014. The Effect of Including an Opt-Out Option in Discrete Choice Experiments. PLoS ONE 9(11), e111805.doi:10.1371/journal.pone.0111805 von Lyxburg, U., Franz, V.H., 2009. A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations and bootstrap. Statistica Sinica 19, 1095-1117.

Westerhof, A., 1997. CO2 in the air. Presented at AGIFORS Reservations and Yield Management Study Group, Melbourne, Australia.

Wilkinson, L., 1999. Statistical methods in psychology journals: guidelines and explanations. American Psychologist 54(8), 594-604.