Estimation and Interpretation of Discrete Choice Models: Methods for Evaluating Multivariate Probability Distributions

by

Jiahui Ying

(Under the Direction of J. Scott Shonkwiler)

Abstract

The discrete choice model is a powerful method that can quantify consumer preferences for both non-market and market goods. In either stated preference surveys with choice experiments or revealed preference data like the retail-level transaction records, discrete outcomes are in the form of multivariate correlated variables. From the view of multivariate probability distributions, this dissertation investigates several methods that improve the estimation and interpretation of discrete choice models. First, for the mixed logit model, Gauss-Hermite integration is found to be a powerful alternative to maximum simulated likelihood when the number of random parameters is moderate($\leq$6). It avoids simulation bias and simulation noise and only incurs controllable approximation error. Further, the Bayesian approach and the block delete jackknife are tested to outperform the Delta method in describing the distribution of mean Willingness to Pay in the mixed logit model. The virtues of a normally distributed cost coefficient is also validated with both empirical and synthetic data set. Finally, in the sense of aggregating discrete choices outcomes in a real market, a utility-consistent count system is developed as a tool to analyze consumer brand choices.

Index words: Mixed logit model, Gauss-Hermite quadrature, mean willingness to pay, Bayesian estimation, block delete jackknife, multivariate Poisson-log Normal, incomplete demand system, over-dispersion, scanner data

Estimation and Interpretation of Discrete Choice Models: Methods for Evaluating Multivariate Probability Distributions

by

Jiahui Ying

B.A., Zhongnan University of Economics and Law, 2013

M.A., Renmin University of China, 2015

M.S., The University of Georgia, 2018

A Doctor of Philosophy Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2019

Estimation and Interpretation of Discrete Choice Models: Methods for Evaluating Multivariate Probability Distributions

by

Jiahui Ying

Approved:

Major Professor: J. Scott Shonkwiler

Committee: Craig E. Landry
John C. Bergstrom

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2019

Dedication

To my dear parents, major professor and committee members.

Acknowledgments

The past four years at the Department of Agricultural and Applied Economics at the UGA has been an academic achievement that I am proud of. For this, I am sincerely grateful to my major professor, Dr. J. Scott Shonkwiler for his years of excellent help, powerful support, and patiently guiding my research. From the first steps of searching a research topic to model building, data analysis, and result estimation, I have learned so much from him that is not limited to the econometric techniques but more about the keen interest in knowledge. All the insights, innovative ideas, and curiosity shared by Dr. Shonkwiler has been inspirational and motivates me in my future research in applied economics.

At the same time, I am truly thankful to my committee members, Dr. Craig Landry and Dr. John Bergstrom, for their constructive suggestions and continued support. Also, I appreciate Dr. Zhen Chen for his kind share of the IRI marketing data set. My deepest appreciation also goes to Dr. Jeffrey Dorfman for his generous help that largely extended my understanding in agricultural futures market.

Specifically, I would like to thank my parents for their unconditional love and support throughout my studies. Although they are in China, their encouragement, understanding, and advice has always been so close, warm, and powerful.

Also, I would thank my sweet friends in Athens who made my life so enjoyable, especially the ones in the same class or office: Yu Chen, Yunhan Li, Wenying Li, Shun Chen, Akash Issar, Mengyao Li, Lei Xian, my close friends Chen Chen, Yan Wang, Xuedong Wu, and the warmest Shonkwiler family: Vanessa, Alyssa and Trevor.

Finally, a special thank to my lovely boyfriend, Charng-Jiun Yu, for all the strength he shares with me in every moment.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

The discrete choice model is a powerful method to measure consumer preferences for both non-market and market goods. In non-market valuation, discrete choice experiments can help to reveal consumers' willingness to pay for interested attributes in either public goods like the natural resource, or a new product that is not available in the market yet. For market good, analysis of revealed preference data such as the retail-level consumption transactions can provide insights to market structures and customer segments. In both cases, multivariate probability distributions are crucial in the calculation of choice probabilities especially when heterogeneities in consumer preferences are considered. This dissertation investigates several methods that improve the estimation and interpretation of two important discrete choice models – the mixed logit model, and the multivariate count data model which can be derived from repeated discrete choices over time (Hellerstein and Mendelsohn, 1993).

As one of the most flexible methods in discrete choice analysis, mixed logit models have been widely applied in various disciplines including non-market valuation, transportation analysis, and health economics, etc. According to Train (2009), the three main advantages of the mixed logit model are 1) allowing for random taste variations as it can handle respondents' tastes that vary with unobserved variables or purely randomly; 2) enable unrestricted substitution patterns by avoiding the restriction of proportionate substitutes; 3) allowing dynamic correlations in unobserved factors over time so that panel data could be analyzed as the lagged response to changes in attributes can be accommodated.

Given the high flexibility allowed for consumer preferences, the estimation of mixed logit models becomes more complex as no closed-form choice probability is available.

Although maximum simulated likelihood estimation is entrenched in most statistical software, researchers have discovered that the number of simulation draws may be too small to ensure estimation accuracy (Czajkowski et al., 2017). Instead of quasi-Monte Carlo simulation, the first chapter explores the application of quadrature methods in approximating choice probabilities in mixed logit models. We extend the previous work by Breffle et al. (2005) which considered Gauss-Hermite quadrature in the estimation of a probit choice model with two uncorrelated random parameters. We generalize this approach to the mixed logit model with repeated choices and with correlated random coefficients. By specifying the exact likelihood function, we avoid simulation error but incur approximation error which depends on the degree of the Hermite orthogonal polynomial used. We show that by appropriately trimming the points of evaluation and rescaling the weights, the number of evaluation points can be substantially reduced without introducing significant approximation error. As a counterpart, this chapter also discusses the possibility of applying another quadrature method, the sparse grid, in mixed logit models as an alternative solution to high-dimensional integration. Our empirical analysis with two available survey data sets suggests Gauss-Hermite quadrature as the preferable method in mixed logit models, especially when the number of random coefficients is relatively small ($\leq$6) and the sample size is small.

For the interpretation of mixed logit models, representation of Willingness to Pay (WTP) has long been debated given the non-zero probability of a zero-denominator when unbounded distributions are applied to the cost coefficient (Hensher and Greene, 2003; Train and Sonnier, 2005; Hess et al., 2005; Daly et al., 2012). However, economic theory indicates the marginal utility of income for normal goods is always positive if the consumption is non-zero. Under this assumption, a well-defined approximation of the moments in WTP – the ratio of two random coefficients – is available according to the theory built in Marsaglia et al. (2006). With this justification, the second chapter explores the distribution of the most important statistic - the mean - of WTP in the mixed logit model with three methods: the classical Delta method, the Bayesian approach with individual-level willingness to pay, and a resampling method using the block delete jackknife. The empirical analysis shows the drawback of the Delta method, reveals the skewness introduced by a log-normally distributed cost coefficient, and validates the virtues of the Bayesian approach and the block delete jackknife. The potential of applying a normally distributed cost coefficient is further validated by a synthetic data set with 2000 respondents.

In terms of consumer brand choice, the third chapter provides a new approach which recognizes that quantities purchased are discrete and over-dispersed and demands may be correlated. As a consequence, we specify a multivariate Poisson-log normal distribution to fit the non-negative count outcomes in consumer demand. As an extension of typical discrete choice models like the mixed logit model that focus on consumers' binomial or multinomial decisions in a single choice, count data models are developed in the sense of aggregating the discrete choice outcomes with the theoretical foundation derived from a repeated application of the discrete choices (Hellerstein and Mendelsohn, 1993). However, most existing models in the multivariate count case are lacking of a consistent underlying framework that can be derived from the well-developed consumer utility theories (Bhat et al., 2015).

To represent consumer preferences in a utility-consistent way, we consider the incomplete demand system for its virtues in fully integrating the extensive commodity selection and intensive derived demand choices within a coherent and consistent model of consumer behavior, as noted by von Haefen (2002). Specifically, a demand specification with log-linear from – a necessary requirement given the exponential link in the Poisson-log normal count data model – is adopted. With the assumption that the prices of all other goods outside the system are quasi-fixed, unconditional price effects and income effects can be computed from the properly specified incomplete demand system. Further, this chapter is unique in applying the count data demand system with the real transaction data at the retail level. Using the panel set of scanner data provided by the IRI markeing data set, we analyze 1927 houshold choices for four major brands in the facial tissue market in Eau Claire, Wisconsin in 2011 and provide insights on market structure and consumer segments.

References

Bhat, C. R., Paleti, R., and Castro, M. (2015). A new utility-consistent econometric approach to multivariate count data modeling. *Journal of Applied Econometrics*, 30(5):806–825.

Breffle, W. S., Morey, E., and Waldman, D. (2005). Gaussian quadrature versus simulation for the estimation of random parameters. In *Applications of Simulation Methods in Environmental and Resource Economics*, pages 341–353. Springer.

Czajkowski, M., Budziński, W., et al. (2017). Simulation error in maximum likelihood estimation of discrete choice models. Technical report.

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1):19–31.

Hellerstein, D. and Mendelsohn, R. (1993). A theoretical foundation for count data models. *American Journal of Agricultural Economics*, 75(3):604–611.

Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30(2):133–176.

Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, 39(2-3):221–236.

Marsaglia, G. et al. (2006). Ratios of normal variables. *Journal of Statistical Software*, 16(4):1–10.

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In *Applications of simulation methods in environmental and resource economics*, pages 117–134. Springer.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

von Haefen, R. H. (2002). A complete characterization of the linear, log-linear, and semilog incomplete demand system models. *Journal of Agricultural and Resource Economics*, 27(2):281–319.

Chapter 2

Simulation Noise versus Approximation Error in Mixed Logit Estimation: Maximum Simulated Likelihood or Gauss-Hermite Integration

* Ying J. and Shonkwiler J. S.. Submitted to Environmental and Resource Economics, 3/11/2019.

Abstract

As one of the most flexible methods in discrete choice analysis, mixed logit models have been widely applied in various disciplines. Although maximum simulated likelihood is entrenched in most statistical software, researchers have discovered that the number of simulation draws may be too small to ensure estimation accuracy. Instead of focusing on quasi-Monte Carlo simulation, this paper explores the application of quadrature methods in approximating choice probabilities in mixed logit models. We extend Gauss-Hermite quadrature to the mixed logit model with correlated coefficients, and improve estimation speed by trimming the low-impact nodes in multi-dimensional integration. As a counterpart, sparse grid integration is also evaluated as a candidate. Our empirical analysis with two available survey data sets validates the merit of Gauss-Hermite quadrature in estimation accuracy and efficiency. Compared to the other two methods, we suggest the Gauss-Hermite quadrature as the preferable method in mixed logit models, especially when the number of random coefficients is relatively small (≤ 6) and the sample size is small.

Keywords: Mixed Logit Model, Numerical Integration, Gauss-Hermite Quadrature, Sparse Grid, quasi-Monte Carlo Simulation

2.1 Introduction

The mixed logit model is a preferred estimator of random utility models due to its flexibility and ability to explain a wide variety of behavioral choices. Since discrete choice models provide constructs for the calculation of willingness to pay (WTP) for many economically important attributes in both market and non-market settings, mixed logit is ideally suited to capture random heterogeneity in in WTP.

The method of Maximum Simulated Likelihood (MSL) has become entrenched as the preferred estimator of mixed logit models. This method has been implemented in several statistical packages and has the attractive property that coefficients may be specified under any distribution that can be randomly generated. Commonly, random coefficients are assumed to be distributed normally or log-normally and it is under these distributions that we consider almost exact maximum likelihood estimation of the mixed logit model.

As is well-known, MSL estimators incur simulation error that depends on the random number generator used and the number of draws. Further, increasing the number of draws does not guarantee that the resulting estimator is closer to the true model. In fact, a recent work (Czajkowski et al., 2017) suggests that tens of thousands of draws may be necessary in order to achieve satisfactory MSL results. On the other hand, exact maximum likelihood estimation suffers from the "curse of dimensionality" due to having to integrate over a multivariate distribution.

This research extends the previous work by Breffle et al. (2005) which considered Gauss-Hermite quadrature in the estimation of a probit choice model with two uncorrelated random parameters. We generalize this approach to the mixed logit model with repeated choices and with correlated random coefficients. By specifying the exact likelihood function, we avoid simulation error but incur approximation error which depends on the degree of the Hermite orthogonal polynomial used. In an application with two lognormally and two normally distributed random coefficients, high order Gauss-Hermite integration using 32 nodes would require 32^4 (over one million) evaluation points.

We show that by appropriately trimming the points of evaluation and rescaling the weights, the number of evaluation points can be substantially reduced without introducing significant approximation error. At the same time, we discuss the possibility of applying another quadrature method, the sparse grid, in mixed logit models as an alternative solution to high-dimensional integration. In empirical applications, trimmed Gauss-Hermite quadrature is demonstrated and compared to MSL estimation using various numbers of scrambled Halton and Sobol pseudo-random draws. We conclude that exact maximum likelihood estimation is a powerful alternative to MSL that practitioners should consider. Also, Gauss-Hermite quadrature is a better choice than sparse grid integration in mixed logit models as it avoids the problem of negative integrated choice probabilities.

2.2 Literature Review

2.2.1 The Mixed Logit Model

The mixed logit model (Revelt and Train, 1998) is a typical discrete choice model under the assumption of utility-maximizing behavior by the decision-maker. Although the traditional multinomial logit model is still quite prevalent in both industrial and academic analyses, the assumption of homogeneous preferences for all respondents is unnecessarily stringent. Also, the independence of irrelevant alternatives (IIA) assumption of logit can be unrealistic in many settings, as it does not allow for different degrees of substitution or complementarity among choices and requires identical cross-price elasticities for a given product, which is obviously implausible (Hausman, 1975).

The mixed logit model is widely preferred given its flexibility in approximating any random utility model (McFadden and Train, 2000). According to Train(2009), the three main advantages of the mixed logit model are 1) allowing for random taste variations as it can handle respondents' tastes that vary with unobserved variables or purely randomly; 2) enable unrestricted substitution patterns by avoiding the restriction of proportionate substitutes; 3) allowing dynamic correlations in unobserved factors over time so that panel data could be analyzed as the lagged response to changes in attributes can be accommodated.

Over the past 15 to 20 years, many studies have applied the mixed logit model in a large range of fields including non-market valuation (e.g., Hensher et al., 2005; Scarpa et al., 2007; Duchesne et al., 2010, etc.), transportation analysis (e.g., Boyd and Mellman, 1980; Cardell and Dunbar, 1980; González-Savignat, 2004; Hess et al., 2004; Shen, 2009; Srikukenthiran et al., 2014; Lee et al., 2016, etc.), and health economics (e.g., Hall et al., 2006; King et al., 2007; Paterson et al., 2008; Regier et al., 2009; Hole, 2008; Hole and Kolstad, 2012, etc.), etc.

2.2.2 Estimation Methods of Mixed logit models

Given that the expression of choice probabilities does not have a closed form in mixed logit models, simulation is widely adopted in numerical evaluations and thus leaves the sampling method a crucial, debatable issue. Before Train (2000), pseudo-random draws (Monte Carlo) were used for almost all simulations in mixed logit models. The efficiencies of pseudo-random draws are always questionable as it can hardly ensure a representative sample from the mixing distribution, especially when the number of draws is small.

An important alternative sampling method is the quasi-Monte Carlo (QMC) method like the Halton sequence and the Sobol sequence. The Halton sequences follow the idea of dividing the unit interval evenly step by step according to the Halton number set, then transfer the Halton sequence into the quasi-random draws by the one-to-one projection with the inverse cumulative function of the mixing distribution. Bhat (2001) first tested Halton sequences for mixed logit models and found it substantially outperforms the pseudo-random draws. Train (2000) further explained the reason as 1) the Halton draws can achieve a fairly even coverage over the domain of the mixing distribution, so that the simulated probabilities vary less over observations; 2) draws in the Halton sequences tend to fill in the spaces that were left empty by the previous observations, so that the simulated probabilities become negatively correlated over observations, and thus reduce the variance in the log-likelihood function.

Following this trend, more researchers investigate the performance of QMC methods with different sampling sequences. For example, the shuffled Halton, scrambled Halton (Daly et al., 2003; Hess and Polak, 2003; Wang and Kockelman, 2008), and the randomized Halton draws (Sándor and Train, 2004; Munger et al., 2012) are further developed and compared with each other. Other sequences like the Sobol sequences, randomized Sobol sequences, the modified Latin hypercube sampling, and the randomized lattice rules are sometimes found to have better performance than Halton sequences in certain circumstances like higher dimensional cases (Garrido, 2003; Hess et al., 2006; Munger et al., 2012).

However, an important problem with the existing QMC sampling methods is that the simulation noise does not necessarily show a monotonic decreasing trend as the number of draws increases. Thus, determining the number of draws to achieve stable status is time-consuming and somewhat subjective. For example, Train (2000) noticed the variances of estimators with 125 Halton draws are unexpectedly larger than that from the 100 Halton draws case. Czajkowski et al. (2017) further found the simulation bias is not negligible, and the number of draws used by many empirical applications is too low for reliable inferences. They recommend 15,000 Sobol draws to achieve the estimation accuracy within 1% of true values.

Crucial information ignored by most existing simulation methods is the coefficients' distribution – only the well-defined QMC algorithms are used to generate random draws in MSL. While for Gauss-Hermite quadrature, evaluation points are carefully selected and weighted based on the coefficients' distribution. Such a "directional" sampling method enables a more comprehensive coverage of the parameters' distribution. A common concern for quadrature techniques is that they may be computationally too burdensome in practice (Albright et al., 1977; Hausman and Wise, 1978). However, Butler and Moffitt (1982) illustrated that such concern is only true for standard quadrature techniques such as trapezoidal integration or its improved variants. Butler and Moffitt (1982) also showed that Gaussian quadrature is extremely efficient with computational feasibility in calculating single-bound or double-bound Tobit models. In 2005, Breffle et al. introduced Gaussian quadrature into the estimation of uncorrelated normally-distributed random parameters. From an empirical analysis with stated preference data on fishing in Green Bay, they showed that quadrature is faster than simulations with pseudo-random draws in obtaining a high level of accuracy. To further extend this method, we will demonstrate how it can be applied in mixed logit models with correlated random parameters and test its performance in comparison to some QMC methods.

Another counterpart to Gaussian quadrature in likelihood approximation is quadrature on sparse grids that originates from Smolyak (1963) and further developed by Krueger and Kubler (2004), Winschel and Krätzig (2004), and Heiss and Winschel (2008). Unlike the product rule in multivariate quadrature where the full grid of points is evaluated, sparse grids integration only uses a subset of the evaluation nodes from the product and rescales the weights appropriately. In other words, it aims to be exact in the class of complete polynomials instead of tensor products of univariate polynomials (Heiss and Winschel, 2008). Such a feature helps sparse grid integration avoid the exponential growth of evaluation nodes. However, this method also has a potential drawback in generating negative weights, which can lead to negative probabilities in likelihood functions. In our analysis, we will also empirically examine the performance of sparse grid integration in mixed logit models.

2.3 Methodology

Like many other econometric models, the likelihood function in the mixed logit model does not have a closed form. Let the utility function of individual i for alternative j be

$$U_{ij} = \beta_i' x_{ij} + \epsilon_{ij}$$

Mixed logit extends the assumption that the coefficient β_i is a fixed true value (for every respondent) to a distribution $f(\beta|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters that determines the distribution of β and needs to be estimated. For example, in terms of a normally distributed coefficient $\beta \sim N(\mu, \sigma)$, parameters μ and σ instead of β are the object of our interest. Given the error term $\epsilon_{ij} \sim EV(0,1)$, the likelihood or the choice probabilities could be expressed as an integration of the logit-form integrand:

$$P_{ij} = \int L_{ij}(\beta) f(\beta|\theta) d\beta = \int \frac{e^{\beta' x_{ij}}}{\sum_{j=1}^{J} e^{\beta' x_{ij}}} f(\beta|\theta) d\beta$$

Clearly, the likelihood cannot be derived analytically so that numerical approximation methods, such as the quadrature methods, are needed. The idea of quadrature methods is to approximate an integral as a weighted sum of the integrand evaluated at a series of nodes. Following Heiss and Winschel (2008) and Skrainka and Judd (2011), for the integral with one dimension,

$$I[g] := \int_{\Omega} w(x) f(x) dx, \quad \Omega \subset \mathbb{R}, \quad w(x) \geq 0 \ \ \forall x \in \Omega$$

the approximation is called a *quadrature formula* as

$$Q[g] := \sum_{k=1}^{R} w_k f(y_k), \quad y_k \in \Omega$$

where y_k is the node and w_k is the corresponding weight, R is the total number of the nodes. For multivariate integration, let $\mathbf{x} = [x_1, x_2, \ldots, x_D]$, then

$$I_D[g] := \int_{\Omega_1} \int_{\Omega_2} \cdots \int_{\Omega_D} g(\mathbf{x}) \tilde{w}(\mathbf{x}) dx_D \ldots d_2 d_1, \quad \Omega \subset \mathbb{R}^d, \quad \tilde{w}(x) \geq 0 \ \ \forall x \in \Omega$$

where $\tilde{w}(x)$ is the joint probability density function of $\mathbf{x}$, and variables in $\mathbf{x}$ are assumed to be independently and identically distributed so that $\tilde{w}(x) = \Pi_{d=1}^{D} w(x_d)$. The approximation $Q[g]$, then, is called the *cubature formula* and it needs to be carefully investigated in order to identify the nodes. During the past 50 years, researchers have explored various quadrature rules aiming to deliver high accuracy at low computational cost (less nodes).

Quadrature approaches, like Gauss-Hermite and the sparse grid integration can deal with moderate dimensional integration with high accuracy and efficiency in general. Following we will introduce their application to the mixed logit model.

2.3.1 Simulation Method

Monte Carlo method is one of the most popular choices for numerical integration with random numbers. Through an appropriate distribution, the nodes are randomly selected and a weight is equally assigned to each node ($w_k = 1/R$), so that the simulated integral is

$$Q[g] = \frac{1}{R}\sum_{r=1}^{R} g(x_r)$$

By the Law of Large Numbers, Monte Carlo method displays $\frac{1}{\sqrt{R}}$ convergence so that the accuracy of the estimator would only be increased at the rate of $\sqrt{R}$ as R increases. Train (2009) further discussed the property of the Maximum Simulated Likelihood (MSL) estimator in the mixed logit model. Under the assumption that the number of draws R rises faster than $\sqrt{n}$, where n is the number of observations, simulation bias disappears asymptotically. For high-dimensional (m) integration, the Monte Carlo method has the advantage of avoiding the exponential increase of evaluating nodes by randomly selecting points in the m-dimensional space.

How to select the random draws to optimize the estimation efficiency of the Monte Carlo Method is of interest and debate. As discussed before, instead of using the pseudo-random draws, quasi-Monte Carlo (QMC) methods with low-discrepancy sequences have a faster rate of convergence close to $O(\frac{1}{R})$ as the nodes are more evenly distributed over the integration space. With these virtues, the QMC method is widely applied in today's mixed logit estimation. In our application, two most commonly used QMC sequence: the scrambled Sobol and Halton quasi-random sequences (in MATLAB) are considered to test the performance of the simulation method.

2.3.2 Gauss-Hermite Integration

Breffle et al. (2005) provided a nice introduction to the use of Hermite orthogonal polynomials to integrate a function of a standard normal random variable using Gauss-Hermite quadrature. Thus our treatment will be brief.

The integral

$$\int_{-\infty}^{\infty} e^{-\varepsilon^2} f(\varepsilon)\, d\varepsilon \approx \sum_{h=1}^{d} w_h f(\varepsilon_h)$$

here the approximation is defined by a Hermite orthogonal polynomial of degree d, $H_d(\varepsilon)$, with associated weights w_h $(h = 1, 2, , d)$. For a standard normal random variable, a change of variable results in

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-\varepsilon^2/2} f(\varepsilon)\, d\varepsilon \approx \sum_{h=1}^{d} w_h^* f(\varepsilon_h^*)$$

where $\varepsilon_h^* = \sqrt{2}\varepsilon_h$ and $w_h^* = w_h/\sqrt{\pi}$. Note that $\Sigma w_h^* = 1$.

A. Mixed Logit with a Single Random Parameter

Let the i^{th} person's utility from alternative j $(j = 1, 2, \ldots, J)$ be written

$$U_{ij} = f(\beta + \sigma\varepsilon)x_{ij} + \theta_{ij} + \varepsilon_{ij}$$

where β and σ are parameters, ε is a standard normal random variable, x_{ij} is an attribute that can vary by individuals and alternatives, θ is composed of other attributes weighted by constant parameters, and ε_{ij} is an independent and identically distributed extreme value random variable. The probability that person i chooses alternative j then becomes

$$P_{ij} = \int_{-\infty}^{\infty} \frac{\exp(\ f(\beta + \sigma\varepsilon)x_{ij}\ +\ \theta_{ij})}{\sum_{k=1}^{J} \exp(\ f(\beta + \sigma\varepsilon)x_{ik}\ +\ \theta_{ik})} \frac{\exp\left(-\varepsilon^2/2\right)}{\sqrt{2\pi}} d\varepsilon$$

When using a Hermite polynomial of degree d, this probability can by approximated by

$$P_{ij} \approx \sum_{h=1}^{d} w_h^* \frac{\exp(\ f(\beta + \sigma\varepsilon_h^*) x_{ij}\ +\ \theta_{ij})}{\sum_{k=1}^{J} \exp(\ f(\beta + \sigma\varepsilon_h^*) x_{ik}\ +\ \theta_{ik})}$$

In a panel context, assume that individual i faces $t = 1, 2, \ldots, T$ choices and that random parameters are constant over all choices faced by the individual. The probability of the observed sequence of T choices by individual i may be expressed as

$$P_{i(T)} = \int \left[\prod_{t=1}^{T} \frac{\exp(\ f(\beta + \sigma\varepsilon) x_{ij(t)}\ +\ \theta_{ij(t)})}{\sum_{k=1}^{J} \exp(\ f(\beta + \sigma\varepsilon) x_{ik(t)}\ +\ \theta_{ik(t)})} \right] \frac{\exp\left(-\varepsilon^2/2\right)}{\sqrt{2\pi}} d\varepsilon$$

Here $j(t)$ indicates the attributes of the alternative that was chosen at the t^{th} choice occasion. The Gauss-Hermite (G-H) quadrature approximation is

$$P_{i(T)} \approx \sum_{h=1}^{d} w_h^* \left[\prod_{t=1}^{T} \frac{exp\big(f\left(\beta + \sigma\varepsilon_h^*\right) x_{ij(t)} + \theta_{ij(t)}\big)}{\sum_{k=1}^{J} exp\big(f\left(\beta + \sigma\varepsilon_h^*\right) x_{ik(t)} + \theta_{ik(t)}\big)} \right].$$

B. Mixed Logit with Mltivariate Normal Correalted Parameters

With m possibly correlated parameters, define

$$V_{ij} = f_1(\beta_1 + \sigma_1\varepsilon_1)x_{1ij} + \cdots + f_m(\beta_m + \sigma_m\varepsilon_m)x_{mij} + \theta_{ij},$$

then

$$P_{ij} = \int_{R^m} \frac{\exp(V_{ij})}{\sum_{k=1}^{J}\exp(V_{ik})} \frac{exp(-1/2E\Sigma^{-1}E')}{(2\pi)^{m/2}|\Sigma|^{1/2}} d_{\varepsilon_1} \ldots d_{\varepsilon_m}$$

where $E = [\varepsilon_1 \ldots \varepsilon_m]$ and Σ is the variance-covariance matrix of E.

As before, for a d-degree Hermite orthogonal polynomial, let $\varepsilon_h^* = \sqrt{2}\varepsilon_h$. Define the set H as the Cartesian product $H = \varepsilon_h^* \times \varepsilon_h^* \ldots \ \times \varepsilon_h^*$. Thus H is of dimension d^m by m. In a similar manner define the set $W = w_h^* \times\ w_h^* \ldots\ \times\ w_h^*$. Next, define w_m as the product of the columns of W such that it is now a d^m by 1 vector whose sum is one. Finally define $E^* = HS$ where S is the (upper triangular) Cholesky decomposition of Σ.

We now write

$$V_{ij}* = f_1(\beta_1 + E_1^*)x_{1ij} + \cdots + f_m(\beta_m + E_m^*)x_{mij} + \theta_{ij}$$

where E_p^* denotes the p^{th} column of E^*. This permits expressing the approximation as

$$P_{ij} \approx w_m\prime \frac{\exp(V_{ij}^*)}{\sum_{k=1}^J \exp(V_{ik}^*)}$$

It is straightforward to extend the approach to panel data:

$$P_{i(T)} \approx w_m' \left[\prod_{t=1}^T \frac{\exp(V_{ij(t)}^*)}{\sum_{k=1}^J \exp(V_{ikt}^*)} \right].$$

C. Trimmed Gauss-Hermite Quadrature

To motivate the rationale for trimming the nodes consider the 20^{th} degree (scaled) Hermite orthogonal polynomials in Table 2.1. Note that the weights, w^*, decline rapidly for larger absolute values of the nodes, ε^*. The fact that they are not zero is a consequence of the possibility that the function of ε^* may be highly convex. However, in the context of random parameters logit, we are dealing with probabilities so that

$$f\,(\varepsilon^*) = \frac{\exp(\ f(\beta + \sigma\varepsilon^*)x_{ij}\ +\ \theta_{ij})}{\sum_{k=1}^J \exp(\ f(\beta + \sigma\varepsilon^*)x_{ik}\ +\ \theta_{ik})} < 1\ \forall\ i, j.$$

This feature of the mixed logit suggests that large absolute values of the nodes contribute little to the log likelihood.

The third column in Table 2.1 shows that if $m = 4$ (four random parameters), the largest weight in the implied w_4^* is 0.046 and the smallest weight 2.47E-51. Since the sum of the 160,000 elements of w_4^* is one, then we know the mean weight is $1/(20^4) = .00000625$. Thus we can consider trimming those nodes which are associated with weights less than some fraction of the mean weight of w_m^*. For the case of w_4^* and $d = 20$, if nodes with weights of one-tenth of the mean, one-hundredth of the mean, and one-thousandth of the mean were trimmed, then 6,832, 10,416, and 14,880 points respectively would need to be evaluated. Clearly this is a substantial reduction from 160,000 points.

Note that once the degree of the Hermite polynomial is chosen and the number of random parameters decided, then the same set of points and weights can be duplicated by any other analyst given the fraction trimmed. This ability to easily replicate estimation is not afforded to MSL since the exact method to generate the random draws may not be available to the analyst. Additionally, under Gauss-Hermite quadrature only one set of nodes and weights is needed for each respondent; whereas typically a different set of random draws is required for each respondent under MSL.

Of course, once the weights are trimmed, they no longer sum to one. We suggest rescaling the weights after trimming to assure they sum to one. Our investigations suggest that for a given degree d, that there is little to be gained by selecting a trimming fraction smaller than one-hundredth of the mean. Precision is uniformly gained by increasing the degree of the orthogonal polynomial given a level of trimming. It is an empirical issue as to the selection of the degree and the fraction trimmed as this will depend on the specific mixed logit model analyzed. We advocate increasing the degree of the orthogonal polynomial until estimated parameters stabilize.

2.3.3 Sparse Grids Integration

In addition to Gauss-Hermite integration, we also consider applying sparse grid in mixed logit models for its good performance in integrating higher dimensional polynomials. The key idea of sparse grid integration is to carefully choose and reweight the evaluation nodes to avoid the "curse of dimensionality" in multivariate quadrature. Instead of approximating a multi-dimensional integral by combining the univariate quadrature rules in a tensor product approach, sparse grid integration carefully identifies the essential evaluations nodes so that the calculation load is greatly reduced. To optimally select the nodes, Smolyak (1963) provided a general rule on how to extend univariate operators to multiple dimensions.

Wasilkowski and Wozniakowski (1995) gave a more explicit expression of the Smolyak rule and Heiss and Winschel (2008) further extended the approach to approximating the likelihood with numerical integration in multiple dimensions. In some recent research, the performance of sparse grid integration is further evaluated in both the BLP (Berry et al., 1995) and mixed logit model (eg., Skrainka and Judd, 2011). More details on the algorithm of sparse grid integration can be found in the Appendix A.

However, one potential problem with sparse grid integration is the possibility of negative weights for some nodes. This means theoretically it is possible to have a negative approximated integral although the integrand is positive everywhere (Heiss and Winschel, 2008). Although Heiss and Winschel mentioned that the effects should be alleviated by increasing the accuracy of the approximation formula and they didn't encounter such problem in their simulation analysis, we address this question in the analysis of mixed logit models with real survey data. In an application, we adopt the sparse grid node and weights generated by the code provided by Heiss and Winschel (2008) in MATLAB.

2.4 Data

For ease of replication, we choose two available data sets to compare the performance of the three numerical integration methods in the mixed logit model. The first data set is provided by Train and Sonnier (2005) on consumers' choice among different types of vehicles (gas, electric and hybrid) under different combinations of price, operating cost, range, and performance. This experiment is a part of the survey that targeted vehicle owners in California. California Air Resources Board has been interested in the promotion of vehicles with fewer pollutant emissions.

The survey contacted respondents randomly throughout the state by telephone and those who planned to purchase a car within the next three years were invited to participate in this survey. 100 participants are included in total. Each participant received a questionnaire with 15 questions of vehicle choices. To make the data set a balanced panel, we adopt the first 10 questions for each respondent. Thus, there are 100×10 choice situations with 3000 alternatives. For each question, three different vehicle types are listed, and the respondents are asked to choose the one vehicle that they prefer the most given the attribute levels. A sample of the data is shown in Table 2.2. Attributes assigned to each alternative include:

(a) Engine type: gasoline, electric, or hybrid

(b) Purchase price (in $\$10,000$)

(c) Operating cost (in dollars per month)

(d) Performance:

   i. High performance: Top speed of 120 mph, and 8 seconds to reach 60 mph

   ii. Middle performance: Top speed of 100 mph, and 12 seconds to reach 60 mph

   iii. Low performance: Top speed of 80 mph, and 16 seconds to reach 60 mph

(e) Range: miles between refueling/recharging

Considering the high correlation coefficient between the purchase price and the operation cost (p-value < 0.001), we drop the operating cost variable in the following analysis. We also combine the middle performance and the low performance as one group to simplify the analysis. The choice experiments were designed to provide wide variation in each attribute and as little covariance among attributes as possible while maintaining plausibility. A summary of the attributes is provided in Table 3.2.

The second data set was introduced by Huber and Train (2001) and further analyzed by Elshiewy et al. (2017). It records 361 respondents' choices on electricity suppliers with the attributes including fixed price, length of contract, type of company, time-of-use rates, and seasonal rates. The survey was conducted by Electric Power Research Instituted. Each respondent was given 12 questions regarding four types of suppliers with different combinations of the attributes listed below:

(a) Fixed price (in cents per kilowatt-hour): 7 or 9 cents per kWh

(b) Length of contract (year): no contract, 1-year contract, or 5-year contract. During the contract period, the utility company guarantees the prices and the consumers will face with a penalty if they switch to other utility carriers

(c) Type of company:

   i. The local utility

   ii. A "well-known company other than the local utility"

   iii. An unfamiliar company

(d) Time-of-use rate: the rate policy that charges 11 cents per kWh from 8 am to 8 pm and 5 cents per kWh from 8 pm to 8 am

(e) Seasonal rates: the rate policy that charges 10 cents per kWh in summer, 8 cents in winter, and 6 cents in spring and fall

A sample of the data is shown in Table 2.4. All the attributes in this data set are categorical variables. The levels of the attributes are allocated randomly over the questions, with summary statistics listed in Table 2.5. Considering the number of respondents is higher in the electricity data set, sparse grid integration is applied in comparison to the Gauss-Hermite integration and quasi-Monte Carlo simulation.

2.5 Empirical Results

2.5.1 Gauss-Hermite Integration v.s. quasi-Monte Carlo Simulation

First, we compare the performance of Gauss-Hermite Integration and the quasi-Monte Carlo simulation with the vehicle data set provided by Train and Sonnier (2005). For the model specification, assume a linear form utility function that the utility of respondent i to choose alternative j in question t is:

$$U_{ijt} = \beta_{price,i} Price_{ijt} + \beta_{Range,i} Range_{ijt} + \beta_{EV,i} ElectricVehicle_{ijt} + \beta_{Hybrid,i} HybridVehicle_{ijt} + \beta_{Perf} Performance + \varepsilon_{ijt}$$

where $\varepsilon_{ijt} \sim EV(0,1)$ to ensure the logit form of the choice probability. To allow potential heterogeneity in respondent preferences, the coefficients of the first four attributes are set to be random. Only the effect of vehicle performance (β_{Perf}) is assumed to be fixed. Also, $\beta_{price,i}$ and $\beta_{Range,i}$ are log-normally distributed to ensure to the positive range of the marginal utility (price enters as the negative of vehicle cost). Parameters for the other two variables, *Electric Vehicle* and *Hybrid Vehicle*, are specified as normally distributed since no other prior information is available for their shape.

As Train and Weeks (2005) pointed out: "Specifying the utility coefficients to be independent implicitly constrains the scale parameter to be constant"(p.5). Consequently we estimate the vehicle choice model with correlated random parameters. To obtain a benchmark for the "true value" of the model specification, the (almost) exact maximum likelihood estimator of with a high degree (64^{th}) Hermite polynomial is adopted, with the nodes whose weights less than one-hundredth of the mean weight($1/64^4$) trimmed and the remaining nodes re-weighted. The result is shown in Table 2.6. Further, a likelihood ratio test of the hypothesis that the six correlations are zero yields a χ^2 test statistic of 24.06 ($p < 0.001$), justifying the specification of correlated random parameters.

For comparison, we estimate the model with both the quasi-Monte Carlo simulation and the Gauss-Hermite integration under differing numbers of draws and nodes, respectively. For the quasi-Monte Carlo simulation, we select two commonly adopted sequences, the scrambled Sobol and the Halton sequences, provided in MATLAB. The study by Czajkowski et al. (2017) gives a detailed description of the Sobol sequences and, in their simulations, Sobol draws were found to slightly outperform Halton draws. Both methods dominated modified Latin hypercube sampling. Under the maximum simulated likelihood estimation, 10,000, 20,000, 40,000, 80,000, and 160,000 scrambled Sobol and Halton draws per respondent were used for model estimation. In the case of Gauss-Hermite integration, a 24^{th} degree (10,416 points), a 32^{th} degree (21,312 points), a 48^{nd} degree (55,440 points), a 52^{th} degree (66,512 points), and a 56^{nd} degree (75,712) Hermite polynomial with weights less than one-tenth of the mean trimmed were used for estimation. The full estimation result is provided in Table 2.7 and 2.8.

As a measure of estimation accuracy, we calculate the percentage absolute deviation of the estimates from the benchmark true value in the (almost) exact ML model with 64^{th} degree Hermite polynomial. In Table 2.7 and 2.8, the darker a cell is marked, the higher the percent absolute deviation observed, or, the estimate accuracy is worse. Figure 2.1 and 2.2 illustrate how the individual parameters and their estimated standard errors behave over different numbers of evaluation nodes (G-H integration) or scrambled random draws.

From Figures 2.1 and 2.2, the performance of the G-H integration is much more stable and accurate than MSL with both the Halton and Sobol sequences. The G-H integration with 66,512 nodes closely approximates the mean of coefficients with all estimators within 1% of the true values. Achieving accurate estimators of the standard errors requires higher degree polynomials as several of the correlations are small and insignificant. The G-H estimated standard errors are within 2% of the true values when 75,712 nodes are used. With regard to MSL estimation, much more draws are needed to achieve the same level of accuracy, and the Halton sequence appears to dominate the Sobol sequence in general. For the mean of coefficients, the Halton sequence achieves reasonably accurate estimates when 80,000 draws are used, where all estimators are within 1% of the true values, except for the variable "Range". However, when the number of draws doubled to 160,000, the estimation accuracy decreased unexpectedly. Thus, the high accuracy in 80,000 draws is just a "lucky" point whose performance can hardly be guaranteed. For the Sobol sequence, convergence speed is much slower as all estimated means are within 4% of the true values when 80,000 draws are used, except for the variable "Range" whose mean estimator locates way far from the true value. With the number of draws doubled to 160,000, the Sobol sequence shows a nice improvement that all estimators are within 2% of the true value. Further, the MSL estimated standard errors converge even more slowly. It takes the Halton sequence 80,000 draws to have all estimated standard errors within 2% of the true values, except for the standard error of variable “Range”. But again the accuracy doesn’t improve as the number of draws is doubled. For the Sobol sequence, it can never achieve an accuracy of having all estimated standard errors to within 2% of the true value even when 160,000 draws are taken.

2.5.2 Gauss-Hermite Integration v.s. Sparse Grid Integration

In this section, we further evaluate the performance of Gauss-Hermite integration in comparison to the Sparse Grid integration which is known for its power accommodating high dimensional integration. The electricity data set is used because it has more respondents and more independent variables. To give a comprehensive view, we also apply the quasi-Monte Carlo simulation with the Halton sequence. The utility function is specified in a linear form of the attributes as:

$$\begin{aligned} U_{ijt} =& \beta_{price,i} Price_{ijt} + \beta_{length,i} ContractLength_{ijt} + \beta_{Local,i} LocalUtility_{ijt} \\ &+ \beta_{wellknown,i} WellKnownUtility_{ijt} + \beta_{Timeuse,i} TimeRate \\ &+ \beta_{seasonal,i} SeasonlRate + \varepsilon_{ijt} \end{aligned}$$

Similarly, $\varepsilon \sim EV(0,1)$ to ensure the logit form of choice probabilities. As shown in this equation, all the six attribute parameters are set to be random to allow for heterogeneity in consumers’ preferences. Thus, a respondent’s choice probability requires a 6-dimensional integration. Similarly, correlated random coefficients are considered as it is more general.

To provide a close approximation to the true model specification, the (almost) exact likelihood estimator is obtained with the G-H integration of 24^{th} degree Hermite polynomial, with the node whose weight less than one-hundredth of the mean trimmed. Note that we choose a lower polynomial degree considering the integration dimension and sample size are increased. For the sparse grid integration, we select the accuracy level k of 15, 18 and 19, with the number of evaluating nodes of 356,797, 1,044,885, 1,425,481, respectively. For the G-H integration, 15 and 20 degree Hermite polynomials are applied, with the number of nodes being 272,821 and 788,992, respectively. In addition, MSL is evaluated as a supplement with the Halton sequences of 3,000, 6,000, and 12,000 draws for each respondent, respectively, that is 1,083,000, 2,155,000, 4,332,000 draws in total. The full estimation result is shown in Tables 2.9 and 2.10, where a darker cell means higher discrepancy between the estimated and the true value.

Again, we plot the percent absolute deviation of the coefficient estimators under the three methods. For the mean of the coefficients in respondents' utility function (Figure 2.3 and Figure 2.4), G-H integration with 272,821 nodes has most estimators within 6% of the true value. When the number of nodes is increased to 788,992, the estimate accuracy is improved for all parameters except for the correlation between the "well-known Utility" and the rate policy "seasonal price". To be more specific, all estimated means are within 2% deviation of the true value with 788,992 nodes, except for three correlations (the one between "well-known utility" and "seasonal price", the one between "fixed price" and "contract length", as well as the one between "contract length" and "time of use"). For sparse grid integration, though, we find the estimation accuracy does not monotonically increase as the number of nodes grows. Although the mean estimators for coefficients and their standard errors are within 3% absolute deviation, estimators for correlations between coefficients are far from the true values. In terms of MSL with the Halton sequence, the estimation accuracy is even less satisfied, especially for the correlations between coefficients.

Such a difference is even clearer for the estimated standard errors. As shown in Figure 2.5, there is a clear and strong converge trend for the G-H estimators to approach the true value as the nodes increased from 272,821 to 788,992. The percent absolute deviations for all standard error estimators are within 3% of the true value when the number of nodes equals 788,992. However, no clear improvements in accuracy are observed for the sparse grid estimators when the number of nodes is increased from 356,797 to 1,425,481, which corresponds to accuracy levels of 15 and 20, respectively. Similarly, no strong converge trend is observed for the MSL method with the Halton sequence (Figure 2.6).

Another problem observed for sparse grid method is the negative choice probabilities that results from the re-weighting algorithm. For the sparse grid integration with the accuracy level $k = 18$, a negative integrated choice probability is observed for respondent 226 in our analysis. This obviously detracts from the usefulness of sparse grid integration within the framework of discrete choice models. Thus, we drop this respondent from the analysis. Heiss and Winschel (2008) mention that such negative approximated integrals can be seen as an evidence of extremely crude approximations that might be solved if accuracy levels are increased. However, in empirical studies where the the integral has the special meaning of choice probabilities, it might be hard for one to foresee this problem and it's difficult to find a criteria that satisfies both the goals of reducing calculation load and avoiding negative integrals.

2.5.3 Discussion

In summary, the empirical test with two choice experiment data sets validates the stable performance of Gauss-Hermite integration in the estimation of the mixed logit models. As we trimmed the negligible nodes, the calculation load of Gauss-Hermite integration is largely reduced so that the application of this method is feasible to moderate dimensional integrals, like the mixed logit model with 4 to 6 random variables. Comparing the three methods, the estimation accuracy is always improved when the number of evaluation nodes increased for Gauss-Hermite integration. While for quasi-Monte Carlo simulation and sparse grid integration, such a convergence trend can hardly be guaranteed. In addition, our results indicate that sparse grid integration may not very suitable for mixed logit models given the possibility of negative integrated choice probabilities.

2.6 Conclusion

This paper explores the potential of numerical integration in the estimation of mixed logit models, validates the feasibility and the satisfactory performance of Gauss-Hermite quadrature in approximating the choice probabilities especially when the number of random parameters is relatively small ($\leq$ 6). We draw our conclusions from empirically comparing the estimation accuracy and efficiency among three candidate methods: quasi-Monte Carlo simulation, G-H quadrature, and spare grid integration. All these approximate solutions to multivariate integrals, with the only difference in how the evaluation nodes are selected and weighted.

For the most commonly adopted estimation method, quasi-Monte Carlo simulation, we find that the problems of slow convergence rates and unstable accuracy are not negligible. Chapter 10 in Train (2009) provides a detailed treatment of the simulation bias and the simulation noise that arise from MSL estimation. For a fixed number of draws, simulation bias increases with the number of respondents (N). However if the number of draws increases faster than $\sqrt{N}$, then Train shows that simulation bias disappears asymptotically. Of course simulation noise decreases as the number of draws increases, and it also decreases with N.

To some extent, then, the rather small size (N=100) of the vehicle example may act to magnify the discrepancies between G-H and MSL estimation. Based on results reported by Czajkowski and Budzinski (2017), it appears that tripling the sample size reduces by approximately 50 percent the number of MSL draws needed for the same level of accuracy. This suggests that even with a sample of 300 respondents, tens of thousands of draws would still be required to achieve parameters and standard deviations within several percent of their true values. While other empirical models or simulation studies could be investigated, the point is that simulation bias and noise will always be a consequence of the maximum simulated likelihood approach. So even if our data set is regarded as a rather limited example, we are confident that our findings would be generally representative of more exhaustive studies.

Ultimately the question becomes how much simulation error is acceptable. Increasing the number of draws used for quasi-Monte Carlo simulation to investigate parameter stability is not a foolproof approach because, while expected simulation bias and noise decrease with increased draws, the empirical outcome is uncertain. This phenomenon can be observed for both Sobol and Halton sampling in the figures provided. The G-H method has the property that overall accuracy will always be improved as the degree of the Hermite polynomial is increased.

Finally, when an analyst is considering MSL estimation there is uncertainty as to which simulation method will work the best. Should every empirical application entertain Halton, Sobol, and modified Latin hypercube sampling in order to validate the results? Further, replicability is compromised unless the analyst has the identical, scrambled draws.

These issues can be avoided completely by instead adopting almost exact ML estimation using Gauss-Hermite quadrature. Given its ease of application and replication, it is a powerful alternative to maximum simulated likelihood as it avoids simulation bias and simulation noise and only incurs controllable approximation error. Certainly this is a desirable trade off.

In terms of the other numerical integration method, sparse grid, estimation accuracy appears questionable plus the integrated choice probability is not always positive. Similar with the simulation method, there is no monotonic relationship between the estimation accuracy and the number of quadrature nodes, so that the selection of the accuracy level in empirical analysis would be arbitrary to some extent. Also, the problem of a negative integrated choice probability is a dead end in terms of the model estimation. These two reasons distance us from applying this method in the mixed logit model. In conclusion, for typical problems in environmental and natural resources economics with a relatively small number of random coefficients and modest sample sizes, we recommend Gauss-Hermite quadrature as worthy of consideration.

References

Albright, R., Lerman, S., and Manski, C. F. (1977). Report on the development of an estimation program for the generalized multinomial probit model. *Rreport prepared by Cambridge Systematics for the Federal Highway Administration.*

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7):677–693.

Boyd, J. H. and Mellman, R. E. (1980). The effect of fuel economy standards on the us automotive market: an hedonic demand analysis. *Transportation Research Part A: General*, 14(5-6):367–378.

Breffle, W. S., Morey, E., and Waldman, D. (2005). Gaussian quadrature versus simulation for the estimation of random parameters. In *Applications of Simulation Methods in Environmental and Resource Economics*, pages 341–353. Springer.

Butler, J. S. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50(3):761–764.

Cardell, N. S. and Dunbar, F. C. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General*, 14(5-6):423–434.

Czajkowski, M., Budziński, W., et al. (2017). Simulation error in maximum likelihood estimation of discrete choice models. Technical report, Faculty of Economic Sciences, University of Warsaw.

Daly, A., Hess, S., and Polak, J. On the performance of the shuffled halton sequence in the estimation of discrete choice models. In *proceedings of the European Transport Conference (ETC)*. Strasbourg, France, 8-10 October, 2003.

Duchesne, T., Fortin, D., and Courbin, N. (2010). Mixed conditional logistic regression for habitat selection studies. *Journal of Animal Ecology*, 79(3):548–555.

Elshiewy, O., Zenetti, G., and Boztug, Y. (2017). Differences between classical and bayesian estimates for mixed logit models: a replication study. *Journal of Applied Econometrics*, 32(2):470–476.

Garrido, R. A. (2003). Estimation performance of low discrepancy sequences in stated preferences. In *the* 10^{th} *International Conference on Travel Behaviour Research*. Lucerne, Switzerland, 10-15 August, 2003.

González-Savignat, M. (2004). Competition in air transport. *Journal of Transport Economics and Policy*, 38(1):77–107.

Hall, J., Fiebig, D. G., King, M. T., Hossain, I., and Louviere, J. J. (2006). What influences participation in genetic carrier testing? results from a discrete choice experiment. *Journal of Health Economics*, 25(3):520–537.

Hausman, J. and Wise, D. (1978). Afdc participation: measured variables or unobserved characteristics, permanent or transitory. Technical report, Department of Economics, MIT.

Hausman, J. A. (1975). Project independence report: A review of us energy needs up to 1985. Technical report, MIT Energy Lab.

Heiss, F. and Winschel, V. (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1):62–80.

Hensher, D., Shore, N., and Train, K. (2005). Households willingness to pay for water service attributes. *Environmental and Resource Economics*, 32(4):509–531.

Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, 39(2-3):221–236.

Hess, S. and Polak, J. (2003). An alternative method to the scrambled halton sequence for removing correlation between standard halton sequences in high dimensions. In *proceedings of the European Regional Science Association conference*.

Hess, S., Train, K. E., and 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: Methodological*, 40(2):147–163.

Hole, A. R. (2008). Modelling heterogeneity in patients preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, 27(4):1078–1094.

Hole, A. R. and Kolstad, J. R. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and wtp space using data from a health-related choice experiment. *Empirical Economics*, 42(2):445–469.

Huber, J. and Train, K. (2001). On the similarity of classical and bayesian estimates of individual mean partworths. *Marketing Letters*, 12(3):259–269.

King, M. T., Hall, J., Lancsar, E., Fiebig, D., Hossain, I., Louviere, J., Reddel, H. K., and Jenkins, C. R. (2007). Patient preferences for managing asthma: results from a discrete choice experiment. *Health Economics*, 16(7):703–717.

Krueger, D. and Kubler, F. (2004). Computing equilibrium in olg models with stochastic production. *Journal of Economic Dynamics and Control*, 28(7):1411–1436.

Lee, J.-K., Yoo, K.-E., and Song, K.-H. A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the seoul-jeju route. *Journal of Air Transport Management*, 56.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Munger, D., LEcuyer, P., Bastin, F., Cirillo, C., and Tuffin, B. (2012). Estimation of the mixed logit likelihood function by randomized quasi-monte carlo. *Transportation Research Part B: Methodological*, 46(2):305–320.

Paterson, R. W., Boyle, K. J., Parmeter, C. F., Neumann, J. E., and De Civita, P. (2008). Heterogeneity in preferences for smoking cessation. *Health Economics*, 17(12):1363–1377.

Regier, D. A., Ryan, M., Phimister, E., and Marra, C. A. (2009). Bayesian and classical estimation of mixed logit: an application to genetic testing. *Journal of Health Economics*, 28(3):598–610.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4):647–657.

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

Scarpa, R., Willis, K. G., and Acutt, M. (2007). Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management*, 50(4):449–466.

Shen, J. (2009). Latent class model or mixed logit model? a comparison by transport mode choice data. *Applied Economics*, 41(22):2915–2924.

Skrainka, B. S. and Judd, K. L. (2011). High performance quadrature rules: How numerical integration affects a popular model of product differentiation. Technical report, No. CWP03/11, Centre for Microdata Methods and Practice (cemmap), London.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Doklady Akademii Nauk*, 148(5):1042–1045.

Srikukenthiran, S., Shalaby, A., and Morrow, E. (2014). Mixed logit model of vertical transport choice in Toronto subway stations and application within pedestrian simulation. *Transportation Research Procedia*, 2:624–629.

Train, K. (2000). Halton sequences for mixed logit. Technical report, Department of Economics, University of California at Berkeley.

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In *Scarpa R. and Alberini A. (ed) Applications of Simulation Methods in Environmental and Resource Economics*, pages 117–134. Springer.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In *Scarpa R. and Alberini A. (ed) Applications of simulation methods in environmental and resource economics*, pages 1–16. Springer.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge University Press.

Wang, X. and Kockelman, K. (2008). Maximum simulated likelihood estimation with spatially correlated observations: a comparison of simulation techniques. In *Sloboda B.W. (ed) Transportation Statistics*, pages 173–194. J.Ross Publishing.

Wasilkowski, G. W. and Wozniakowski, H. (1995). Explicit cost bounds of algorithms for multivariate tensor product problems. *Journal of Complexity*, 11(1):1–56.

Winschel, V. and Krätzig, M. (2010). Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality. *Econometrica*, 78(2):803–821.

Table 2.1: 20^{th} Degree nodes and Weights for Gaussian-Hermite Quadrature

i	ε_i^*	w_i^*		$(w_i^*)^4$	
1	-5.387	2.229E-13	(<0.00001)	2.470E-51	(<0.00001)
2	-4.604	4.399E-10	(<0.00001)	3.746E-38	(<0.00001)
3	-3.945	1.086E-07	(<0.00001)	1.391E-28	(<0.00001)
4	-3.348	7.803E-06	(<0.00001)	3.706E-21	(<0.00001)
5	-2.789	2.283E-04	0.0002	2.718E-15	(<0.00001)
6	-2.255	3.244E-03	0.0032	1.107E-10	(<0.00001)
7	-1.739	2.481E-02	0.025	3.789E-07	(<0.00001)
8	-1.234	1.090E-01	0.109	1.412E-04	0.00014
9	-0.737	2.867E-01	0.287	6.754E-03	0.0068
10	-0.245	4.622E-01	0.462	4.565E-02	0.046
11	0.245	4.622E-01	0.462	4.565E-02	0.046
12	0.737	2.867E-01	0.287	6.754E-03	0.0068
13	1.234	1.090E-01	0.109	1.412E-04	0.00014
14	1.739	2.481E-02	0.025	3.789E-07	(<0.00001)
15	2.255	3.244E-03	0.0032	1.107E-10	(<0.00001)
16	2.789	2.283E-04	0.0002	2.718E-15	(<0.00001)
17	3.348	7.803E-06	(<0.00001)	3.706E-21	(<0.00001)
18	3.945	1.086E-07	(<0.00001)	1.391E-28	(<0.00001)
19	4.604	4.399E-10	(<0.00001)	3.746E-38	(<0.00001)
20	5.387	2.229E-13	(<0.00001)	2.470E-51	(<0.00001)

Table 2.2: Sample of the Vehicle Data Set

Question	ID	Choice	Price	Operating cost	Range	Electric	Gas	Hybrid	Perf1	Perf2
1	1	0	4.676	47.43	0	0	0	1	0	0
	1	1	5.721	27.43	1.3	1	0	0	1	1
	1	0	8.796	32.41	1.2	1	0	0	0	1
2	1	1	3.377	4.89	1.3	1	0	0	1	1
	1	0	9.034	30.19	0	0	0	1	0	1
	1	0	5.71	27.16	1.8	1	0	0	1	1

Table 2.3: Summary Statistics of the Vehicle Data Set

Continuous Variables						
	Variable	N	Mean	Std Dev	Minimum	Maximum
	Price (in $10,000)	3000	3.57	1.75	0.70	9.72
	Range (in 100 miles)	3000	0.42	0.65	0.00	2.00
	Operating cost (in $)	3000	33.23	15.58	2.59	72.29
Categorical Variables						
	Variable	Level			Frequency	Percent (%)
	Engine Type					
		Electric			974	32.47
		Gas			1008	33.60
		Hybrid			1018	33.93
	Perfomance					
		High Perf			990	33.00
		Middle and low Perf			2010	67.00

Table 2.4: Sample of the Electricity Data Set

Question	ID	Choice	Alternative	Fixed price	Contract Length	Local Company	Wellknown Company	Time of Day Rate	Seasonal Rate
1	1	0	1	7	5	0	1	0	0
	1	0	2	9	1	1	0	0	0
	1	0	3	0	0	0	0	0	1
	1	1	4	0	5	0	1	1	0
2	1	0	1	7	0	0	1	0	0
	1	0	2	9	5	0	1	0	0
	1	1	3	0	1	1	0	1	0
	1	0	4	0	5	0	0	0	1

Table 2.5: Summary Statistics of the Electricity Data Set

Variable	Levels	Frequency	Percent(%)
Fixed Price			
	7 cents per kWh	3863	24.39
	9 cents per kWh	3975	25.09
	no fixed price	8002	50.52
Contract Length			
	1-year contact length	4999	31.56
	5-year contract length	5539	34.97
	no contract length	5302	33.47
Company Type			
	local company	3228	20.38
	unfamilar company	6285	39.68
	wellknown company	6327	39.94
Time of Day Rate			
	no	11882	75.01
	yes	3958	24.99
Seasonal Rate			
	no	11796	74.47
	yes	4044	25.53

Table 2.6: (Almost) Exact Maximum Likelihood Estimation for the Vehicle Data Set
(Correlated Random Coefficients)
(Gauss-Hermite Quadrature with 156,816 Points)

Parameter	Coefficient	Std. Error	Z-value
Pricelog	-0.7955***	0.1951	-4.0774
Rangelog	-0.3128	0.3925	-0.7969
EV	-1.7627***	0.4395	-4.0107
Hybrid	1.1203***	0.2208	5.0738
Performance	0.6308***	0.1081	5.8353
SE-Price	1.2248***	0.1848	6.6277
SE-Range	0.8768***	0.252	3.4794
SE-EV	1.2163***	0.4737	2.5677
SE-Hybrid	1.4175***	0.2489	5.6951
CorPr-Range	0.6834**	0.3269	2.0905
CorPr-EV	-0.2757	0.4642	-0.5939
CorPr-Hybrid	0.6397***	0.1524	4.1975
CorRange-EV	0.0981	0.6016	0.1631
CorRange-Hyb	0.838**	0.4218	1.9867
CorEV-Hybrid	0.3348	0.3555	0.9418

Table 2.7: Estimated Means for the Vehicle Data Set

Gauss-Hermite Estimated Means

Coefficient	GH64-156,816	GH24-10,416	GH32-21,312	GH48-55,440	GH52-66,512	GH56-75,712
Price	-0.794	-0.787	-0.796	-0.794	-0.793	-0.793
Range	-0.313	-0.301	-0.313	-0.314	-0.314	-0.313
EV	-1.760	-1.754	-1.765	-1.758	-1.759	-1.760
Hybrid	1.123	1.128	1.119	1.125	1.125	1.124
Performance	0.631	0.632	0.631	0.631	0.631	0.631
SE-Price	1.225	1.226	1.223	1.232	1.231	1.228
SE-Range	0.877	0.855	0.880	0.878	0.879	0.877
SE-EV	1.216	1.236	1.204	1.216	1.214	1.215
SE-Hybrid	1.419	1.420	1.416	1.422	1.420	1.419
Cor:Pr&Range	0.682	0.669	0.697	0.680	0.682	0.682
Cor:Pr&EV	-0.272	-0.243	-0.294	-0.268	-0.272	-0.272
Cor:Pr&Hybrid	0.641	0.642	0.637	0.643	0.642	0.641
Cor:Range&EV	0.100	0.097	0.113	0.100	0.101	0.101
Cor:Range&Hyb	0.837	0.806	0.859	0.837	0.838	0.838
Cor:EV&Hybrid	0.336	0.366	0.321	0.338	0.336	0.336

Sobol Estimated Means

Coefficient	GH64-156,816	Sobol 10,000	Sobol 20,000	Sobol 40,000	Sobol 80,000	Sobol 160,000
Price	-0.794	-0.791	-0.794	-0.794	-0.794	-0.794
Range	-0.313	-0.316	-0.319	-0.314	-0.311	-0.311
EV	-1.760	-1.756	-1.760	-1.760	-1.761	-1.760
Hybrid	1.123	1.122	1.123	1.124	1.124	1.123
Performance	0.631	0.630	0.631	0.631	0.631	0.631
SE-Price	1.225	1.220	1.226	1.228	1.228	1.228
SE-Range	0.877	0.884	0.886	0.877	0.874	0.874
SE-EV	1.216	1.181	1.221	1.219	1.226	1.218
SE-Hybrid	1.419	1.414	1.423	1.418	1.419	1.419
Cor:Pr&Range	0.682	0.703	0.685	0.677	0.674	0.681
Cor:Pr&EV	-0.272	-0.304	-0.274	-0.268	-0.263	-0.269
Cor:Pr&Hybrid	0.641	0.637	0.642	0.642	0.642	0.642
Cor:Range&EV	0.100	0.142	0.089	0.091	0.088	0.099
Cor:Range&Hyb	0.837	0.880	0.835	0.831	0.823	0.835
Cor:EV&Hybrid	0.336	0.313	0.336	0.338	0.344	0.339

Halton Estimated Means

Coefficient	GH64-156,816	Halton 10,000	Halton 20,000	Halton 40,000	Halton 80,000	Halton 160,000
Price	-0.794	-0.797	-0.794	-0.794	-0.794	-0.794
Range	-0.313	-0.315	-0.321	-0.315	-0.313	-0.311
EV	-1.760	-1.761	-1.756	-1.760	-1.759	-1.760
Hybrid	1.123	1.122	1.124	1.123	1.123	1.123
Performance	0.631	0.631	0.631	0.631	0.631	0.631
SE-Price	1.225	1.232	1.225	1.226	1.227	1.227
SE-Range	0.877	0.880	0.889	0.880	0.876	0.874
SE-EV	1.216	1.221	1.197	1.223	1.213	1.219
SE-Hybrid	1.419	1.424	1.424	1.421	1.419	1.419
Cor:Pr&Range	0.682	0.682	0.692	0.681	0.681	0.678
Cor:Pr&EV	-0.272	-0.269	-0.290	-0.269	-0.272	-0.267
Cor:Pr&Hybrid	0.641	0.645	0.639	0.642	0.641	0.642
Cor:Range&EV	0.100	0.088	0.117	0.088	0.104	0.097
Cor:Range&Hyb	0.837	0.827	0.852	0.830	0.839	0.832
Cor:EV&Hybrid	0.336	0.340	0.328	0.340	0.335	0.340

Table 2.8: Estimated Standard Errors for the Vehicle Data Set

Gauss-Hermite Estimated Standard Errors

Coefficient	GH64-156,816	GH24-10,416	GH32-21,312	GH48-55,440	GH52-66,512	GH56-75,712
Price	0.195	0.195	0.197	0.195	0.194	0.194
Range	0.392	0.385	0.399	0.391	0.393	0.393
EV	0.439	0.440	0.439	0.439	0.439	0.439
Hybrid	0.221	0.226	0.221	0.221	0.222	0.222
Performance	0.108	0.108	0.108	0.108	0.108	0.108
SE-Price	0.185	0.196	0.185	0.187	0.190	0.189
SE-Range	0.252	0.245	0.269	0.242	0.252	0.254
SE-EV	0.473	0.472	0.470	0.471	0.471	0.472
SE-Hybrid	0.249	0.251	0.248	0.250	0.250	0.249
Cor:Pr&Range	0.327	0.309	0.327	0.323	0.322	0.324
Cor:Pr&EV	0.463	0.439	0.475	0.455	0.458	0.461
Cor:Pr&Hybrid	0.152	0.154	0.152	0.152	0.153	0.153
Cor:Range&EV	0.602	0.599	0.607	0.604	0.605	0.603
Cor:Range&Hyb	0.421	0.406	0.428	0.413	0.415	0.418
Cor:EV&Hybrid	0.355	0.331	0.369	0.347	0.351	0.353

Sobol Estimated Standard Errors

Coefficient	GH64-156,816	Sobol 10,000	Sobol 20,000	Sobol 40,000	Sobol 80,000	Sobol 160,000
Price	0.195	0.195	0.194	0.195	0.195	0.195
Range	0.392	0.396	0.392	0.397	0.395	0.395
EV	0.439	0.441	0.439	0.440	0.440	0.440
Hybrid	0.221	0.222	0.221	0.221	0.221	0.221
Performance	0.108	0.108	0.108	0.108	0.108	0.108
SE-Price	0.185	0.202	0.184	0.183	0.187	0.188
SE-Range	0.252	0.262	0.238	0.258	0.261	0.262
SE-EV	0.473	0.703	0.429	0.468	0.484	0.478
SE-Hybrid	0.249	0.257	0.249	0.249	0.249	0.250
Cor:Pr&Range	0.327	0.541	0.292	0.311	0.333	0.332
Cor:Pr&EV	0.463	0.723	0.427	0.449	0.468	0.469
Cor:Pr&Hybrid	0.152	0.172	0.150	0.151	0.152	0.153
Cor:Range&EV	0.602	0.997	0.543	0.597	0.609	0.609
Cor:Range&Hyb	0.421	0.793	0.352	0.409	0.433	0.431
Cor:EV&Hybrid	0.355	0.479	0.334	0.350	0.355	0.360

Halton Estimated Standard Errors

Coefficient	GH64-156,816	Halton 10,000	Halton 20,000	Halton 40,000	Halton 80,000	Halton 160,000
Price	0.195	0.197	0.194	0.195	0.195	0.195
Range	0.392	0.408	0.395	0.395	0.394	0.391
EV	0.439	0.444	0.438	0.440	0.439	0.439
Hybrid	0.221	0.221	0.221	0.221	0.221	0.221
Performance	0.108	0.108	0.108	0.108	0.108	0.108
SE-Price	0.185	0.199	0.180	0.188	0.186	0.186
SE-Range	0.252	0.297	0.249	0.258	0.258	0.250
SE-EV	0.473	0.546	0.442	0.461	0.475	0.469
SE-Hybrid	0.249	0.252	0.248	0.248	0.249	0.249
Cor:Pr&Range	0.327	0.363	0.297	0.310	0.326	0.324
Cor:Pr&EV	0.463	0.540	0.440	0.448	0.469	0.455
Cor:Pr&Hybrid	0.152	0.155	0.149	0.151	0.152	0.152
Cor:Range&EV	0.602	0.663	0.576	0.578	0.602	0.595
Cor:Range&Hyb	0.421	0.506	0.348	0.395	0.420	0.410
Cor:EV&Hybrid	0.355	0.389	0.338	0.346	0.357	0.349

Table 2.9: Estimated Means for the Electricity Data Set

Coefficients	GH-24 (0.01) 2,107,328	GH-15 (0.05) 272,821	GH-20 (0.05) 788,992	Sparse Grid k=15 356,797	Sparse Grid K=18* 1,044,885	Sparse Grid K=19 1,425,481	MSL Halton 3,000	MSL Halton 6,000	MSL Halton 12,000
Price	-1.054	-1.055	-1.053	-1.059	-1.075	-1.091	-1.049	-1.037	-1.049
Length	-0.257	-0.257	-0.256	-0.255	-0.258	-0.260	-0.259	-0.256	-0.252
Local	2.696	2.677	2.695	2.729	2.741	2.601	2.711	2.685	2.694
Well-known	2.083	2.073	2.081	2.060	2.054	2.045	2.099	2.063	2.076
Time-of-use	-9.961	-9.819	-10.006	-10.041	-10.109	-10.327	-9.931	-9.827	-9.904
Seasonal	-10.168	-10.127	-10.178	-10.294	-10.152	-10.482	-10.153	-10.024	-10.124
SD Price	0.883	0.881	0.881	0.886	0.887	0.897	0.878	0.886	0.877
SD Length	0.450	0.448	0.450	0.443	0.446	0.447	0.448	0.451	0.444
SD Local	2.397	2.397	2.396	2.297	2.335	2.423	2.429	2.398	2.392
SD Well-known	1.771	1.767	1.770	1.851	1.792	1.788	1.785	1.761	1.760
SD Time-of-use	7.874	7.781	7.819	7.945	8.010	7.813	7.859	7.988	7.816
SD Seasonal	7.428	7.373	7.393	7.548	7.473	7.367	7.461	7.590	7.361
Cor: Price&Length	0.099	0.104	0.095	0.136	0.018	0.104	0.061	0.108	0.089
Cor: Price&Local	0.492	0.500	0.493	0.523	0.447	0.468	0.468	0.538	0.476
Cor: Price&W-known	0.347	0.355	0.349	0.346	0.303	0.342	0.303	0.367	0.333
Cor: Price&TOU	0.909	0.906	0.907	0.913	0.912	0.908	0.909	0.920	0.908
Cor: Price&Seasonal	0.945	0.944	0.945	0.942	0.943	0.942	0.944	0.949	0.943
Cor: Length&Local	0.234	0.230	0.235	0.214	0.287	0.250	0.281	0.249	0.240
Cor: Length&W-known	0.136	0.132	0.137	0.070	0.155	0.170	0.180	0.151	0.134
Cor: Length &TOU	0.087	0.100	0.076	0.111	0.045	0.103	0.074	0.104	0.075
Cor: Length&Seasonal	0.054	0.060	0.047	0.075	-0.006	0.045	0.029	0.064	0.043
Cor: Local&W-known	0.786	0.786	0.786	0.798	0.772	0.790	0.793	0.783	0.785
Cor: Local&TOU	0.486	0.513	0.488	0.460	0.402	0.508	0.460	0.525	0.475
Cor: Local&Seasonal	0.464	0.482	0.465	0.452	0.386	0.463	0.426	0.491	0.445
Cor: W-known&TOU	0.336	0.356	0.342	0.314	0.242	0.350	0.300	0.345	0.330
Cor: W-known&Seasonal	0.302	0.317	0.305	0.290	0.218	0.296	0.261	0.305	0.285
Cor: TOU&Seasonal	0.931	0.929	0.930	0.928	0.934	0.932	0.929	0.936	0.928

* Respondent 226 is excluded since negative integrated choice probability is observed

Table 2.10: Estimated Standard Errors for the Electricity Data Set

Coefficients	G-H24 (0.01)	GH-15 (0.05)	GH-20 (0.05)	Sparse Grid k=15	Sparse Grid K=18*	Sparse Grid K=19	MSL Halton	MSL Halton	MSL Halton
	2,107,328	272,821	788,992	356,797	1,044,885	1,425,481	3,000	6,000	12,000
Price	0.0707	0.071	0.070	0.069	0.080	0.074	0.070	0.070	0.071
Length	0.03	0.030	0.030	0.027	0.031	0.029	0.030	0.031	0.030
Local	0.1773	0.176	0.177	0.202	0.176	0.189	0.176	0.173	0.178
Well-known	0.1395	0.140	0.139	0.130	0.133	0.136	0.138	0.137	0.139
Time-of-use	0.6072	0.633	0.606	0.659	0.686	0.648	0.601	0.608	0.619
Seasonal	0.5969	0.600	0.594	0.590	0.671	0.642	0.591	0.596	0.601
SD Price	0.0736	0.075	0.073	0.072	0.068	0.075	0.072	0.075	0.073
SD Length	0.0297	0.029	0.030	0.029	0.030	0.029	0.031	0.030	0.029
SD Local	0.1758	0.177	0.176	0.159	0.152	0.178	0.178	0.188	0.175
SD Well-known	0.1427	0.144	0.143	0.145	0.125	0.143	0.143	0.152	0.145
SD Time-of-use	0.6056	0.755	0.595	0.566	0.528	0.585	0.585	0.615	0.604
SD Seasonal	0.5981	0.656	0.595	0.557	0.539	0.591	0.580	0.602	0.596
Cor: Price&Length	0.0936	0.096	0.094	0.097	0.090	0.081	0.085	0.096	0.098
Cor: Price&Local	0.0864	0.086	0.086	0.069	0.072	0.069	0.077	0.102	0.084
Cor: Price&W-known	0.0976	0.106	0.097	0.066	0.074	0.078	0.105	0.108	0.104
Cor: Price&TOU	0.0177	0.030	0.017	0.014	0.015	0.017	0.016	0.015	0.018
Cor: Price&Seasonal	0.0105	0.012	0.011	0.010	0.011	0.012	0.010	0.010	0.011
Cor: Length&Local	0.0823	0.085	0.082	0.104	0.070	0.075	0.078	0.091	0.082
Cor: Length&W-known	0.0866	0.089	0.087	0.076	0.076	0.082	0.083	0.092	0.085
Cor: Length &TOU	0.0898	0.100	0.090	0.081	0.086	0.077	0.084	0.091	0.094
Cor: Length&Seasonal	0.0943	0.097	0.094	0.091	0.086	0.079	0.085	0.097	0.098
Cor: Local&W-known	0.0407	0.041	0.041	0.035	0.036	0.040	0.040	0.041	0.042
Cor: Local&TOU	0.0827	0.107	0.081	0.066	0.068	0.063	0.075	0.101	0.083
Cor: Local&Seasonal	0.0882	0.095	0.088	0.068	0.070	0.072	0.080	0.103	0.087
Cor: W-known&TOU	0.0947	0.151	0.093	0.064	0.075	0.076	0.103	0.107	0.108
Cor: W-known&Seasonal	0.0985	0.122	0.098	0.065	0.070	0.080	0.108	0.106	0.106
Cor: TOU&Seasonal	0.0137	0.021	0.014	0.014	0.012	0.013	0.013	0.013	0.014

* Respondent 226 is excluded since negative integrated choice probability is observed

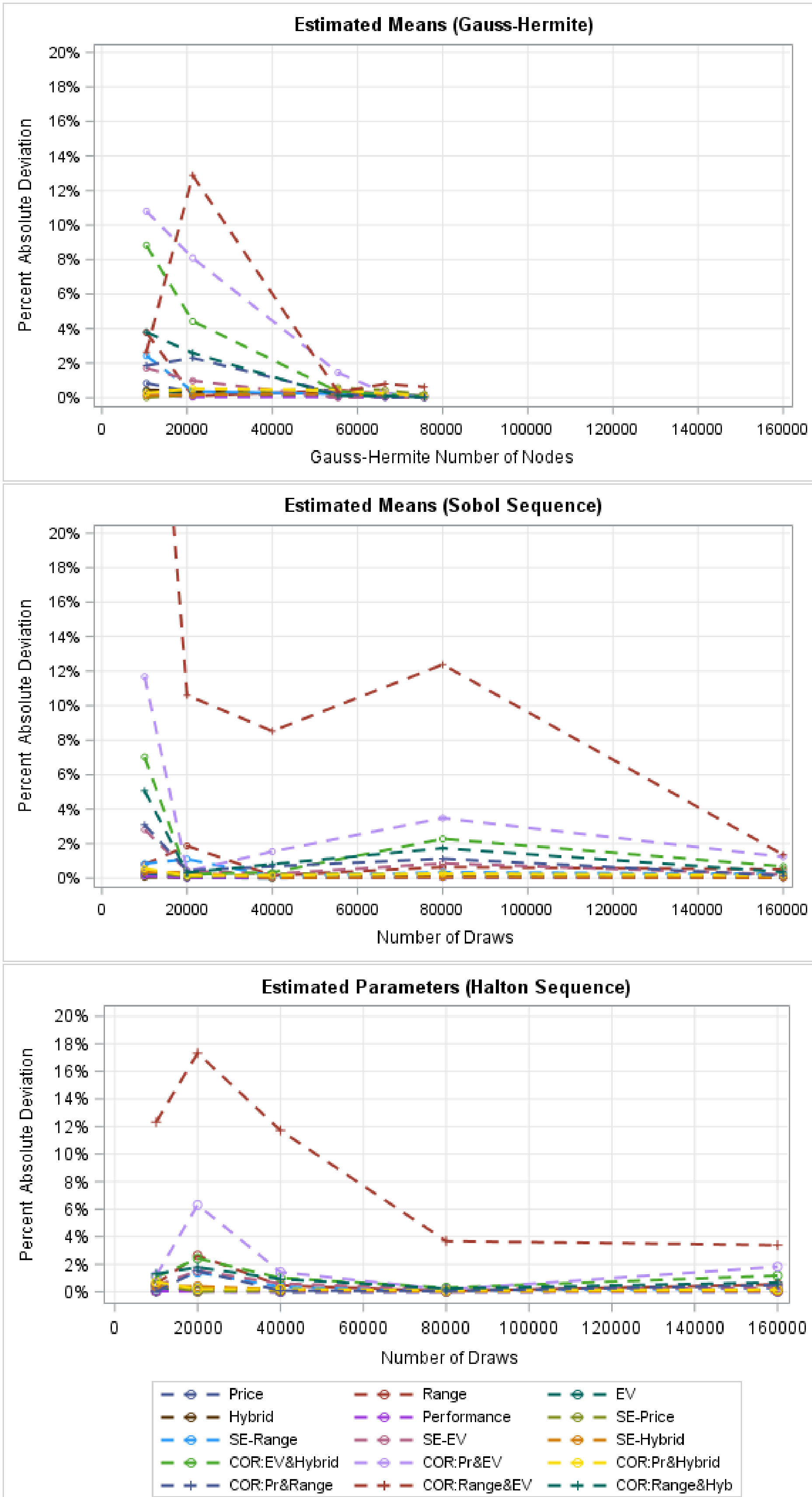


Figure 2.1: Estimated Means of the Vehicle Data Set

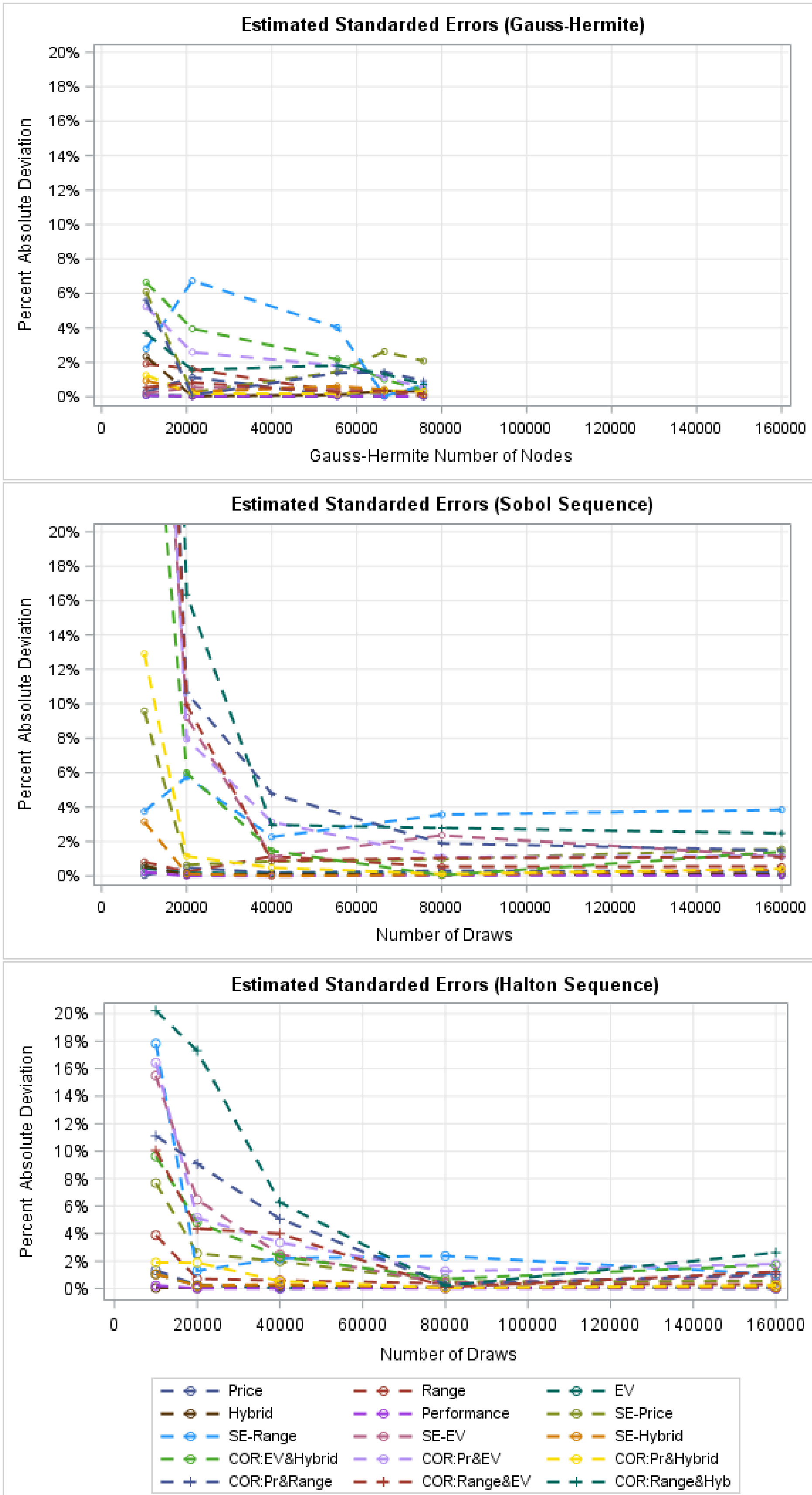


Figure 2.2: Estimated Standard Errors of the Vehicle Data Set

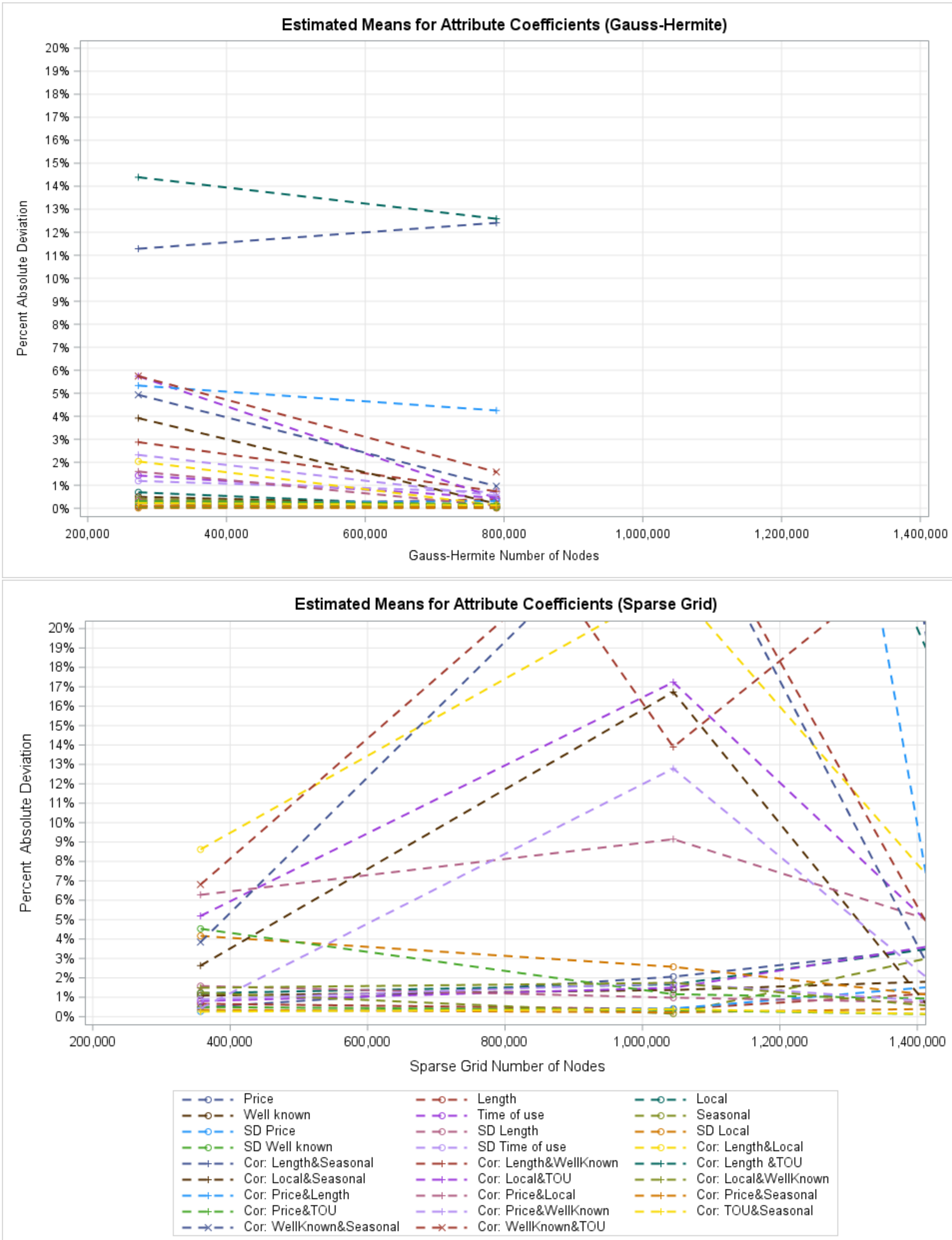


Figure 2.3: Estimated Means of the Electricity Data Set (G-H vs. Sparse Grid)

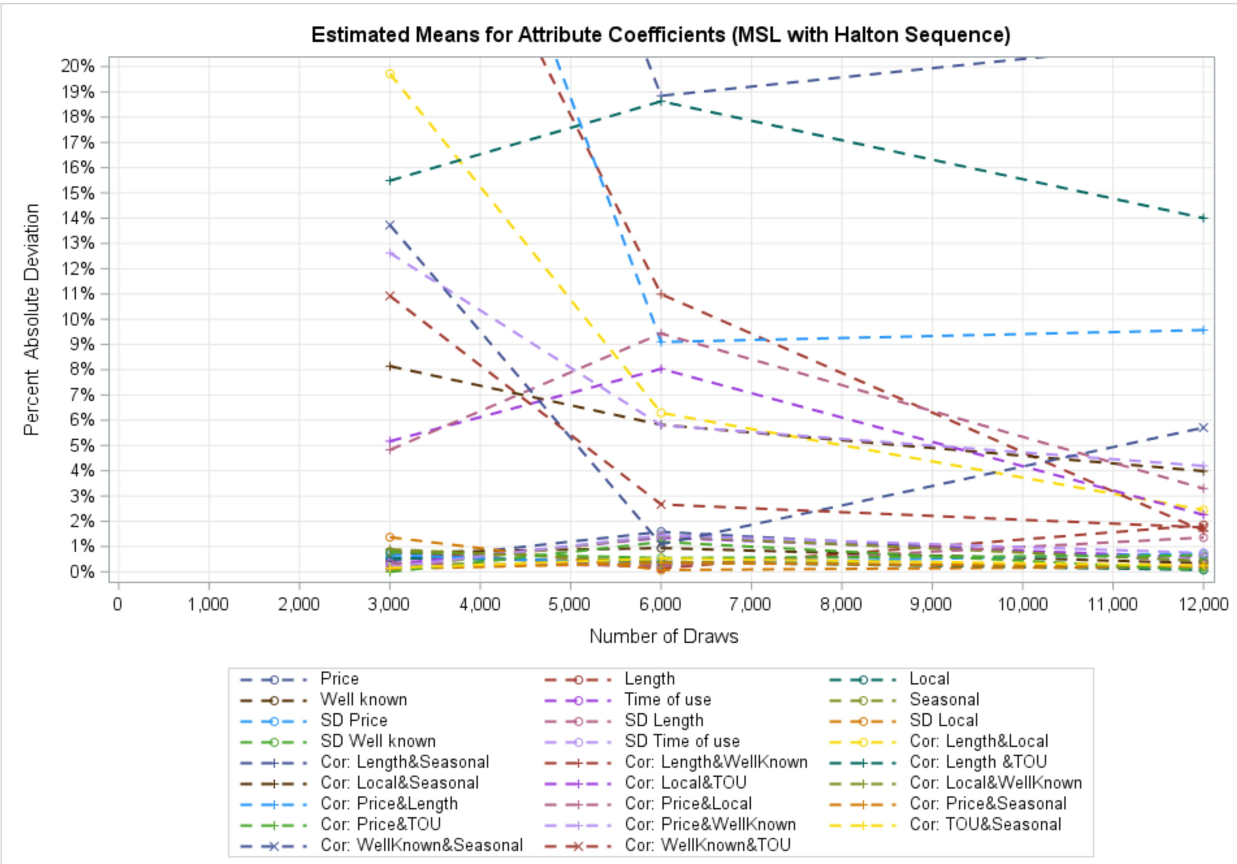


Figure 2.4: Estimated Means of the Electricity Data Set (MSL with Halton Sequence)

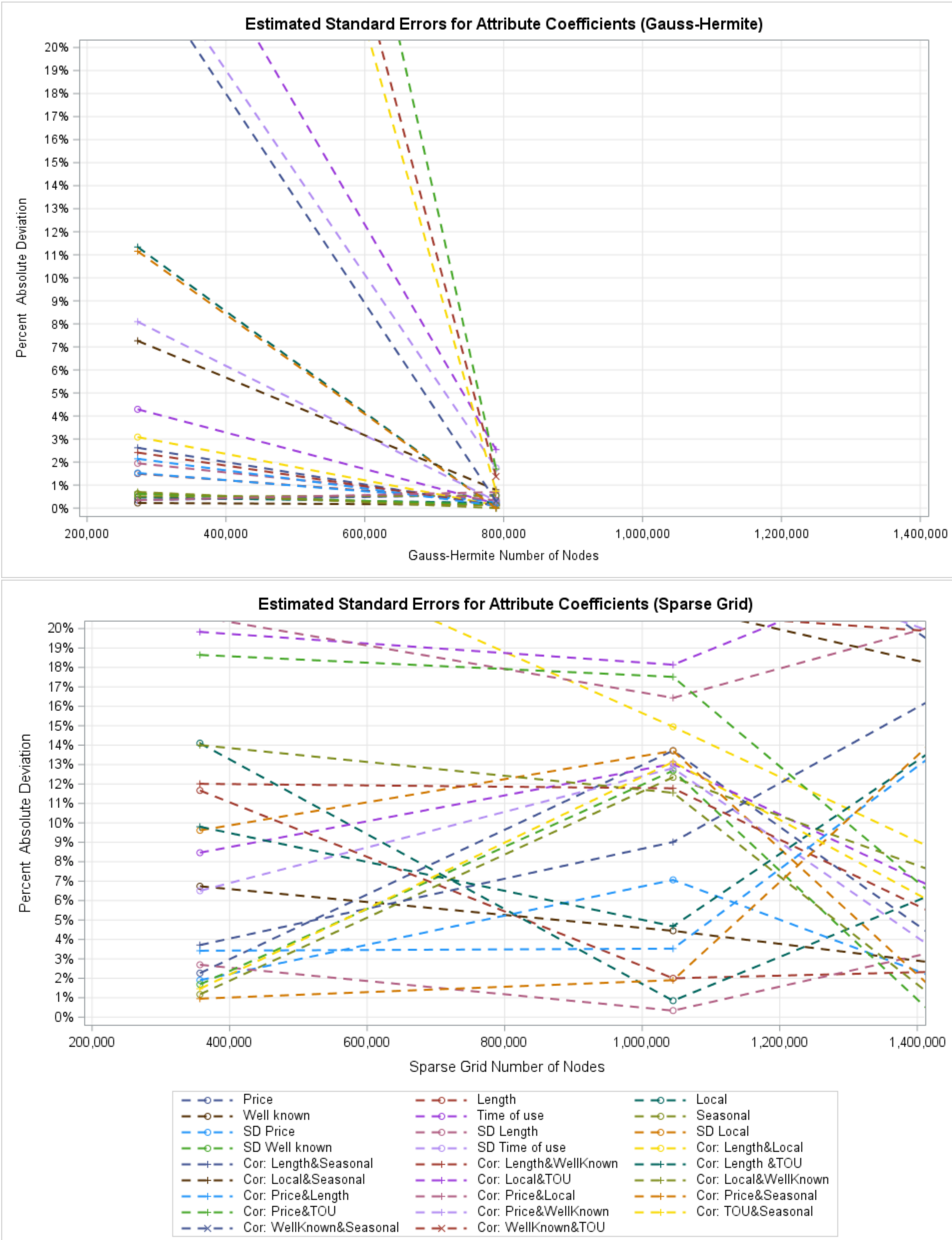


Figure 2.5: Estimated Standard Errors of the Vehicle Data Set (G-H vs. Sparse Grid)

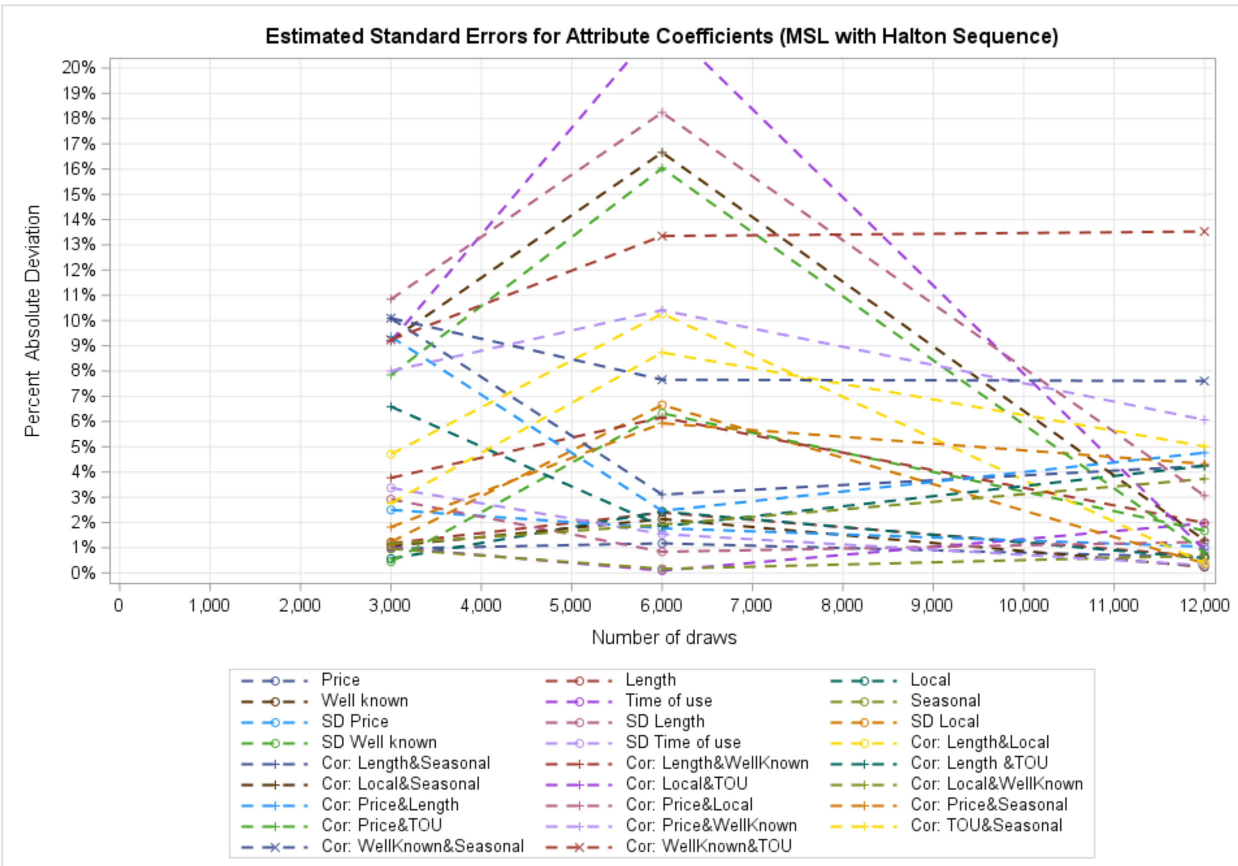


Figure 2.6: Estimated Standard Errors of the Electricity Data Set (MSL with Halton Sequence)

Chapter 3

The Distribution of Mean Willingness to Pay in Mixed Logit Models: Classical, Bayesian and Resampling Approaches

* Ying J. and Shonkwiler J. S.. To be submitted to American Journal of Agricultural Economics.

Abstract

Representation of Willingness to Pay (WTP) in the mixed logit model has long been debated given the non-zero probability of a zero-denominator when unbounded distributions are applied to the cost coefficient. However, economic theory indicates the marginal utility of income for normal goods is always positive if the consumption is non-zero. With this assumption, a well-defined approximation of the moments in WTP, which is a ratio of two normal variables, is available. This paper explores the distribution of mean WTP with three methods: the classical Delta method, the Bayesian approach with individual-level WTP, and a resampling method using the block delete jackknife. The empirical analysis shows the drawback of the Delta method, reveals the skewness introduced by a log-normally distributed cost coefficient, and validates the virtues of the Bayesian approach and the block delete jackknife. The potential of applying a normally distributed cost coefficient and the accuracy of the Bayesian approach is further validated by a synthetic data set with 2000 respondents.

Keywords: Mean Willingness to Pay, Mixed Logit Model, Bayesian Individual-level WTP, Block Delete Jackknife, Delta Method

3.1 Introduction

As one of the most popular methods in discrete choice modeling, the mixed logit model has been widely applied to derive consumers' willingness to pay (WTP) for goods and services in the field of non-market valuation, transportation, health economics, etc. WTP is the maximum price at or below which a consumer will definitely buy one unit of a product – that is – the marginal rate of substitution (MRS) between income and the quantity expressed by the attribute given the utility level is fixed (Small and Rosen, 1981). For a linear utility function, the point estimation of WTP could be simply expressed as the ratio between the coefficient for the attribute of interest and the coefficient for the cost (price) variable. For the traditional multinomial logit model, WTP is a fixed value given both coefficients in the WTP ratio are fixed under the assumption of "the same preference for all respondents". In contrast, WTP in the mixed logit model is represented by a random distribution as the coefficients in the utility function could be random. To this end, mixed logit models are more flexible and realistic by allowing more variation in consumers' taste, unrestricted substitution patterns and correlations in unobserved factors over time (Train, 2009).

However, adding more flexibility also brings more complexities in the calculation of willingness to pay. A common concern with the mixed logit model is the non-zero probability of a zero denominator in the WTP ratio when the cost coefficient follows a distribution that spans over zero. For example, the normal distribution is one of the most popular distributions for coefficients in the utility function when no prior information is available on the shape of individuals' marginal utility. However, the range of the normal distribution overlaps zero so that a random draw from a normally distributed cost coefficient could be extremely close to zero. In that case, an extremely large WTP might be observed and severely plague the welfare analysis. Due to this concern, traditional simulation approaches like the Krinsky and Robb (K&R) procedure proposed by Hensher and Greene (2003) could be biased given the the first two moments may not exist for a WTP defined as the ratio of normal variables.

More specifically, Daly et al. (2012) demonstrate that some popular parameterizations for the cost coefficient, such as normal, truncated normal, uniform, and triangular, can imply infinite moments for WTP if the probability density is positive when the cost coefficient equals to zero. They also point out that simulation can serve to mask the non-existence of the moments in WTP ratio by providing finite simulated moments.

To avoid the non-existance of moments in the WTP ratio, most existing studies advocate using bounded and sign restricted distributions to shift the distribution away from zero. Distributions like log-normal, censored normal, and the Johnson S_B distribution (Train and Sonnier, 2005) are widely applied. However, imposing bounds or specific shapes on the cost coefficient could be arbitrary (Daly et al., 2012). It may also fail to reflect the shape of the real data (Hensher and Greene, 2003; Cirillo and Hetrakul, 2010; Hess et al., 2005). Empirical studies have found that applying the log-normal distribution may introduce biased cost coefficients due to the heavy tails and consequently over-estimate the WTP ratio (Balcombe et al., 2009; Rigby et al., 2009; Hole and Kolstad, 2012). Additionally, the sampling variability of WTP measures under a lognormal distribution may be extreme and, in fact, this phenomenon has caused investigations into welfare measures derived from mixed logit models specified in willingness to pay space.

Another approach is to divide the attribute coefficient by the mean of cost coefficient to avoid extreme measures in WTP ratio. A main draw back of this method is it sacrifices the variability in WTP measures, and the ratio may not be interpreted as the mean of WTP but a WTP derived from the coefficients of the "average individual" for each parameter (Sillano and de Dios Ortúzar, 2005). Also, for most empirical analyses, simply fixing the price parameter could be flawed due to the unrealistic assumption of the same marginal utility of income for every respondent – no heterogeneity exists (Meijer and Rouwendal, 2006; Scarpa and Rose, 2008; Daly et al., 2012).

Then, what is the analyst to do if a normally distributed (unbounded, no sign restricted) price parameter empirically fits the data much better than a log-normally distributed one? To answer this question, the most important pre-condition is to ensure the stable performance of the moments for WTP ratio. The ratio of two correlated normal random variables has long been discussed in the statistics literature (Geary, 1930; Fieller, 1932; Box, 1958; Marsaglia, 1965; Hinkley, 1969). The ratio z of two centred normal variables is a Cauchy variable, and the ratio z of two arbitrary normal variables lead to a Cauchy-like distribution (Cedilnik et al., 2004). It is well recognized that either the Cauchy or the Cauchy-like distribution does not have finite moments of order greater than or equal to one. This property is also addressed by Daly et al. (2012) to show the inappropriateness of a normally distributed cost coefficient.

However, Marsaglia (2006) points out a fact that is frequently observed but rarely investigated: when handling general ratios, in theory, none of the moments exist yet practical considerations suggest there should be approximations whose adequacy can be verified. Further, his research suggests many of the ratios of normal variables encountered in practice can themselves be taken as normally distributed. Further, if the denominator is always positive, the ratio of two joint normal variables is approximately normally distributed with a well-defined mean and variance.

Then, do we have the basis to motivate a non-zero cost-coefficient in consumers' indirect utility functions? The answer is yes, if we focus on normal goods and ignore the possibility of zero consumption for a respondent for all survey questions. In microeconomic theory, the partial derivative of indirect utility to the i^{th} price or cost variable is $\frac{\partial U^*}{\partial cost_i} = -\lambda x_i^*$, where λ is the marginal utility of income $\frac{\partial U}{\partial M}$, and x_i^* is the compensated or Hicksian demand for the good. For a normal good and for the representative population (ignoring the extremely wealthy whose marginal utility from income may be negligible), the marginal utility of income is positive. If we also exclude the extreme situation where a respondent chooses the opt-out alternative "none of the above" for all choice experiment questions in a survey, which means the respondent is either not serious with the survey or cares nothing about the topic surveyed, we will have a non-zero Hicksian demand x^* so that the marginal indirect utility with respect to price or cost is non-zero. In fact, only when respondents make some choices can we extract information about their underlying utility function.

With the justification of stable approximated moments for the WTP ratio, this paper advocates examining the distribution of the most important statistic – the mean – of WTP in mixed logit models under different distributions for the cost coefficient, like the normal and log-normal distribution. From the central limit theorem, the normal distribution is the most appropriate for a random variable with no prior information. It also has the attribute of equivalent mean and median. The log-normal distribution, on the other hand, is advantageous in ensuring the sign of the cost coefficient, while the potential drawback is the shape with heavy tails and difference between the mean and the median.

In calculating the distribution of mean WTP, we apply the following three approaches. The first approach is the Delta method, what we also label as the "classical method" given its wide application in measuring the distribution of a function of variables. Bliemer and Rose (2013) derived the formula of the Delta method in random coefficient models. However, their empirical application confuses the standard error of the sample mean with the standard error of the sample. In our application, we distinguish confidence intervals of the mean WTP from confidence intervals of the general WTP. Taking the mean WTP as a statistic of interest, the standard error of the estimator would be $\frac{std(WTP)}{\sqrt{n}}$. Beside, the credible interval of the mean WTP can also be achieved by posterior draws of individual-level coefficients from the Bayesian approach. Thirdly, resampling methods are explored in describing the variations of mean WTP. The block-delete jackknife is suggested considering the potential correlations among questions answered by the same respondent in the survey data.

As a criteria to examine the estimators of mean WTP, the conditional logit model has the valuable property that under proper specification of the conditional mean, it yields consistent estimators even with distributional misspecification (Gourieroux et al.,1984). Also, the variances of the multinomial logit estimators can be consistently estimated using robust methods (White, 1982). Further, since the Independence of Irrelevant Alternatives (IIA) holds at individual level in both the mixed logit model (Hahn et al., 2017) and the conditional logit model, the welfare measures from the conditional logit model (which is averages of the individuals in the sample – but the individuals follow IIA) could be treated as a counterpart to the "mean WTP" from the mixed logit model. Thus, WTP estimators derived from the conditional logit model should provide a reasonable approximation to the true value. Furthermore, a synthetic data set is generated and evaluated to further test the validity of our conclusion. Given the real utility function is known, the theoretical true moment of WTP can be calculated and serve as the criterion to examine the performance of our estimators.

The rest of the paper is organized as follows. The next section introduces the three methods of evaluating mean WTP. Section three introduces the data we apply in the empirical analysis. Section four presents the empirical results with a discussion. Section five applies a robustness check with a synthetic data set, and the conclusion section summarizes key findings of this paper.

3.2 Methodology

3.2.1 Mixed Logit Model and the Classical Approach (Delta Method)

Let's define the utility that decision maker i obtains from alternative j to be U_{ij}, $i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, J$. U_{ij} includes an observed part V_{ij} and an unobserved part ε_{ij}, so that $U_{ij} = V_{ij} + \varepsilon_{ij}$. V_{ij} is the systematic part of the utility function and is completely deterministic to the decision maker and captured by the researcher. It is a linear or non-linear function form of both the price and amenity attributes of interest. ε_{ij} captures the factors that affect utility but are not included in V_{ij}, and is generally assumed to be distributed independently and identically (iid) extreme value type 1 ($EV1$ or Gumble) distributed. Usually, V_{ij} is defined in a linear form $V_{ij} = \sum_{k=1}^{K} \beta_k x_{ijk}$, where x_{ijk} is the attribute including price with the total number of K. Under such specification, a closed-form expression of the logit form probability in the multinomial logit model can be derived as below.

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{j=1}^{J} e^{V_{ij}}} = \frac{e^{\sum_{k=1}^{K} \beta_k x_{ijk}}}{\sum_{j=1}^{J} e^{\sum_{k=1}^{K} \beta_l x_{ijk}}}$$

The mixed multinomial logit model extends the traditional multinomial logit by relaxing the restriction of homogeneous preference, assuming the cost coefficient β_k to follow a distribution instead of being a single value to account for potential heterogeneity. Correspondingly, the probability of respondent i choosing alternative j is

$$P_{ij} = \int \frac{e^{\sum_{k=1}^{K} \beta_k x_{ijk}}}{\sum_{j=1}^{J} e^{\sum_{k=1}^{K} \beta_l x_{ijk}}} f(\beta_1, \ldots, \beta_K) d\beta$$

Willingness to pay is defined as the monetary value of the population averaged maximum utility associated with amenity changes. It is the ratio of the marginal utility of the attribute to the marginal utility of its cost.

$$WTP_k = \frac{\partial U_j / \partial x_{jk}}{\partial U_j / \partial cost_j} = \frac{\beta_k}{\beta_{price}}$$

Where β_k is the coefficient for amenity k, β_{price} is the coefficient for the price. In the mixed logit model, both the numerator and denominator are random variables. Following Severini (2005), the asymptotic distribution of a ratio of two possibly correlated random means can be shown to be normally distributed with a well-defined variance. For example, if β_k and β_{price} are assumed to be normally distributed, we can write $\beta_k = \mu_k + \sigma_k \varepsilon$ and $\beta_{price} = \mu_{price} + \sigma_{price} \varepsilon$, where ε is standard normally distributed, and the mean of WTP can be written as

$$\theta = E\left(WTP_K\right) = \frac{E\left(\mu_k + \sigma_k \epsilon\right)}{E\left(\mu_p + \sigma_p \epsilon\right)} = \frac{E\left(\mu_k + \gamma_k\right)}{E\left(\mu_p + \gamma_p\right)}, \quad and \quad E\left(\gamma_k, \gamma_p\right) = \sigma_{ap}$$

To investigate the variance of $E\left(WTP_K\right)$, the Delta method is applied. Since σ_k $(k = a, p)$ is estimated by $V\left(\sigma_k\right) = \sigma_k^2 + \Sigma_{\sigma k}$, where Σ is the variance-covariance matrix of the estimated parameters, we have

$$V(\theta) \approx \left\{ V(\mu_k + \sigma_k \epsilon) + \frac{\mu_k^2}{\mu_p^2} V(\mu_p + \sigma_p \epsilon) - \frac{2\mu_k}{\mu_p} \left[cov(\mu_k, \mu_p) + cov(\gamma_k, \gamma_p) \right] \right\} / n\mu_p^2$$

or with the variance-covariance matrix Σ,

$$V(\theta) \approx \left\{ \Sigma_{\mu_k} + \sigma_k^2 + \Sigma_{\sigma k} + \frac{\mu_k^2}{\mu_p^2} \left(\Sigma_{\mu_p +} \sigma_p^2 + \Sigma_{\sigma p} \right) - \frac{2\mu_k}{\mu_p} \left[\Sigma_{\mu_k, \mu_p} + \sigma_{kp} \right] \right\} / n\mu_p^2$$

where n is the number of total respondents since each respondent′s parameters are perfectly correlated over their responses. We treat such approach as the classical approach to describe the distribution and the confidence interval of mean WTP in mixed logit models. However, the Delta method can be badly biased (Efron, 1981 p.595) and it is equivalent to the infinitesimal jackknife (Efron, 1981, p589-99). Given that the infinitesimal jackknife is a limiting form of a delete-1 jackknife, we expect Delta methods standard errors to be biased as a result of the unreasonable assumptions of the data structure.

3.2.2 Bayesian Approach with the Individual-level WTP

The Hierarchical Bayes model is applied to obtain posterior distributions of hyper-parameters that shape the distribution of parameter β_k in the utility function. Refer to Train (2002), the funciton and corresponding probability (likelihood) of decision maker i $(i = 1, 2, \ldots, n)$ to choose alternative j $(j = 1, 2, \ldots, J)$ in question t $(t = 1, \ldots T)$ is:

$$u_{ijt} = \beta_i' x_{ijt} + \varepsilon_{ijt}; \quad L(y_i|\beta_i) = \prod_t \left(\frac{e^{\beta_i' x_{ijt}}}{\sum_j \beta_i' x_{ijt}} \right) \quad (normal)$$

$$u_{ijt} = \left(e^{\beta_i}\right)' x_{ijt} + \varepsilon_{ijt}; \quad L(y_i|\beta_i) = \prod_t \left(\frac{e^{\left(e^{\beta_i}\right)' x_{ijt}}}{\sum_j e^{\left(e^{\beta_i}\right)' x_{ijt}}} \right) \quad (log-normal)$$

where β_i is the vector of coefficients for respondent i. By assuming β_i to be normally distributed, the conjugate posterior distribution can be applied and thus simplifies the estimation. Also, rewrite the likelihood using the hyper-parameters mean (b) and the variance covariance matrix (W), the likelihood function is $L(y_i|b,W) = \int L(y_i|\beta_i)\,\phi(\beta_i|b,W)\,d\beta_i$. Assuming the respondents have independent choices from each other, the joint posterior distribution of b, W, can be written as the product of posterior distribution of all the n respondents:

$$K(b,W|Y) = \prod_n L(y_i|b,W)k(b,W)$$

Although direct draws from the posterior distribution are possible with the Metropolis-Hastings algorithm, the calculation process is too time-consuming in simulating the likelihood $L(y_i|b,W)$ for each respondent i. At the same time, the property of simulated means of the posterior draws would be affected given the probability cannot be captured without simulation (Train, 2009). Instead, treating β_i as a parameter along with b and W can largely simplify the process by introducing the Gibbs sampling method. For all the n respondents, the joint posterior distribution of b, W, and β_i can be expressed as

$$K(b,W,\beta_i|Y) \propto \prod_n L(y_i|\beta_i)\,\phi(\beta_i|b,W)\,k(b,W)$$

Conditional posteriors on $b|W,\beta_i$ and $W|b,\beta_i$ can be derived conveniently when diffuse priors are applied. Set the prior on b as normal with an unboundedly large variance, the posterior on b is a $N(\bar{\beta}, W/N)$, where $\bar{\beta}$ is the sample mean of the β_i. With an inverted Wishart prior for W, the posterior on W is inverted Wishart with $K+N$ degrees of freedom and an updated scale matrix that combines the information from β_i and b. Then, the M-H method can be used to draw the posterior β_i for all the n respondent by comparing a random draw of a uniform variable μ with the ratio F: If $\mu \leq F$, accept $\tilde{\beta}_i^1$; if $\mu > F$, reject $\tilde{\beta}_i^1$.

$$F = \frac{L(y_n|\tilde{\beta}_i^1)\phi(\tilde{\beta}_i^1|b,W)}{L(y_n|\tilde{\beta}_i^0)\phi(\tilde{\beta}_i^0|b,W)}$$

With the Gibbs sampling for conditional posteriors for b and W, we avoid the simulation of the likelihood in the M-H method.

For each decision maker, m random posterior draws could be obtained for each coefficient in the utility function (m is very large, like 5000). In each posterior draw, dividing the coefficient on amenity k over the coefficient of the cost variable provides m draws of WTP for each respondent, so that the mean WTP for each individual is known. Then the overall sample distribution of mean WTP for all the n decision makers is also available.

Note that since the prior mean (b) of the cost coefficient is set to follow a diffuse normal distribution, statistically, it is possible to obtain some posterior draws of the cost coefficient that are extremely close to zero. However, economically, an extremely small cost coefficient indicates the respondent places negligible or zero weight on the cost variable (Sillano and de Dios Ortúzar, 2005), which means he or she has a zero marginal utility of income. This is apparently unrepresentative in terms of the normal good we discuss in this paper. We should distinguish the statistical assumptions from economic assumptions. Thus, we suggest a criterion to drop those draws with extremely small cost coefficients to ensure the validity in an economic sense. Further, the possible reason for negligible marginal income utility could be the respondent's lack of interest in the survey, survey fatigue, or the lack of a budget constraint in the survey questions.

3.2.3 Resampling Approach with the Block Delete Jackknife

Resampling methods are another approach to obtain the variability of an estimator by using subsamples from the original dataset. The Jackknife is one of the most commonly used resampling methods that could be applied to construct reasonably reliable confidence intervals for a wide variety of estimators.

Instead of applying the most widely used method, $delete-1$ Jackknife, which is typically suitable for identical and independent data, we advocate to adopt the block delete jackknife in estimating the confidence interval of mean WTP in mixed logit models. The reason is, respondents are almost always given more than one choice question in a choice experiment survey, so grouping the answers from the same respondent is natural as the preference of the same respondent is assumed to be constant for all questions he or she answered. In other words, it is reasonable to assume the independence among answers from different respondents, but not necessarily the same (independent) for the questions answered by the same respondent. The delete-1 jackknife fails to recognize this and thus its limiting form, the infinitesimal jackknife, which is equivalent to the Delta method that we mentioned before, is also biased. Further, Shao and Wu (1989) show the block delete jackknife works properly when the estimators are non-smooth. Thus, we will develop the delete-m Jackknife standard deviation of mean WTP in the future to describe the confidence interval of the mean WTP.

For delete-m jackknife, suppose the sample is divided into g mutually exclusive and independent groups of equal size at a time (extension to unequal group sizes is trivial), then the jackknife bias corrected estimator is calculated g times by using all but one group of observations. In a mixed logit model with balanced data, suppose we have n decision makers and T questions for each decision maker ($L = nT$ observations in all), then the $delete - T$ jackknife estimator of WTP is the mean of all pseudo-values estimators of block delete jackknife

$$WTP_{Jackknife(T)} = mean\left(\widetilde{WTP}_{(j)}\right) = \frac{1}{n}\sum_{j=1}^{n}\widetilde{WTP}_{(j)}$$

$$where \qquad \widetilde{WTP}_{(j)} = n\widehat{WTP}_L - (n-1)\,\widehat{WTP}_{(j)}, j = 1, \ldots, n$$

$\widehat{WTP}_{(j)}$ is the WTP calculated with the j^{th} block of T questions removed (individual j), $\widehat{WTP}_L$ is the WTP calculated with all L observations. The variance estimator of the pseudo-value estimators of block delete jackknife is

$$var\left(\widetilde{WTP}_{(j)}\right) = \frac{1}{n-1}\sum_{j=i}^{n}\left(\widetilde{WTP}_{(j)} - \frac{1}{n}\sum_{j=1}^{n}\widetilde{WTP}_{(j)}\right)^2$$

It should be noted that, we will need $\frac{var\left(\widetilde{WTP}_{(j)}\right)}{n}$instead of $var(\widetilde{WTP}_j)$ to construct the confidence interval of the mean WTP estimated from delete-T jackknife estimator as

$$var\left(\bar{X}\right)=var\left(\frac{Y_1+Y_2+\ldots+Y_n}{n}\right)=\frac{var\left(Y_1\right)+\ldots+var\left(Y_n\right)}{n^2}=\frac{\sigma_Y^2}{n}$$

Thus,

$$var\left(WTP_{Jackknife(T)}\right)=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{n-1}\left(\widetilde{WTP}_{(j)}-\frac{1}{n}\sum_{k=1}^{n}\widetilde{WTP}_{(k)}\right)^2$$

$$=\frac{n-1}{n}\sum_{j=1}^{n}\left(\widehat{WTP}_{(j)}-W\bar{T}P_{(T)}\right)^2$$

With the mean and sample variance of the block-delete jackknife estimator of WTP known, the 95% confidence interval for $WTP_{Jackknife(T)}$ can be constructed based on the Central Limit Theorem that the distribution of the sample mean converges to the standard normal distribution function as $n \rightarrow \infty$.

Similarly, the block delete jackknife variance of the maximum likelihood (ML) estimator could be derived to compare with the variance from the classical Delta Method. Let WTP^* denote the ML estimator of WTP over the entire sample and let WTP_j^* denote the ML estimator of WTP with the j^{th} block deleted, a more direct form of the variance is

$$var\left(WTP^*\right)=\frac{(n-1)}{n}\sum_{j=1}^{n}\left(WTP_j^*-WTP^*\right)^2$$

3.3 Data

For ease of replication, we choose a data set on consumers choice among different types of vehicles (gas, electric and hybrid) under different combinations of price, operating cost, range, and performance. The data was provided by Train and Sonnier (2005). This experiment is a part of the survey that targeted vehicle owners in California. Californias Air Resources Board has been interested in the promotion of vehicles with fewer pollutant emissions.

The survey contacted respondents randomly throughout the state by telephone and those who planned to purchase a car within three years were invited to participate in this survey.100 participants are included in total. Each participant received a questionnaire with 15 questions of vehicle choice. To make the data set a balanced panel, we adopt the 10 first choice experiments for each respondent. Thus, there are 100×10 choice situations with 3000 alternatives. For each question, three different vehicle types are listed, and the respondents are asked to choose the one vehicle that they prefer the most given the attributes assigned to the different vehicles. A sample of the data is shown in Table 3.1. Attributes assigned to each alternative include,

1. Engine type: gasoline, electric, or hybrid
2. Purchase price (in $\$10,000$)
3. Operating cost in dollars per month
4. Performance:
   (a) High performance: Top speed of 120 mph, and 8 seconds to reach 60 mph
   (b) Middle performance: Top speed of 100 mph, and 12 seconds to reach 60 mph
   (c) Low performance: Top speed of 80 mph, and 16 seconds to reach 60 mph
5. Range: miles between refueling/recharging

Considering the high correlation coefficient between the purchase price and the operation cost (p-value < 0.001), we drop the operation cost variable in the following analysis. We also combine the middle performance and the low performance as one group to simplify the analysis.

The choice experiments were designed to provide wide variation in each attribute and as little covariance among attributes as possible while maintaining plausibility. A summary of the continuous variables is provided in Table 3.2. Other variables, including the dependent variable *Choice*, the independent variables *Range*, *Electric*, *Hybrid*, and *Performance 1 (high)*.

3.4 Empirical Results

3.4.1 Multinomial Logit Model

Maximum likelihood estimation (MLE) results for the multinomial logit model are shown in Table 3.3. As mentioned before, the conditional logit model should yield consistent estimators even under distributional mis-specification if the conditional mean is properly specified. Moreover, variance of estimators can be consistently estimated using robust methods. Thus, the mean WTP from the basic logit model can be seen as a reasonable approximation to the true value that can be used to compare with the estimators from the mixed logit model.

The Mean WTP premium for Hybrid vehicle (versus gas vehicle) is

$$E\left(\widehat{WTP}_{Hybrid}\right) = E\left(\frac{\hat{\beta}_{Hybrid}}{\hat{\beta}_{price}}\right) = 1.399$$

According to the Delta method, the variance of the mean WTP is

$$var\left(\widehat{WTP}_{Hybrid}\right) = var\left(\frac{\hat{\beta}_{Hybrid}}{\hat{\beta}_{price}}\right) = \left(-\frac{\hat{\beta}_{Hybrid}}{\hat{\beta}^2_{Price}}, \frac{1}{\hat{\beta}_{price}}\right)\Sigma_{PH}\left(-\frac{\hat{\beta}_{Hybrid}}{\hat{\beta}^2_{Price}}, \frac{1}{\hat{\beta}_{price}}\right)'$$

where Σ_{PH} the robust variance covariance matrix of the price and hybrid coefficients, so that the standard error is $std(E(\widehat{WTP}_{Hybrid})) = 0.2274$.

We also estimate the standard error of mean WTP through the jackknife method. Delete-1 jackknife results in a sample standard error of 0.2272, which is very close to the Delta methods result. This is because the Delta method is equivalent to the infinitesimal jackknife which is a limiting form of delete-1 jackknife. In contrast, the standard error estimated from the block delete jackknife is 0.318. As we discussed before, the block delete jackknife is more appropriate in terms of the survey data given the dependence among questions answered by the same respondent. Thus, it is possible that the Delta method underestimates the standard error of mean WTP in the basic logit model.

3.4.2 Mixed Logit Model

As an alternative to the maximum simulated likelihood estimation (MSL), we employ the Gauss-Hermite (G-H) quadrature approach in the estimation of the mixed logit model. In this approach, we extend the Gauss-Hermite integration method that was first introduced by Breffle et al. (2005) in the estimation of a Probit choice model. We briefly outline this approach in the Appendix B. The distribution of mean WTP in mixed logit models is calculated with each of the three methods: the classical (Delta method) approach, the Bayesian individual-level WTP approach, and the block delete jackknife approach. We estimated the mixed logit model under both uncorrelated and correlated random parameter settings, with both normal and log-normal distributed cost coefficient, respectively.

Case I: Uncorrelated Random Parameters

(1) Normal Distributed Cost Coefficient

We first set the coefficients for all the four attributes: price, range, electric, and hybrid to be normally distributed. In this case, WTP for hybrid vehicle is a normal to normal ratio. Maximum likelihood estimation results with Gauss-Hermite Quadrature is provided in Table 3.4. With the maximum likelihood (ML) estimators, the mean WTP is calculated to be 1.3414 and the sample standard error of mean WTP from the Delta method is 0.1709, while the sample standard error of mean WTP calculated by the block delete jackknife method is 0.2560, which is much larger than that from the Delta method.

The Bayesian posterior means and standard errors of the hyper-parameters b and W are provided in Table 3.5. The Bayesian result is very close to the MLE result in both the coefficient means and standard errors, validating the coherence between the two approaches. At the same time, we calculate the the mean and standard error of the mean WTP for the 100 respondents through a sample of 5000 posterior draws on conditional posteriors for both β_{hybrid} and β_{price}. For each respondent, we average the β_{hybrid} and β_{price} in the 5000 draws to obtain their mean marginal utility on hybrid vehicle and their mean marginal utility of income. As shown in Table 3.6, we drop respondents whose mean price coefficient in the 5000 posterior draws is smaller than a criterion $\hat{\beta}_{price}$ as they place negligible or zero weight on the price variable. We suggest choosing $\hat{\beta}_{price_i} = 0.1$ given the standard error of mean WTP tends to be stable in this region. In this case, respondents$'$ mean of WTP on hybrid cars (versus gas cars) is 1.717, with the sample standard error of 0.2333. To give a more comprehensive view, Figure 3.1 shows the distribution of mean WTP under different criteria $0.01 < \hat{\beta}_{price_i} < 0.2$ in Table 6, with the dashed blue line marks the mean WTP with the screen criterion $\hat{\beta}_{price_i} = 0.1$.

On the other hand, we can obtain the distribution of mean WTP by calculating the mean WTP for all respondents in each draw first, then plot the mean WTP in the 5000 posterior draws. Figure 3.2 shows the histogram of mean WTP when we have $\hat{\beta}_{price_i} = 0.1$ as the criteria to screen respondents in each posterior draw. The mean WTP is 1.7127, with the standard error of 0.2332, which is very close to the result from the first approach that averages each respondent's mean marginal utility of attribute first.

(2) Log-normal Distributed Cost coefficient

Similarly, maximum likelihood estimation result for the mixed logit model with lognormally distributed cost and range coefficients is provided in Table 3.7. The Bayesian posterior means and standard errors of the hyperparameters are listed in Table 3.8. Again, the ML estimators are quite close to the Bayesian results. With the ML estimators, the mean WTP of hybrid is calculated to be 1.1417, the sample standard error of ML estimator from the Delta method is 0.1717, and 0.2353 from the block delete jackknife method.

For the Bayesian approach, the mean WTP of hybrid (compare to gas) calculated from individual-level WTP is listed in Table 3.9. Similarly, we take the criterion of $\hat{\beta}_{price_i} = 0.1$ and drop the respondents whose posterior mean of price coefficient is less than 0.1. But note that no respondent is dropped in this case. The mean of WTP is 2.1147 with the sample standard error of 0.195. From the second approach that average the mean WTP in each posterior draw first, we have the mean WTP 2.2547, with the standard error of 0.2993. Figure 3.3 and 3.4 provide the distribution of mean WTP with the criteria $0.01 < \hat{\beta}_{price_i} < 0.2$ in approach 1, and the histogram of mean WTP under the second approach with $\hat{\beta}_{price_i} = 0.1$, respectively.

Case II: Correlated Random Parameters

(1) Normal Distributed Cost Coefficient

Assume that all the four random parameters are correlated and are normally distributed, the MLE result is provided in Table 3.10, and the Bayesian posterior mean and variance of hyperparameters are listed in Table 3.11. Still, the Bayesian result is quite close to that of ML estimator (for both the mean and standard error), except for the standard errors of parameters posterior variances. This may be caused by the diffuse prior we adopt with large prior variance. With the ML estimator, the mean WTP is estimated to be 1.4158, and the Delta method standard error of mean WTP is 0.1798, which is much smaller than the block delete jackknife standard error of 0.2469.

Mean WTP calculated with individual-level WTP estimators is provided in Table 3.12. With the criteria of $\hat{\beta}_{price_i} = 0.1$ and deleting respondents whose posterior mean of the price parameter is less than 0.1, 92 respondents are retained and the mean of WTP is 1.5605 with the sample standard error of 0.2258. The distribution of mean WTP under different criteria is shown in Figure 3.5. For the second approach that averages the mean WTP within each posterior draw first, we have the mean WTP of 1.6822 and the standard error of 0.2529. The empirical distribution of mean WTP is presented in Figure 3.6.

(2) Log-normal Distributed Cost Coefficient

We also apply the lognormal distribution to both the price and range variables in the case of correlated random coefficients. The MLE and Bayesian results are provided in Table 3.13 and Table 3.14, respectively. It can be found that mean estimator in MLE is very close to that from the Bayesian approach, except for the standard errors of parameters posterior variances. With the ML estimator, the mean WTP of hybrid is 1.1693 (versus gas cars), with the Delta method sample standard error of 0.1779, and the block delete jackknife sample standard error of 0.2346.

Further, the mean WTP based on individual-level WTP is summarized in Table 3.15. Note that with the criteria $\hat{\beta}_{price_i} = 0.1$, no respondent is dropped again, and the mean WTP of all the 100 respondents is 1.6649 with sample standard error of 0.2423. The distribution of mean WTP with criteria $0.01 < \hat{\beta}_{price_i} < 0.2$ is provided in Figure 3.7. The second approach that averages the mean WTP within each posterior draw provides mean WTP of 1.6822, with the standard error of 0.2529 The distribution of mean WTP shown in Figure 3.8.

Discussion

A brief comparison on the estimators of mean WTP from the logit and the mixed logit model are provided in Table 3.16. Take the multinomial logit model estimator as a reasonable approximation of the true value of mean WTP, we find applying the log-normal distribution to the cost coefficient will always underestimate the mean WTP from the maximum likelihood estimation. This is not surprising considering the log-normal distribution has a heavy tail compared to the normal distribution so that a lower mean WTP is derived with a larger denominator. On the other hand, such a characteristic of the log-normal distribution also weakens the performance of the Bayesian individual-level approach since the criteria on $\hat{\beta}_{price_i}$ can hardly screen out any respondent when all respondents' marginal utility of income is increased due to the shape of the distribution. In our empirical results, all the 100 respondents are retained when the cost coefficient is log-normally distributed, but some of them may have $\hat{\beta}_{price_i}$ that is only slightly larger than the criterion so that the inclusion of them results in a larger mean WTP compared to the situation with a normally distributed cost coefficient. It should be noted that, the choice of the screening criteria $\hat{\beta}_{price_i}$ needs to be carefully evaluated. Instead of presenting a single value, a range of criteria and the corresponding mean WTP would be preferred.

Second, the Delta method is found to be too conservative in estimating the standard error of mean WTP. In both the multinomial logit and the mixed logit models, the standard error from the Delta method is lower than that from the block delete jackknife. The Delta method is equivalent to the infinitesimal jackknife, a limiting form of the delete-1 jackknife that ignores the fact that not all choices in the data set are mutually independent. Instead, the choices from the same respondent are also assumed to be perfectly correlated. To this extent, the block delete jackknife method has a much more reasonable assumption on the independence between groups (the questions of the same respondent) instead of questions, and thus provides a closer estimated standard error to that from the basic logit model.

3.5 Robustness Check

Our empirical result shows the potential and merit of applying a normally distributed cost coefficient in the mixed logit model: comparing with the sign-restricted distributions (like the log-normal distribution), normal distribution has the merit of avoiding the skewness from the strong assumption on the shape of cost coefficient. However, since we use the multinomial logit result as the approximation of the true value of mean WTP, one may still be concerned about the accuracy of such a criterion. In the robustness check, we will use a synthetic data set that generated from a known utility function to examine the validity of our conclusion. Instead of estimating the true distribution of mean WTP with the observed data, we can apply the method introduced by Marsaglia (2006) to derive the theoretical approximation of the moments for mean WTP given that the real distribution of both the numerator and the denominator in the WTP ratio is known and actually, is set by us. Then, the theoretical true value of mean WTP can be used to evaluate the performance of the estimators derived from the mixed logit model. We will use the Bayesian approach to run the estimation given its good performance in avoiding unrepresentative respondents.

Assume the utility function is a linear form:

$$U_{ijt} = \beta_{price,i} Price_{ijt} + \beta_{attribute1,i} Attribute1_{ijt} + \beta_{attribute2,i} Attribute2_{ijt} + \varepsilon_{ijt}$$

where $\varepsilon \sim EV(0,1)$ to ensure the logit form of the probability function. To allow the heterogeneity in respondents' preference, we set all the three coefficient $\beta_{price}, \beta_{attribute1}$, and $\beta_{attribute2}$ to be normally distributed. As defined, WTP for attribute 1 is calculated by the ratio $\frac{\beta_{attribute1}}{\beta_{price}}$. Refering to Marsaglia (2006), the ratio of two jointly distributed normal variates z, w can be transformed into the form $\frac{a+x}{b+y}$, where x and y are independent standard normal variates and a, b are non-negative constants. More specifically, for a given ratio $\frac{z}{w}$, there are constants r and s such that

$$r(\frac{z}{w} - s) = r\frac{z - sw}{w} \text{ is distributed as } \frac{a+x}{b+y} \text{ and } \frac{z}{w} \text{ is distributed as } \frac{1}{r}(\frac{a+x}{b+y}) + s$$

where

$$b = \frac{\mu_w}{\sigma_w}; \quad a = \pm\frac{\mu_z/\sigma_z - \rho\mu_w/\sigma_w}{\sqrt{1-\rho^2}}; \quad r = \frac{\sigma_w}{\pm\sigma_z\sqrt{1-\rho^2}} \quad (\text{ } a \text{ and } b \text{ shall have the same sign})$$

Then, two practical rules are developed to ensure the existance of approximating moments for the ratio: (1) if $a < 2.256$ and $4 < b$, the ratio $\frac{a+x}{b+y}$ is approximately normlaly distributed with mean $\mu = a/(1.01b - 0.2713)$ and variance $\sigma^2 = (a^2+1)/(b^2 + 0.108b - 3.795) - \mu^2$. Correspondingly, $\frac{z}{w} \sim N(\frac{1}{r}\mu + s, \frac{1}{r^2}\sigma^2)$; (2) If one can ensure z is always positive, then the ratio z/w approximates to a well defined normal distribution. To simplify the data generating process, we use the rule (1) and assume the distribution of β_p, β_1 and β_2 are mutually independent:

$$\beta_{price} \sim N(3, 0.25); \quad \beta_{attribute1} \sim N(2, 16); \quad \beta_{attribute2} \sim N(2, 1)$$

Then theoretically, the distribution of mean WTP for the attribute 1 could be derived:

$$a = \frac{\mu_{\beta_1}}{\sigma_{\beta_1}} = \frac{2}{4} = 0.5; \quad b = \frac{\mu_{\beta_p}}{\sigma_{\beta_p}} = \frac{3}{0.5} = 6; \quad r = \frac{0.5}{4} = \frac{1}{8}$$

$$\frac{a+x}{b+y} \sim N(a/(1.01b - 0.2713), \quad (a^2+1)/(b^2 + 0.108b - 3.795) - \mu^2) = N(0.086, 0.0305)$$

$$\text{so that} \quad \frac{\beta_{attribute1}}{\beta_{price}} \sim N(\frac{1}{r}0.086, \quad (\frac{1}{r})^2 0.0305) = N(0.691, 1.958)$$

Thus, the distribution of sample mean of WTP is $\hat{\mu} \sim N(0.691, 1.968/n)$ where n is the number of respondent in the sample. If we set n=2000 in our synthetic data set, the true mean WTP is 0.69, the true standard error of mean WTP is 0.0313.

To construct a synthetic data set, a choice experiment survey is first designed based on the fractional factorial method provided by Kuhfeld (2010) to achieve the highest design efficiency. Given the choice task design in Table 3.17, 100% design efficiency can be achieved when total number of runs (all possible combinations of factor levels) n=8 or 16. To ensure the highest design efficiency, we select n=16 and group them into 4 questions (the relative D-efficiency=100). Table 3.18 shows the survey generated by SAS.

Then, a synthetic data set is generated by simulating each respondent's choice based on the fundamental assumption in random utility models: respondent will always choose the alternative that provides the highest utility in each question. With n=2000, we have 2000 respondents' answers simulated. Then, the Bayesian method is used to estimate the utility function under the framework of the mixed logit model. The result is shown in Table 3.19. The estimated mean and variance of the coefficients are very close to the true value we set, and validates the efficiency of our survey design and the accuracy of the Bayesian estimation.

Then, we empirically calculate the distribution of mean WTP with 5000 draws of conditional posteriors of individual-level coefficients. Similar with our analysis before, one approach is to average each respondent's $\beta_{attribute1}$ and β_{price} through the 5000 posterior draws first to have the mean WTP for all the 2000 respondents. Then the distribution of mean WTP can be calculated. In this approach, the minimum $\hat{\beta}_{price,i}$ is 1.94, which is far from 0 so that we don't need to drop any respondent. The mean WTP is calculated to be 0.6786, and the estimated standard error of mean WTP is 0.0256. The result is very close to the theoretical true distribution we derived. The other approach is to calculate the mean WTP in each of the 5000 posterior draws first, and then describe the distribution of the 5000 mean WTP. The 95% credible interval of mean WTP is shown in Figure 3.9, with the mean WTP of 0.6955, and stand error of mean WTP of 0.0285, almost the same with the first approach and very close to the theoretical true value.

Note that, in Figure 3.9, although almost all the density of mean WTP concentrates on the normal shape in the left-hand side, we still observe a very light spike in the region around mean WTP = 1.33. This is obviously caused by a few draws whose β_{price} is very small and definitely out of the 95% confidence interval. Since one would expect the mean WTP to capture the preference for the most representative population, instead of the most wealthy, we think it is reasonable to conclude the distribution of mean WTP has been very well captured by the normal distribution on the left hand-side: the small spike on the right-hand side is negligible and won't influence the approximation of the normality in mean WTP.

Thus, the robustness check further validates the feasibility of a normally distributed cost coefficient in the mixed logit model. Although the problem of non-existence of moments may still exist in the view of statistics, we found both theoretical and empirical evidence that the approximation distribution of the ratio performs well and follows a normal distribution. The potential small spike should not influence the estimation of mean WTP as long as we can ensure the coefficients satisfy either one of the two rules we defined before. Intuitively, if the marginally utility has a mean far from zero and have a small standard error, or if we can ensure it is always positive in an economic sense (which holds for normal good if consumption is not zero), we can expect a stable performance of mean WTP when a normal distribution is applied to the cost coefficient.

3.6 Conclusion

This paper explores the estimation of mean Willingness to Pay (WTP) in the mixed logit model under various distributional assumptions and calculation approaches. Different than most former studies which mainly focus on evaluating the statistical features of WTP, we advocate carefully examining the performance of the WTP ratio with a full consideration of the underneath economic assumptions. Our research with both empirical and synthetic data sets shows the potential and merit of applying a normally distributed cost coefficient in the WTP ratio. We also validate the good performance of the Bayesian approach and the Block delete jackknife approach in capturing the distribution of mean WTP.

By narrowing down the scope to the normal goods and ruling out the possibility of zero consumption for all survey questions, we first validate the rationality of a positive marginal utility of income in an economic sense – the condition to ensure stable moment approximations of the WTP as a ratio of two normal variables. Further, we apply both the normal and log-normal distributions to the cost coefficient in our estimation of the mixed logit model. With an empirical survey data on consumers' choices of hybrid vehicles, we examine the performance of three methods in estimating the distribution of mean WTP – the most important statistic. Also, we consider that it should be the standard error of sample mean that used to construct the confidence interval of mean WTP, instead of the standard error of the sample.

The empirical analysis first shows the limitation of the Delta method in determining the standard error of mean WTP given it is equivalent to infinitesimal jackknife, a limiting form of the delete-1 jackknife, which assumes the mutual independence between all observations. We argue that the block delete jackknife has a more reasonable assumption of independence among respondents rather than among questions (observations). The empirical result validates our expectation and underscores the difference between these two methods: the Delta method tends to underestimate the standard error of mean WTP.

Further, our results reveal the skewness caused by applying the log-normal distribution to the cost coefficient. Comparing the result from the basic multinomial logit model that we treat as a reasonable approximation of the true value of mean WTP, the log-normal distribution leads to a larger estimator of the price coefficient and thus lowers the value of mean WTP in maximum likelihood estimation. Although the sign-restricted distributions like the log-normal distribution can avoid negative or zero cost coefficient in statistical theory, the strong assumption on the shape of marginal utility of income may not fit the empirical data well and thus should be carefully considered.

Finally, we validate the performance of the Bayesian individual-level WTP approach in capturing the distribution of mean WTP. From the perspective of individual preference, it is reasonable to set a criterion to drop respondents or posterior draws for which a negligible or zero marginal utility of income is placed. Depending on the criteria assumed, a series of distributions of mean WTP could be obtained, instead of just one distribution. We suggest to select a range of criteria where the standard error of mean WTP tends to be stable in stead of just selecting one value.

In addition to the empirical analysis with a real survey data set, a synthetic data set is simulated to further examine the feasibility of a normally distributed cost coefficient with the Bayesian individual-level WTP approach. We find that, if the marginal utility of income has a mean far from zero and has relative a small standard error, or if we can ensure the marginal utility of income is always positive in economic sense (which holds for normal good if the consumption is not all zero), we can expect a stable performance of the mean WTP when a normally distributed cost coefficient is applied. In our analysis with the synthetic data set, the estimated mean WTP from the mixed logit model is very close to the theoretical value we calculated with the practical rules developed by Marsaglia (2006). Although we do observe a very slight spike that highly deviates from the majority density of mean WTP, it is economically reasonable to ignore such small spike given the mean WTP should present a general picture of the preference for the population, rather than the preference of the small group of perhaps wealthy respondents. On the other hand, for most empirical survey data, such a spike would be more likely due to respondents' inattention to money: they do not face with a real budget constraint given it is only a hypothetical choice instead of a real decision in their life. In this sense, we should put even less concern on such a spike.

References

Balcombe, K., Chalak, A., and Fraser, I. (2009). Model selection for the mixed logit with bayesian estimation. *Journal of Environmental Economics and Management*, 57(2):226–237.

Bliemer, M. C. and Rose, J. M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological*, 58:199–214.

Box, G. E. (1958). A note on the generation of random normal deviates. *Ann. Math. Stat.*, 29:610–611.

Breffle, W. S., Morey, E., and Waldman, D. (2005). Gaussian quadrature versus simulation for the estimation of random parameters. In *Applications of Simulation Methods in Environmental and Resource Economics*, pages 341–353. Springer.

Cedilnik, A., Kosmelj, K., and Blejec, A. (2004). The distribution of the ratio of jointly normal variables. *Metodoloski zvezki*, 1(1):99.

Cirillo, C. and Hetrakul, P. (2010). Continuous random coefficient logit models: a comparison of parametric and non-parametric methods to estimate individual preferences over cybernetic transportation systems. In *presentation at the 89th Annual Meeting of the Transportation Research Board, Washington, DC*.

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1):19–31.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, 68(3):589–599.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, 24(3/4):428–440.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93(3):442–446.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700.

Hahn, J., Hausman, J. A., and Lustig, J. (2017). Specification test on mixed logit models. Technical report, cemmap working paper.

Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30(2):133–176.

Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, 39(2-3):221–236.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika*, 56(3):635–639.

Hole, A. R. and Kolstad, J. R. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and wtp space using data from a health-related choice experiment. *Empirical Economics*, 42(2):445–469.

Kuhfeld, W. F. (2010). *Marketing Research Methods in SAS.* Citeseer.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, 60(309):193–204.

Marsaglia, G. (2006). Ratios of normal variables. *Journal of Statistical Software*, 16(4):1–10.

Meijer, E. and Rouwendal, J. (2006). Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics*, 21(2):227–244.

Rigby, D., Balcombe, K., and Burton, M. (2009). Mixed logit model performance and distributional assumptions: preferences and gm foods. *Environmental and Resource Economics*, 42(3):279–295.

Scarpa, R. and Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, 52(3):253–282.

Severini, T. A. (2005). *Elements of distribution theory*, volume 17. Cambridge University Press.

Shao, J. and Wu, C. J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics*, 17(3):1176–1197.

Sillano, M. and de Dios Ortúzar, J. (2005). Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning A*, 37(3):525–550.

Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, pages 105–130.

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In *Applications of simulation methods in environmental and resource economics*, pages 117–134. Springer.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Table 3.1: Sample of the Vehicle Data Set

Question	**ID**	**choice**	**price**	**Operating cost**	**Range**	**Electric**	**Gas**	**Hybrid**	**Perf1**	**Perf2**
1	1	0	4.676	47.43	0	0	0	1	0	0
	1	1	5.721	27.43	1.3	1	0	0	1	1
	1	0	8.796	32.41	1.2	1	0	0	0	1
2	1	1	3.377	4.89	1.3	1	0	0	1	1
	1	0	9.034	30.19	0	0	0	1	0	1
	1	0	5.71	27.16	1.8	1	0	0	1	1

Table 3.2: Summary Statistics of the Vehicle Data Set

Continuous Variables						
	Variable	N	Mean	Std Dev	Minimum	Maximum
	Price (in $10,000)	3000	3.57	1.75	0.70	9.72
	Range (in 100 miles)	3000	0.42	0.65	0.00	2.00
	Operating cost (in $)	3000	33.23	15.58	2.59	72.29
Categorical Variables						
	Variable	Level			Frequency	Percent (%)
	Engine Type					
		Electric			974	32.47
		Gas			1008	33.60
		Hybrid			1018	33.93
	Perfomance					
		High Perf			990	33.00
		Middle and low Perf			2010	67.00

Table 3.3: MLE Estimation of the Conditional Logit Model

Parameter	**Estimate**	**Std.Error_Robust**	**Est/SER**
Price	0.4263***	0.0444	10.278
Range	0.7033***	0.2214	3.231
Electric	-1.3285***	0.3163	-4.18
hybrid	0.5965***	0.123	4.983
Perf1	0.4634***	0.0911	5.121

* Log likelihood at convergence: -946.3609
* Note we use negative of the price variable in the model

Table 3.4: MLE with High Dimensioned Gauss-Hermite Quadrature of the Mixed Logit Model (Case I and normal cost coefficient)

Parameter	Estimate	Std. Error	Est/SER
Price	0.6454	0.082	7.868
Range	0.6661	0.3004	2.218
Electric	-1.4472	0.426	-3.397
hybrid	0.8658	0.1653	5.236
Perf1	0.6248	0.1062	5.882
SE(Price)	0.5253	0.0798	6.582
SE(Range)	0.8648	0.2543	3.401
SE(Electric)	0.7396	0.5386	1.373
SE(Hybrid)	0.8044	0.1931	4.166

* Log likelihood at convergence: -898.9314

* Note we use negative of the price variable in the model

Table 3.5: Bayesian Posterior Mean and Variance of the Mixed Logit Model (Case I and normal cost coefficient)

Parameter	N	Estimation/Mean	SE/StDv	Est/SE
Price	5000	0.6495	0.0851	7.6322
Range	5000	0.7761	0.2804	2.7678
Electric	5000	-1.6331	0.3945	-4.1397
hybrid	5000	0.8695	0.1644	5.2889
Perf1	5000	0.4428	0.1035	4.2783
Var(Price)	5000	0.3134	0.0868	3.6106
Var(Range)	5000	0.7438	0.335	2.2203
Var(Electric)	5000	0.8305	0.5487	1.5136
Var(Hybrid)	5000	0.6938	0.2997	2.315

Table 3.6: Mean WTP from the Bayesian Individual-level WTP (Case I and normal cost coefficient)

Criteria of $\hat{\beta}_{price}$	Number of remained respondents	Mean WTP	Std(mean WTP)
0.01	100	1.9652	0.7497
0.02	99	2.3101	0.6724
0.03	98	1.7794	0.4173
0.04	98	1.7794	0.4173
0.05	97	1.5308	0.3386
0.06	97	1.5308	0.3386
0.07	95	1.765	0.2316
0.08	94	1.7542	0.2338
0.09	93	1.717	0.2333
0.1	93	1.717	0.2333
0.11	93	1.717	0.2333
0.12	92	1.6258	0.2171
0.13	92	1.6258	0.2171
0.14	91	1.5912	0.2167
0.15	90	1.7308	0.1676

Table 3.7: MLE with High Dimensioned Gauss-Hermite Quadrature of the Mixed Logit Model (Case I and lognormal cost coefficient)

Parameter	**Estimate**	**Std. Error**	**Est/SER**
Price	-0.7845	0.173	-4.535
Range	-0.3096	0.3773	-0.82
Electric	-1.6714	0.3927	-4.256
hybrid	0.898	0.164	5.476
Perf1	0.6286	0.1059	5.937
SE(Price)	1.0434	0.1583	6.59
SE(Range)	0.5815	0.2635	2.207
SE(Electric)	1.1173	0.2977	3.753
SE(Hybrid)	0.7722	0.1962	3.937

* Log likelihood at convergence: -896.1257

* Note we use negative of the price variable in the model

Table 3.8: Bayesian Posterior Mean and Variance of the Mixed Logit Model (Case I and lognormal cost coefficient)

Parameter	**N**	**Est/Mean**	**SE/StDv**	**Est/SE**
Price	5000	-0.8397	0.1807	-4.6469
Range	5000	-0.4999	0.4293	-1.1645
Electric	5000	-1.6365	0.3781	-4.3282
hybrid	5000	0.8686	0.1659	5.2357
Perf1	5000	0.4285	0.1017	4.2134
Var(Price)	5000	1.1734	0.3708	3.1645
Var(Range)	5000	0.5896	0.4289	1.3747
Var(Electric)	5000	1.1482	0.5908	1.9435
Var(Hybrid)	5000	0.6351	0.2825	2.2481

Table 3.9: Mean WTP from the Bayesian Individual-level WTP (Case I and lognormal cost coefficient)

Criteria of $\hat{\beta}_{price}$	Number of remained respondents	Mean WTP	Std(mean WTP)
0.01	100	2.1147	0.195
0.02	100	2.1147	0.195
0.03	100	2.1147	0.195
0.04	100	2.1147	0.195
0.05	100	2.1147	0.195
0.06	100	2.1147	0.195
0.07	100	2.1147	0.195
0.08	100	2.1147	0.195
0.09	100	2.1147	0.195
0.1	100	2.1147	0.195
0.11	100	2.1147	0.195
0.12	100	2.1147	0.195
0.13	100	2.1147	0.195
0.14	99	2.0799	0.1938
0.15	99	2.0799	0.1938

Table 3.10: MLE with High Dimensioned Gauss-Hermite Quadrature of the Mixed Logit Model (Case II and normal cost coefficient)

Parameter	Estimate	Std. Error	Est/SER
Price	0.698	0.0957	7.296
Range	0.8578	0.3601	2.382
Electric	-1.6556	0.5182	-3.195
hybrid	0.9881	0.2146	4.605
Perf1	0.6543	0.1097	5.963
SE(Price)	0.6527	0.0998	6.541
SE(Range)	1.3841	0.453	3.056
SE(Electric)	1.9382	0.7004	2.767
SE(Hybrid)	1.4416	0.2382	6.052
Corr(Price, Range)	0.4012	0.3006	1.335
Corr(Price, Electric)	-0.0888	0.3217	-0.276
Corr(price, Hybrid)	0.5205	0.1418	3.672
Corr(Range, Electric)	-0.4948	0.2956	-1.674
Corr(Range, Hybrid)	0.2334	0.3382	0.69
Corr(Electric, Hybrid)	0.5379	0.2754	1.953

* Log likelihood at convergence: -886.5121

* Note we use negative of the price variable in the model

Table 3.11: Bayesian Posterior Mean and Variance of the Mixed Logit Model (Case II and normal cost coefficient)

Parameter	N	Est	SE	Est/SE
Price	5000	0.7219	0.1002	7.2046
Range	5000	0.8047	0.3323	2.4216
Electric	5000	-1.6097	0.4628	-3.4782
hybrid	5000	1.0066	0.2134	4.717
Perf1	5000	0.4776	0.1099	4.3458
Var(Price)	5000	0.5383	0.1382	3.8951
Var(Range)	5000	1.6826	0.9147	1.8395
Var(Electric)	5000	2.1601	1.411	1.5309
Var(Hybrid)	5000	2.0996	0.68	3.0876
Variance-Covariance	**Price**	**Range**	**Electric**	**hybrid**
Price	0.5383	0.2279	0.0724	0.4766
Range	0.2279	1.6826	-0.4521	0.838
Electric	0.0724	-0.4521	2.1601	0.8442
hybrid	0.4766	0.838	0.8442	2.0996

Table 3.12: Mean WTP from the Bayesian Individual-level WTP (Case II and normal cost coefficient)

Criteria of $\hat{\beta}_{price}$	Number of remained respondents	Mean WTP	Std(mean WTP)
0.01	100	0.0772	1.8679
0.02	98	1.8299	0.7002
0.03	96	2.1938	0.6242
0.04	95	1.6265	0.2633
0.05	94	1.6217	0.266
0.06	94	1.6217	0.266
0.07	94	1.6217	0.266
0.08	94	1.6217	0.266
0.09	92	1.5605	0.2258
0.1	92	1.5605	0.2258
0.11	89	1.7314	0.179
0.12	88	1.7759	0.1753
0.13	88	1.7759	0.1753
0.14	88	1.7759	0.1753
0.15	88	1.7759	0.1753

Table 3.13: MLE with High Dimensioned Gauss-Hermite Quadrature of the Mixed Logit Model (Case II and lognormal cost coefficient)

Parameter	**Estimate**	**Std. Error**	**Est/SER**
Price	-0.7871	0.1952	-4.032
Range	-0.3012	0.3849	-0.783
Electric	-1.7535	0.4396	-3.989
hybrid	1.1276	0.226	4.989
Perf1	0.6319	0.1082	5.839
SE(Price)	1.2255	0.1961	6.25
SE (Range)	0.8553	0.2453	3.487
SE (Electric)	1.2363	0.472	2.62
SE (Hybrid)	1.4204	0.2513	5.652
Corr(Price, Range)	0.669	0.3087	2.167
Corr (Price, Electric)	-0.243	0.4391	-0.553
Corr (price, Hybrid)	0.6425	0.154	4.173
Corr (Range, Electric)	0.0973	0.5991	0.162
Corr (Range, Hybrid)	0.8056	0.4055	1.986
Corr (Electric, Hybrid)	0.3661	0.3312	1.105

* Log likelihood at convergence: -884.1523
* Note we use negative of the price variable in the model

Table 3.14: Bayesian Posterior Mean and Variance of the Mixed Logit Model (Case II and lognormal cost coefficient)

Parameter	**N**	**Est/Mean**	**SE/StDv**	**Est/SE**
Price	5000	-0.8505	0.1947	-4.36826
Range	5000	-0.8305	0.523	-1.58795
Electric	5000	-1.4722	0.4131	-3.56379
hybrid	5000	1.0476	0.2138	4.899906
Perf1	5000	0.4381	0.1061	4.129123
Var(Price)	5000	1.6132	0.4919	3.279528
Var(Range)	5000	1.2586	0.6834	1.841674
Var(Electric)	5000	1.7391	0.8575	2.028105
Var(Hybrid)	5000	1.8643	0.6609	2.82085
Variance-Covariance	**Price**	**Range**	**Electric**	**Hybrid**
Price	1.6132	0.4244	0.0097	0.9451
Range	0.4244	1.2586	0.0327	0.7175
Electric	0.0097	0.0327	1.7391	0.7448
Hybrid	0.9451	0.7175	0.7448	1.8643

Table 3.15: Mean WTP from the Bayesian Individual-level WTP (Case II and lognormal cost coefficient)

Criteria of $\hat{\beta}_{price}$	Number of remained respondents	Mean WTP	Std(mean WTP)
0.01	100	1.6649	0.2423
0.02	100	1.6649	0.2423
0.03	100	1.6649	0.2423
0.04	100	1.6649	0.2423
0.05	100	1.6649	0.2423
0.06	100	1.6649	0.2423
0.07	100	1.6649	0.2423
0.08	100	1.6649	0.2423
0.09	100	1.6649	0.2423
0.1	100	1.6649	0.2423
0.11	99	1.7583	0.2259
0.12	98	1.7397	0.2274
0.13	98	1.7397	0.2274
0.14	98	1.7397	0.2274
0.15	97	1.7611	0.2288

Table 3.16: Summary of the Mean WTP for Hybrid Car

Model	Case	Distribution of the Cost Coefficient	Maximum Likelihood Estimation			Hierarchical Bayes							
				Std (Mean WTP)		Approach 1				Approach 2			
			E (Mean WTP)	Delta method	Block Delete Jackknife	Criterion β_{price}	n of respondent remained	E (Mean WTP)	Std (Mean WTP)	Criterion $beta_{price}$	n of Posterior Draws	E (Mean WTP)	Std (Mean WTP)
Multinomial Logit	/	/	1.399	0.2274	0.318	/	/	/	/	/	/	/	/
Mixed logit	Uncorrelated Random Parameters	Normal	1.3414	0.1709	0.256	0.1	93	1.717	0.2333	0.1	5000	1.7127	0.2332
		Lognormal	1.1417	0.1717	0.2353	0.1	100	2.1147	0.195	0.1	5000	2.2547	0.2993
	Correlated Random Parameters	Normal	1.4158	0.1798	0.2469	0.1	92	1.5605	0.2258	0.1	5000	1.6822	0.2529
		Lognormal	1.1693	0.1779	0.2346	0.1	100	1.6649	0.2423	0.1	5000	2.0085	0.3465

Table 3.17: Product Attributes in the Synthetic Data Set

Factor	Possible values of independent variable
Attribute 1	$\{0, 1\}$
Attribute 2	$\{0, 1\}$
Price	$\{1.99, 2.99, 3.99, 4.99\}$

Table 3.18: Survey Designed for the Synthetic Data Set

Question	**Alternative**	**Attribute1**	**Attribute2**	**Price**
1	1	0	1	3.99
	2	1	1	1.99
	3	1	0	4.99
	4	0	0	2.99
2	1	0	0	4.99
	2	0	1	2.99
	3	1	1	1.99
	4	1	0	3.99
3	1	0	1	3.99
	2	1	1	4.99
	3	0	0	1.99
	4	1	0	2.99
4	1	0	0	1.99
	2	1	0	3.99
	3	1	1	2.99
	4	0	1	4.99

Table 3.19: Bayesian Posterior Mean and Variance of the Mixed Logit Model (Synthetic Data with Uncorrelated and Normal Coefficients)

Parameter	**N**	**Estimate**	**Std. Error**	**Est/SER**
Attribute1	5000	2.0167	0.1289	15.6455
Attribute2	5000	2.0238	0.1101	18.3815
Price	5000	2.9901	0.0913	32.7503
Var(attribute1)	5000	15.0529	1.3255	11.3564
Var(attribute2)	5000	0.7884	0.2873	2.7442
Var(price)	5000	0.347	0.0843	4.1163
Variance-Covariance		**Attribute1**	**Attribute2**	**Price**
Attribute1		15.0529	-0.1702	0.0711
Attribute2		-0.1702	0.7884	0.0549
Price		0.0711	0.0549	0.347

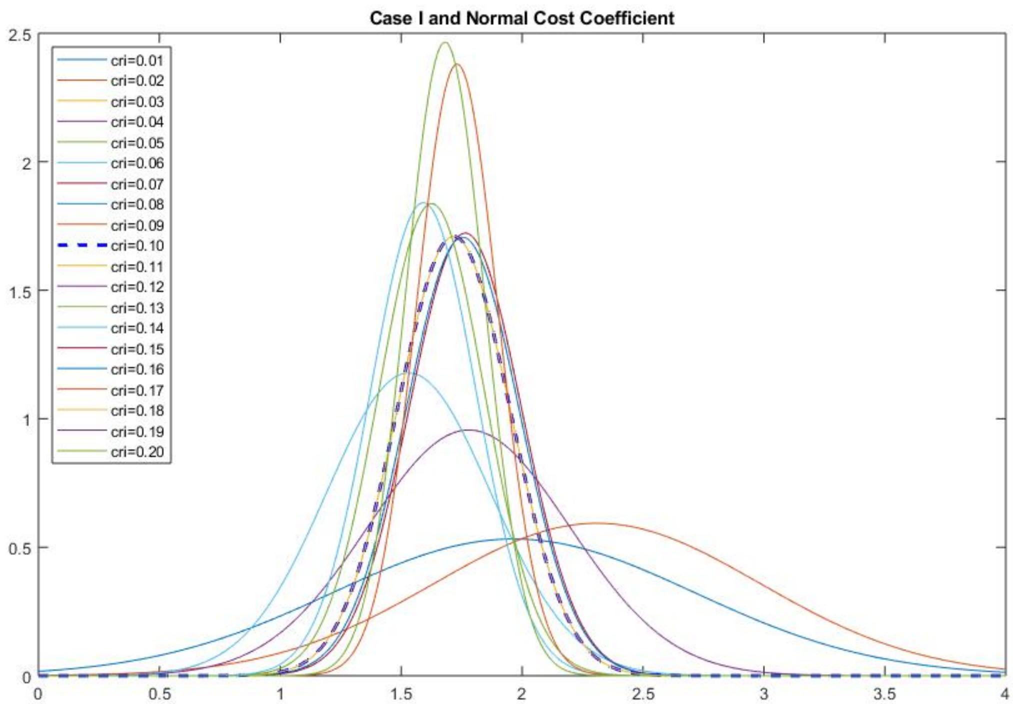


Figure 3.1: Distribution of Mean WTP with the Criterion
$0.01 < \hat{\beta}_{price_i} < 0.20$
(Uncorrelated Random Coefficients & Normal Cost Coefficient)

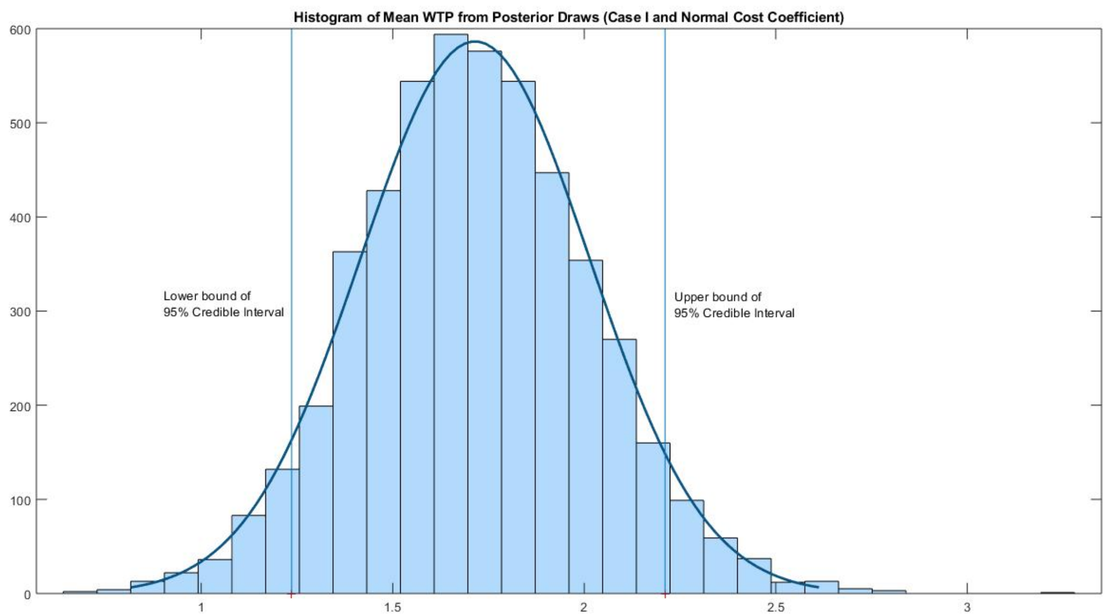


Figure 3.2: Distribution of Mean WTP with $\hat{\beta}_{price_i} = 0.1$
(Uncorrelated Random Coefficients & Normal Cost Coefficient)

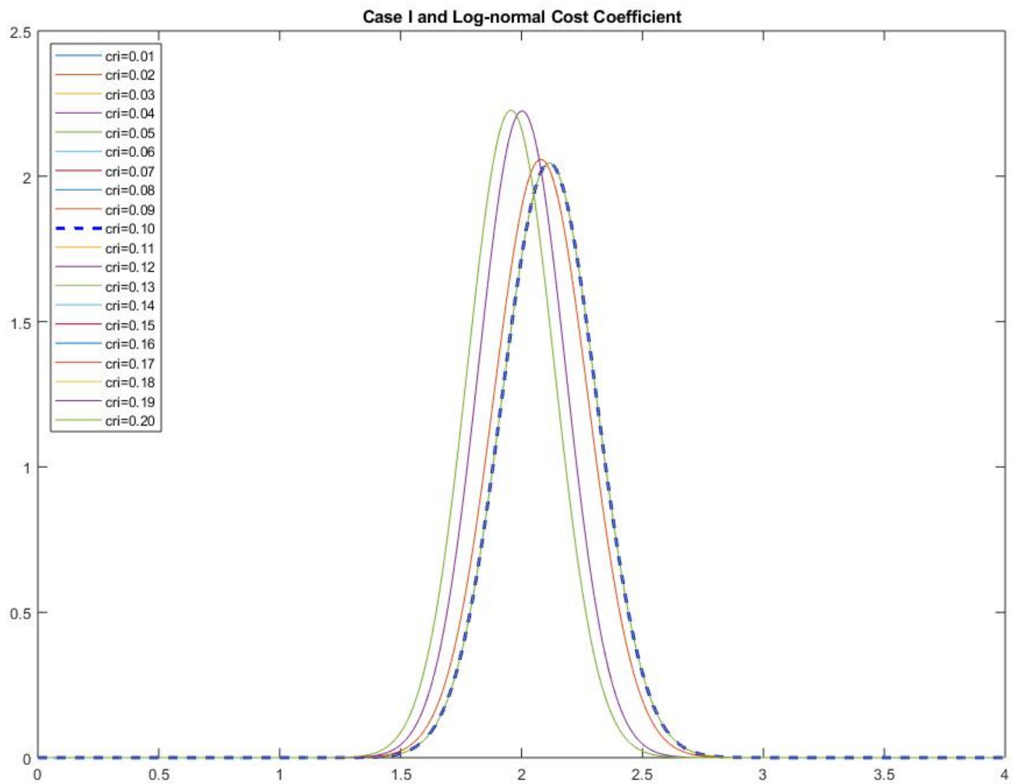


Figure 3.3: Distribution of Mean WTP with the Criterion
$0.01 < \hat{\beta}_{price_i} < 0.20$
(Uncorrelated Random Coefficients & Log-normal Cost Coefficient)

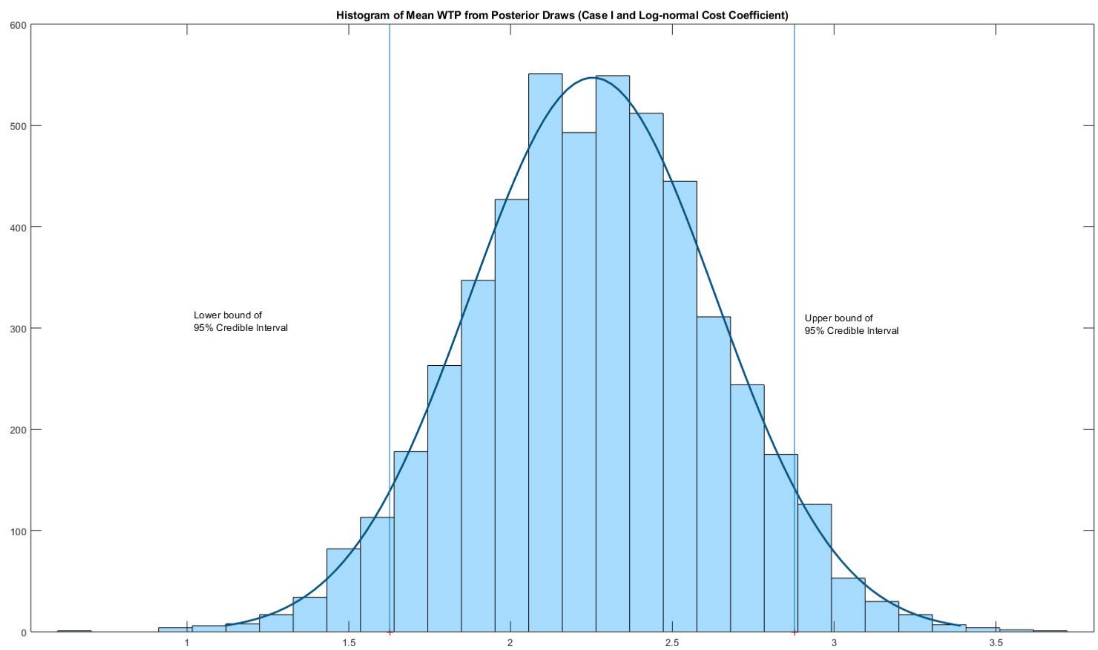


Figure 3.4: Distribution of Mean WTP with $\hat{\beta}_{price_i} = 0.1$
(Uncorrelated Random Coefficients & Log-normal Cost Coefficient)

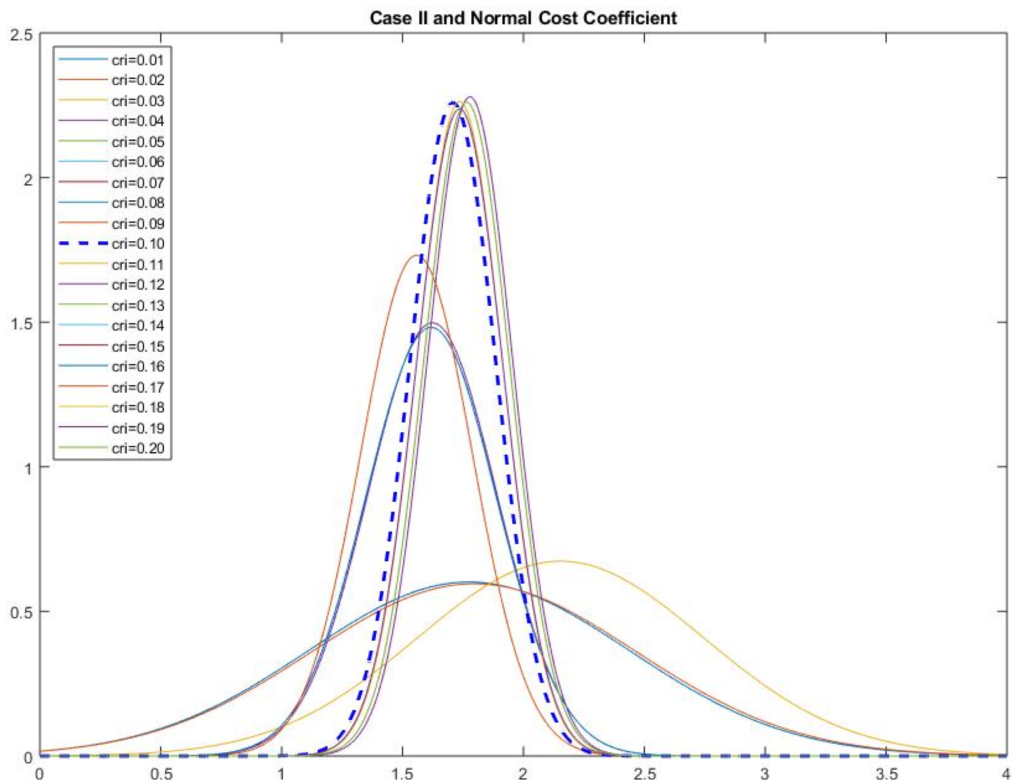


Figure 3.5: Distribution of Mean WTP with the Criterion
$0.01 < \hat{\beta}_{price_i} < 0.20$
(Correlated Random Coefficients & Normal Cost Coefficient)

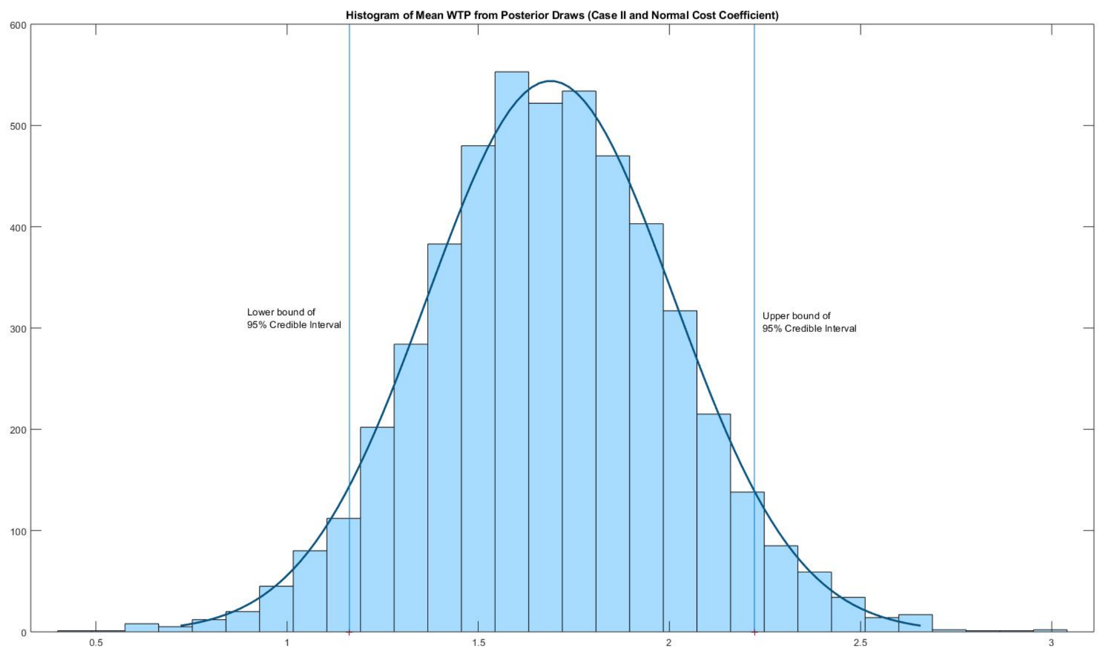


Figure 3.6: Distribution of Mean WTP with $\hat{\beta}_{price_i} = 0.1$
(Correlated Random Coefficients & Normal Cost Coefficient)

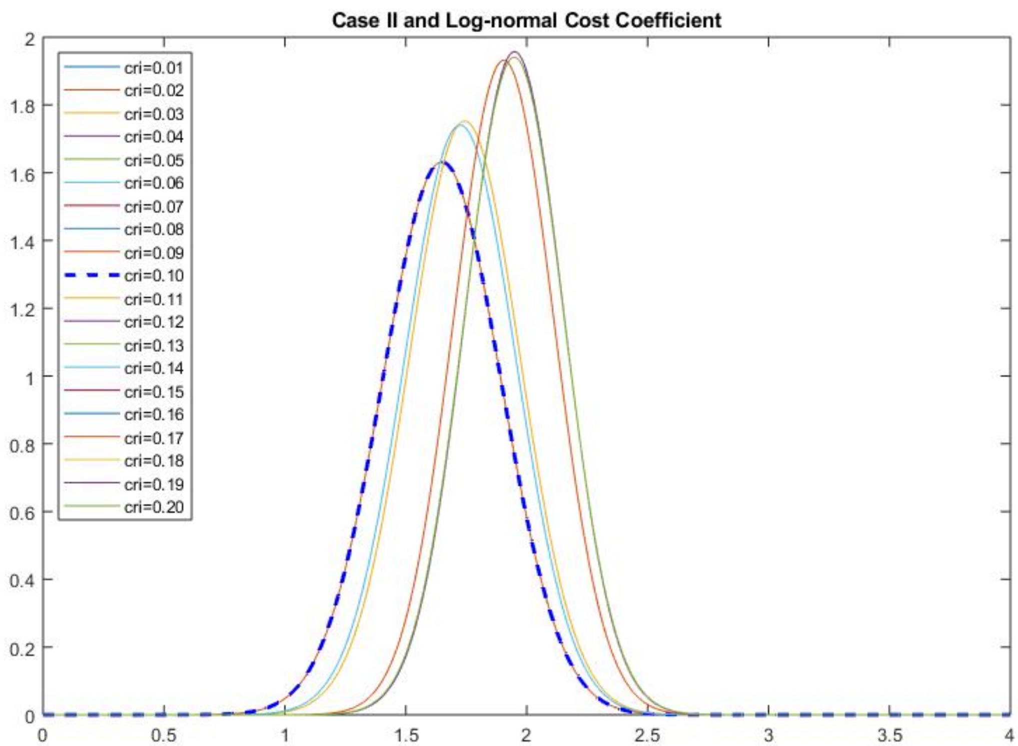


Figure 3.7: Distribution of Mean WTP with the Criterion
$0.01 < \hat{\beta}_{price_i} < 0.20$
(Correlated Random Coefficients & Log-normal Cost Coefficient)

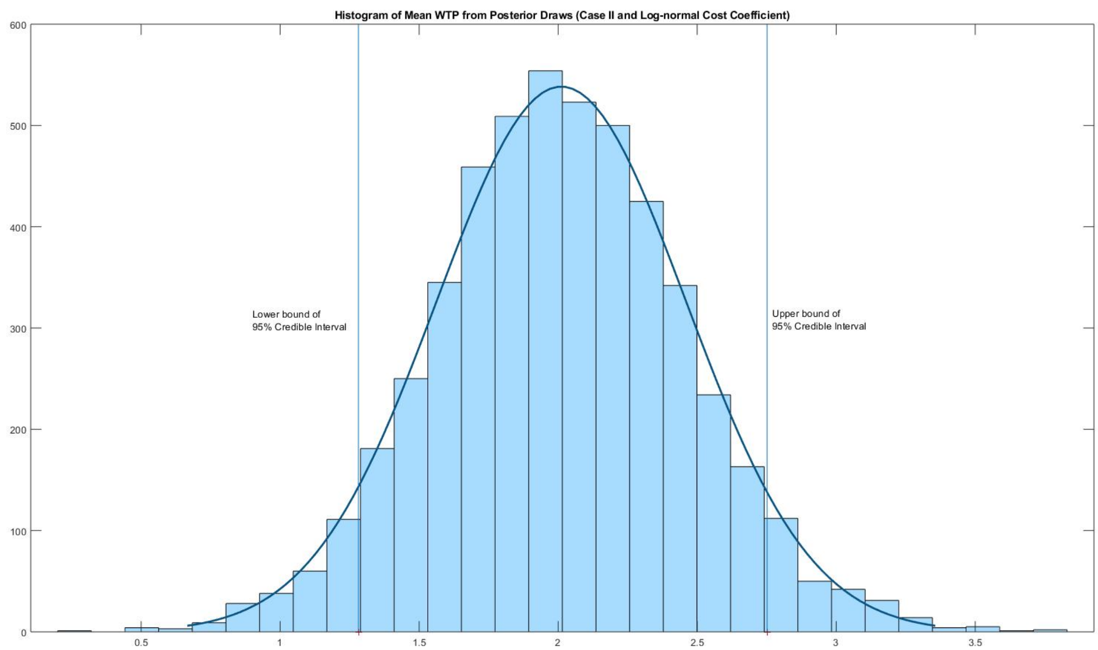


Figure 3.8: Distribution of Mean WTP with $\hat{\beta}_{price_i} = 0.1$
(Correlated Random Coefficients & Log-normal Cost Coefficient)

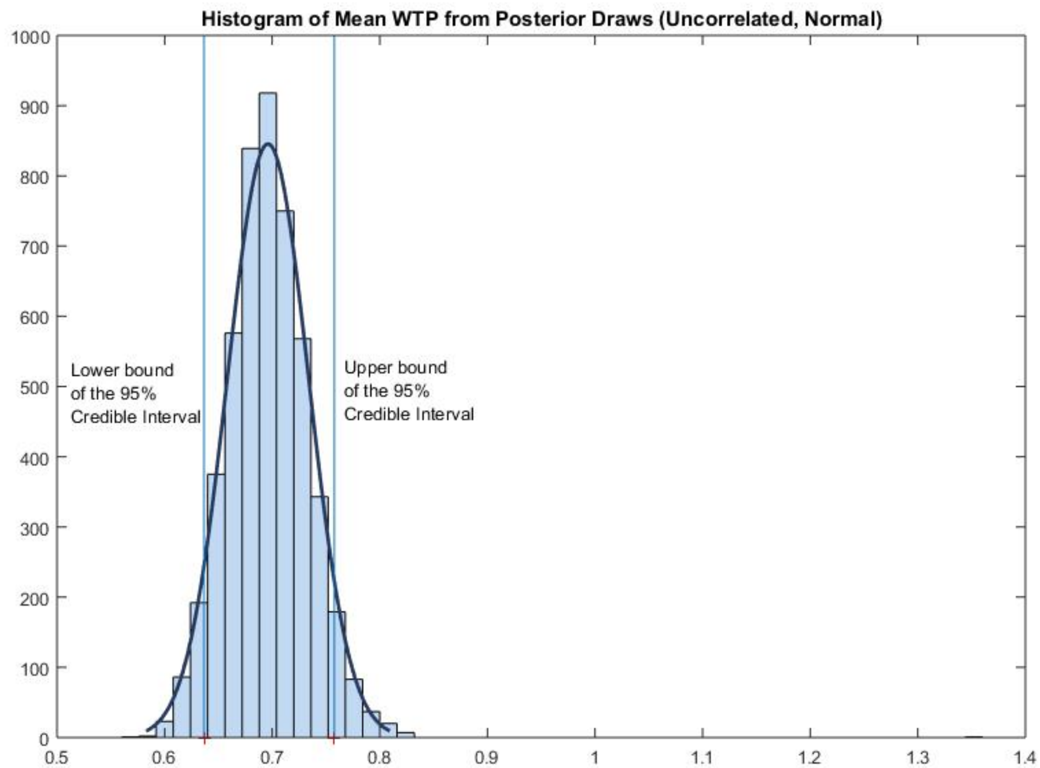


Figure 3.9: Distribution of Mean WTP from the Synthetic Data set
(Uncorrelated Random Coefficients & Normal Cost Coefficient)

Chapter 4

Estimation of Brand Choice Using a Multivariate Poisson-log Normal Incomplete Demand System

* Ying J. and Shonkwiler J. S.. To be submitted to Journal of Agricultural and Applied Economics.

Abstract

This paper provides a new approach which recognizes that quantities purchased are discrete and over-dispersed and demands may be correlated. As a consequence we specify a multivariate Poisson-log normal distribution. To represent preferences in an utility-consistent way, we consider the incomplete demand system specification as it allows log-linear and semi-log forms – a necessary requirement given the exponential link in the Poisson-log normal count data model. With the assumption that the prices of all other goods outside the system are quasi-fixed, unconditional price elasticities and income elasticities can be computed from the properly specified incomplete demand system. Using a facial tissue consumption data of 1927 households in Eau Claire, Wisconsin in 2011 from the IRI Marketing Data Set, we analyze amounts purchased of four major brands and provide insights on market structure and consumer preferences.

Keywords: Multivariate Poisson-log Normal Distribution, Incomplete Demand System, Over-dispersion, Scanner Data, Count Data Model

4.1 Introduction

Consumer brand choice is of keen interest for academics and practitioners to understand consumer preferences and market structures The typical information studied by most existing demand analyses has concentrated on the estimation of elasticities and the prediction of market shares using demand models in both the product space (Stone, 1954; Theil, 1965; Barten and Turnovsky, 1966; Christensen et al., 1975; Deaton and Muellbauer, 1980, etc.) and the characteristics space (Berry et al., 1995; Berry and Pakes, 2002, etc.). However, for some important economic variables in the form of non-negative integers such as recreational demand, health care demand, consumption of special goods like cigarettes, estimation of household labor supplies, and analysis of market entry decisions among oligopolistic firms, inference on the discrete quantity demanded – the count outcomes – is quite valuable.

Unlike the mature demand models constructed with market shares, demand systems with count outcomes are much less developed. For count data models, the multivariate framework is much less discussed compared to the univariate one. In terms of univariate count data, in-depth studies have shed light on problems like the large portion of zero observations with inflated-zero models and hurdle models based on distributions like Poisson or negative binomial (Mullahy, 1986; Hellerstein and Mendelsohn, 1993; Englin and Shonkwiler, 1995; Winkelmann, 2000; Hidayat and Pokhrel, 2010; Hidayat and Pokhrel, 2010; Bach et al., 2018). In comparison, only a small portion studies have investigated correlated count models that are applicable for cases like multiple related counted outcomes (like brand choices) or a panel of individual choices over time.

So far, the major method to model multivariate count outcomes is a "mixing approach" that allows flexible correlations among different counts by introducing randomness to the parametrization of the mean of multivariate count distributions such as the multivariate Poisson distribution. Aitchison and Ho (1989) first suggested the mathematical form of the multivariate Poisson-Log normal distribution (MPLN), and later Shonkwiler (1995) noted that "...only the MPLN distribution can both reproduce any arbitrary correlation structure and account for overdispersion". After that, Chib and Winkelmann (2001) developed the Bayesian approach with Markov Chain Monte Carlo to summarize the posterior distribution of the parameters and latent effects in the MPLN count data model. Further Egan and Herriges (2006) extended the MPLN model to control for on-site sampling biases in survey data; Haque et al. (2010) incorporated Bayesian inference and suggested the Hierarchical Poisson-log Normal and Hierarchical Poisson Gamma model to study the number of motorcycle crashes. As noted, the *advantage* of this "mixing structure" model (like MPLN) is the unrestricted correlation structure of the counts – the dependency could be either positive or negative – in contrast of other restrictive forms like the seemingly uncorrelated negative binomial (SUNB) models in which only positive correlations are permitted among the counts (Winkelmann, 2000; Chib and Winkelmann, 2001; Egan and Herriges, 2006; Bach et al., 2018; etc.). However, criticisms of this approach also arise from the cumbersome calculation of the multidimensional integration in the likelihood function caused by the mixing components. Existing approaches, either maximum simulated likelihood or Bayesian estimation, are found to be less satisfactory given (1) possible flat areas are likely to occur in the likelihood function when the number of free parameters increases so that the maximum simulated likelihood could perform poorly, and (2) multiple and time-consuming Metropolis-hasting steps needed for the Bayesian estimation due to the lack of standard forms of conditional posterior distributions, respectively.

Beside MPLN, another counterpart approach that developed in recent years is the "ordered-response model" that maps the underlying continuous latent variables with a set of count outcomes through a multivariate normal distribution and a bunch of thresholds. Different from the mixing approach mentioned above, the ordered-response model first estimates the conditional means of a vector of latent variables through a continuous distribution (usually, the multivariate normal distribution), and then maps the estimators back into discrete outcomes through a series of estimated cutoff points. As noted by Meyer (1998), efficiency losses due to the estimation of even a large number of thresholds in the ordered-response model is small. Given this, applications like Scott and Kanaroglou (2002) and Bhat and Srinivasan (2005) have extended the ordered-response model from a trivariate case to a modeling system that can accommodate large numbers of count outcomes. The merit of this approach is the flexible underlying probability function for the latent variables that can avoid the concern of excess zeros in count outcomes. Also, the correlations among count outcomes are flexible in general and the Bayesian method could be simplified if linear conditional means are used (Herriges et al., 2008). However, econometric challenges still exist in the estimation of the ordered-response system due to the convergence problem and imprecision in estimates. For example, the Bayesian estimation approach proposed in this framework is criticized to be "cumbersome, requires extensive simulation and is time-consuming – the convergence assessment becomes very difficult as the number of dimension increase" (Ferdous et al., 2010). As an alternative, Ferdous et al. (2010) proposed the composite marginal likelihood (CML) as an alternative estimation method.

A crucial problem of the existing multivariate count approaches, as pointed in Bhat et al. (2015), is that the models are lacking of a consistent underlying framework that can be derived from the well-developed consumer utility theories. In other words, the existing approaches are more about an application of statistical theory rather than a demand model derived from consumer theories – the lack of key economic information like the substitution effects, income effects and welfare analysis is a concern. To address this problem, two approaches have been investigated in recent years to enhance the existing multivariate count data models, especially the ordered response model. One approach is advocated by Herriges et al. (2008) that combined the ordered response models with a demand system that consists of linear specified latent demand equations. This work is encouraging as it applies multivariate count models in the framework of demand systems, which is much more comprehensive and informative in analyzing consumer behavior. However, although this model is flexible in allowing own-price, cross-price, and non-price determinants of demand, it is still flawed in not satisfying the integrability conditions that required in utility-theoretic demand systems. The other approach is a combination of discrete choice and count data model that represented by Bhat et al. (2015), using the maximum composite marginal likelihood. Clearly, the first approach is under the scope of "product space" and the second is under the "characteristics space", counterpart to what has been developed in the conventional demand analysis framework. Of these two approaches, we think the demand system approach is in a more comprehensive and solid view. As indicated by von Haefen (2002), compared to the discrete choice random utility maximization approach, demand system approaches are appealing because they "fully integrate the extensive commodity selection and intensive derived demand choices within a coherent and consistent model of consumer behavior". Following this, we think it is of great value to further extend the demand system with count data outcomes.

In this paper, we are interested in developing a utility-consistent incomplete demand system with count outcomes modeled with the MPLN distribution to understand consumer preference from their brand choices using retail-level data. As a count data system in the "product space", our research can be seen as an extension of the work by Herriges et al. (2008) but different in the selection of count distributions, the specification of the demand function, as well as the estimation method. The merits of using the MPLN distribution to model the expected count outcomes include: first, it allows flexible correlation structures among different count outcomes – which is in our case – the purchase amount of different brands; second, it can satisfy the integrability conditions required by utility theories underlying the incomplete demand system so that the welfare measures will be exact; third, it can accommodate severe over-dispersion in consumer brand choices. In terms of the estimation method, we adopt a numerical integration approach – Gauss-Hermite integration – rather than the classical methods of either maximum simulated likelihood or Bayesian approach. As mentioned before, the major concern of the mixing count data model with the MPLN distribution is the cumbersome estimation that is caused by the lack of a closed-form integrated likelihood function. However, given that the conditional mean of the Poisson distribution is assumed to be log-normally distributed, Gauss-Hermite integration is applicable. As we found in the first chapter, Gauss-Hermite integration is always more accurate and efficient compared to the quasi-Monte Carlo simulation when the dimension of integration is not too high (≤ 6). Given this, we estimate the count system with Gauss-Hermite integration and test its performance through empirical studies.

Further, different from most existing studies in count data analysis, this paper is unique in applying the count data demand system with the real transaction data at the retail level. Traditional data used in count data analysis are mostly survey data in topics like recreational demand, car accidents, health care utilization, etc. (Chib and Winkelmann, 2001; Egan and Herriges, 2006; Herriges et al., 2008; Whitehead et al., 2010; Haque et al., 2010; Bhat et al., 2015). Beyond that, we think the power of count data analysis should be expanded to larger scale consumer data sets in a wider range of areas. With the emergence of scanner technology from the 1970s, consumer purchase data has been disaggregated into the household level and product level so that the analysis of differentiated products, especially the estimation of demand systems and demand parameters is much more convenient and powerful (Cotterill, 1994; Baron and Lock, 1995; Bronnenberg et al., 2008). Unlike the stated preference data or the revealed preference data collected through survey, revealed preference data from real market transactions has the advantage of avoiding biases caused by either sample selection, contextual differences between the survey scenarios and actual purchase sets, or the inconsistency between what people state and what people do (Ben-Akiva and Morikawa, 1990; Brownstone et al., 2000; Swait and Andrews, 2003; Allenby et al., 2005; Brooks and Lusk, 2010; Ellickson et al., 2017). With these merits, our study applies the MPLN model with incomplete demand system into a retail-level panel data set of facial tissue consumption in Eau Claire, Wisconsin from the IRI marketing data set.

In summary, our study contributes to several aspects of literature. First, we enhance the mixing approach of the multivariate count data model by inserting it into a utility-consistent framework and estimate it through an innovative numerical method of of Gauss-Hermite integration. Second, we enhance the existing demand system by contributing a more diverse model specification with count outcomes. This would be especially important for researchers interested in certain economic values in the form of non-negative integers. Third, we expand the application of the count data model, or the demand system with count data, from survey data to scanner data, providing a reasonable solution to accommodate the common problem of over-dispersion in this high quality data.

The following sections are arranged as below: the second section describes our approach in developing the incomplete demand system with Poisson-log Normal distribution; the third section introduces the data we use for the empirical analysis; the forth section presents the empirical result and the last section is the discussion and conclusion of our paper.

4.2 Methodology

4.2.1 Poisson-log Normal Distribution

The Poisson distribution has been widely applied in modeling the number of events occurring in a fixed interval of time or space given that the frequency of each event is constant and independent from the time since the last event happened (Haight, 1967). In our case, we use the Poisson distribution to model the number of facial tissue boxes a household purchases during a fixed interval of time. When the univariate Poisson is expanded to multivariate space, the probability mass function was first developed by McKendrick (1916) and Wicksell (1916), independently. However, such a multivariate extension, as shown statistically, can not support a flexible correlation structure that is desired for brand choices. Take the bivariate Poisson distribution as an example, let $y_1 = x_1 + z$ and $y_2 = x_2 + z$ so that we allow correlations between y_1 and y_2. Then, the joint probability mass function, as noted by Inouye et al. (2017), could be written as

$$P(y_1, y_2|x_1, x_2, z) = exp(-\lambda_1 - \lambda_2 - \lambda_0) \times \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_1^{y_2}}{y_2!} \sum_{z=0}^{min(y_1,y_2)} \begin{pmatrix} y_1 \\ z \end{pmatrix} \begin{pmatrix} y_2 \\ z \end{pmatrix} z! \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^z .$$

Under this specification, the correlation coefficient resides in the following range

$$corr(y_1, y_2) \in \left(0, min\left\{\frac{\sqrt{\lambda_1 + \lambda_0}}{\sqrt{\lambda_2 + \lambda_0}}, \frac{\sqrt{\lambda_2 + \lambda_0}}{\sqrt{\lambda_1 + \lambda_0}}\right\}\right).$$

Given $\lambda_0, \lambda_1, \lambda_2$ are all non-negative, y_1 and y_2 can only be either independent or positively dependent. This is not reasonable as consumers who purchase a large amount of facial tissue in brand A may reduce their consumption of brand B in the following period if they are satisfied with their experience in brand A.

To allow a more flexible correlation structure among the count outcomes, a mixed Poisson distribution is developed by allowing the location parameter in Poisson distribution to be a random variable. Let the cumulative density function of the location parameter to be $G(\lambda|\boldsymbol{\varphi})$, the unconditional mass function of the observed count y is

$$P(Y = y) = \int f(y|\lambda) d_{G(\lambda|\boldsymbol{\varphi})} = \int \frac{e^{-\lambda}\lambda^y}{y!} d_{G(\lambda|\boldsymbol{\varphi})}$$

Aitchison and Ho (1989) introduced the multivariate log normal distribution into the mixing distribution $G(\lambda|\phi)$ to retain the rich covariance structures in the normal mixture approach, and to accommodate the practical consideration of a positive range of λ. For the univariate case, let $ln(\lambda) \sim N(\mu, \sigma^2)$, the Poisson-log normal probability mass function is a integral over the probability density function of λ

$$P(Y = y) = \int_{R_+} \frac{e^{-\lambda}\lambda^y}{\lambda y!} \frac{e^{0.5(ln(\lambda))-\mu)^2/\sigma^2}}{\sigma\sqrt{2\pi}} d\lambda, \quad y = 0, 1, 2, \ldots$$

Evaluation of this integral is made difficult due to the requirement that $\lambda > 0$. Introducing the reparameterization that $ln(\lambda) = \mu + \sigma\epsilon$, where $\epsilon \sim N(0, 1)$. With the corresponding Jacobian of transformation, $exp(\mu + \sigma\epsilon) \times \sigma = \lambda\sigma$, the probability mass function is

$$P(Y = y)) = \int_{-\infty}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} \frac{e^{-.5\epsilon^2}}{\sqrt{2\pi}} d\epsilon, \quad \lambda = e^{\mu + \sigma\epsilon}$$

Extending the Poisson-log normal to the multivariate case, the mixing function follows a multivariate log-normal distribution

$$g(\boldsymbol{\lambda}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{\frac{1}{2}J}(\lambda_1\lambda_2 \ldots \lambda_J)^{-1}|\boldsymbol{\Sigma}|^{\frac{1}{2}} exp\{-\frac{1}{2}ln(\boldsymbol{\lambda} - \boldsymbol{\mu})'\boldsymbol{\Sigma^{-1}}(ln(\boldsymbol{\lambda} - \boldsymbol{\mu})\}$$

Again, reparameterazing with $ln(\lambda) = \boldsymbol{\mu} + \boldsymbol{\Sigma\epsilon}$, where $\boldsymbol{\epsilon} \sim N_J(\mathbf{0}, \boldsymbol{\Sigma})$, we have the joint probability of consumer's choice of J brands expressed as

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \int \ldots \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \frac{e^{-\lambda_j} \lambda_j^{y_j}}{y_j!} \frac{exp\left[-\frac{1}{2}\boldsymbol{\epsilon\Sigma}^{-1}\boldsymbol{\epsilon}\right]}{(2\pi)^{J/2}|\boldsymbol{\Sigma}|^{1/2}} d\epsilon_1 \ldots d\epsilon_J$$

$$= \int \ldots \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \frac{e^{exp(\mu_j + \sigma_{jj}\epsilon_j)} exp(\mu_j + \sigma_{jj}\epsilon_j)^{y_j}}{y_j!} \frac{exp\left[-\frac{1}{2}\boldsymbol{\epsilon\Sigma}^{-1}\boldsymbol{\epsilon}\right]}{(2\pi)^{J/2}|\boldsymbol{\Sigma}|^{1/2}} d\epsilon_1 \ldots d\epsilon_J$$

As noted by Aitchison and Ho (1989), under such specification, we have

$$E(y_j) = exp(\mu_j + \frac{1}{2}\sigma_{jj}) = \lambda_j$$

$$var(y_j) = \lambda_j + \lambda_j^2\{exp(\sigma_{jj}) - 1\}$$

$$corr(y_i, y_j) = \frac{exp(\sigma_{ij}) - 1}{[\{exp(\sigma_{ii}) - 1 + \lambda_i^{-1}\}\{exp(\sigma_{jj}) - 1 + \lambda_j^{-1}\}]^{\frac{1}{2}}}$$

Apparently, the variance of the MPLN distribution converges to its mean when $\sigma_{jj} \to 0$, which means the variation on the location parameter is very low, or, there is no heterogeneity in the brand consumption counts for households. Otherwise, when $\sigma_{jj} \neq 0$, we will always have $var(y_j) > E(y_j)$ so that the marginal distribution of the purchase amount for brand j (y_j) has overdispersion relative to the pure Poisson. Also, the correlation coefficient between the purchase amount of different brands could be either positive, 0, or negative, depending on the sign of the covariance of the underlying multivariate normal distribution σ_{ij}. These properties make the MPLN distribution very desirable in fitting retail level purchase data with severe overdispersion.

4.2.2 Incomplete Demand System

To further ensure our model to be utility consist, we incorporate the incomplete demand system in specifying the vector of expected counts in the MPLN distributions. The incomplete demand system, which was proposed by Epstein (1982) and further developed by LaFrance and Hanemann (1989) and von Haefen (2002), is a less restrictive demand system that focuses on a group of goods that form a subset of the household budget, assuming the price of other goods to be quasi-fixed. The incomplete demand system specification is an attractive alternative to continuous models since it permits log-linear and semi-log models, which is a necessary requirement given the exponential link in the Poisson-log normal count data model.

In the incomplete demand system, the Marshallian demand functions are defined in the following form

$$y_i = y_i\left(\boldsymbol{p}, \boldsymbol{q}, I, \boldsymbol{\beta}\right), \qquad i = 1, \ldots, n,$$

where y_i is the Marshallian demand for good (brand) i, $\boldsymbol{p}$ is a vector of prices for the goods (brands) of our interest, $\boldsymbol{q}$ is a vector of prices (assumed to be quasi-fixed) for the other goods in the economy, I is the consumer's income, and $\boldsymbol{\beta}$ is the vector of the parameters in the demand functions.

Since we are interested in estimating the expected counts (the means) of the MPLN model, the demand (as opposed to share) with count outcomes is preferred. Among the twenty four available demand specifications derived by LaFrance and Hanemann (1985, 1986, 1990) and von Haefen (2002), the following specification is chosen because its in a log-linear form that consistent with the link funciton in the MPLN distribution.

$$y_j = \alpha_j\left(\boldsymbol{q}\right)\left\{\prod_{j=1}^{J} {p_j}^{\beta_{jk}}\right\} Income^{\gamma_j}, \; \forall\; j = 1, 2, \ldots, J$$

To ensure a consistent Hicksian welfare measure from the incomplete demand system, integrability problems need to be addressed to ensure the system is consistent with the rational preference ordering and that a rational individual always maximizes the utility given a linear budget constraint. Refering to the weak integrability concept from LaFrance and Hanemann (1986), an incomplete demand system shall satisfy the four conditions of (a) the demand functions are zero homogeneous in price and income; (b) nonnegative purchase amount $\boldsymbol{x} \geq \mathbf{0}$; (c) total expenditure of the individual demand on the subset of goods is strictly less than total income; (d) the Slutsky substitution matrix is symmetric (the change in Marshallian demand x_i with respect to a price change in p_j equals to the change in Marshallian demand x_j with respect to the same level of price change in $p_i \Rightarrow \mathrm{S}_{ij} = S_{ij}$) and negative semidefinite (the eigenvalues of the Skustky matrix are nonpositive).

Applying the four integrability conditions to the log-linear demand function we choose, von Haefen (2002) derived the four parameter restrictions as below

$$\alpha_j(\boldsymbol{q}) > \mathbf{0}$$

$$\beta_j < 0$$

$$\beta_{jk} = 0 \, \forall \, j \neq k$$

$$\gamma_j = \gamma \, \forall \, j$$

These restrictions ensure the exact welfare measures in the incomplete demand system, but also add a high cost of assuming no cross-price effect among different brands and assuming the income effect to be the same for all brands. As suggested by Herriges et al. (2008), the unrestricted demand systems that does not satisfying the integrability conditions may still be considered for the merit of flexibility in price and income effects. For the unrestricted form, it should be noted that the welfare measures are based on consumer surplus approximations instead of the Hicksian indirect utility function. With this consideration, we will estimate our model for both with- and without- the integrability restrictions.

For the specification of household demand for facial tissue, the expected purchase count is defined in a log-linear form of the prices of the J brands, the household income, and three important demographic factors that we think will influence households' demand on facial tissue: household size, whether the household have kids in house, and whether the household belongs to the elder group. Note that since the expected count of the MPLN distribution is reparameterized as $ln(\boldsymbol{\lambda}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_J(\mathbf{0}, \boldsymbol{\Sigma})$, we will give specifications for both the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Define the expected household purchase amount of household i on brand j as

$$\lambda_{ij} = E\left(x_{ij}\right) = \left[\left(\prod_{k=1}^{J} Price_k{}^{\beta_{jk}}\right) exp(\alpha_j + \tau_j HouseholdSize_i + \psi_j KidStatus_i + \delta_j Elder_i)\right] Income_i{}^{\gamma_j}$$

$$i = 1, \ldots, N; j = 1, \ldots, J$$

The mean ($\boldsymbol{\mu}$) of the log-normal distribution of $\boldsymbol{\lambda}$ can be written as:

$$E(ln(\lambda_j)) = \mu_{ij} = \alpha_j + \tau_j HouseholdSize_i + \psi_j KidStatus_i + \delta_j Elder_i + {}_{k=i}^{J} \beta_{jk} ln(Price_k) + \gamma_j ln(Income_i)$$

where the effect between the independent variables price, income and the dependent variable of expected household demand is in a log-log form. In this way,

$$\beta_{jk} = \frac{\partial ln(\lambda_j)}{\partial ln(Price_k)} = \frac{\partial \lambda_j}{\partial Price_k} \frac{Price_k}{\lambda_j} = \varepsilon_{j,k}$$

$$\gamma_j = \frac{\partial ln(\lambda_j)}{\partial ln(Income_i)} = \frac{\partial \lambda_j}{\partial Income_i} \frac{Income_i}{\lambda_j} = \varepsilon_{w,j}$$

Thus, coefficient β_{jk} is the price elasticity between brand k and j, and γ_j is the income elasticity of brand j. Note that this is different with the conventional Poisson count data model in which the price/income coefficients are marginal effects instead of elasticities.

For the diagonal elements in the variance-covariance matrix ($\boldsymbol{\Sigma}$) of $\boldsymbol{\lambda}$, define

$$\sigma_{ijj}^2 = \exp(\eta_{1j} HouseholdSize_i + \eta_{2j} Income_i)$$

for each household. For off-diagonal elements in $\boldsymbol{\Sigma}$, we will treat each of σ_{ij} as a single parameter. In this way, we can account for heteroskedasticities in expected purchase counts. Also, since the household income in real data sets may not be available or accurate, we use household total expenditures on all goods as an indicator to approximate household income in each period.

4.2.3 Gauss-Hermite Integration

As an alternative to the classical estimation methods like the maximum simulated likelihood (MSL) and the Bayesian approach, we employ the Gauss-Hermite quadrature in the estimation of the incomplete demand system. As mentioned before, the estimation of MPLN count data model is criticized to be cumbersome due to the convergence problem in the MSL method and the inconvenience in posterior sampling in the Bayesian approach. The merit of Gauss-Hermite integration is the careful selection of evaluation nodes that is helpful in delivering a more accurate and efficient estimation. More details about Gauss-Hermite integration can be found in chapter 1.

4.3 Data

For the empirical analysis, we use the scanner data provided by the IRI (Information Resources, Inc.) marketing data set. The IRI marketing data set is a weekly transaction data in both the grocery and drug chain with the Universal Product Codes (UPC), the price, transaction time, purchase unit, and other information like advertising features and displays, etc. It includes a broad array of markets (50 cities) across the United States from 2001 to 2012, covering 30 categories of food and daily necessities that account for approximately 25% - 30% of the consumer packaged goods sales in a grocery store (Bronnenberg et al., 2008; Kruger, 2017). Among the 50 markets, 48 are standard markets and 2 are Behavior Scan markets with household ID: one is Eau Claire, Wisconsin, and the other is Pittsfield, Massachusetts. Given the high quality and wide coverage of the data, the IRI marketing data set has been intensively adopted in existing research that over 100 publications have been generated from it by 2016.

To select a product with high homogeneity in packaging so that the purchase quantities among different brands are more comparable, we choose the facial tissue market in which the product size is more uniform. Also, considering the high dimensioned summation and tremendous possible outcomes for the joint probability, we decide to take the lastest year (2011) as an example for our analysis. To obtain the household demographic information, we choose Eau Claire, Wisonsin from the two panel Behavior Scan markets given this city has a bigger population compared to the other city. Further, out of the three available outlets in the data set: grocery, drug store, and mass, we focus on the grocery store as it accounts for the majority of transactions (89.85% in 2011) in the facial tissue market .

To construct the data for the estimation of the incomplete demand system, we merge the grocery store transaction data, the product attribute data, and the household demographic data together and make the purchase amount (the number of boxes) of facial tissue uniform by the volume equivalence for each transaction. Given the purchase cycle of facial tissue market is pretty long – 70 months according to Bronnenberg et al. (2008) – we aggregate the facial tissue purchase amounts by 6 months to avoid too many zero observations. In this way, we have two purchase periods, January to June and July to December, for each household in 2011.

For the brand in facial tissue market, we have three major brands and a "private label" brand that is a collection of a series of grocery store owned brands in the data set. Other than these four, there are still some brands with very small market shares. To protect the information of the specific brand, we name the four major brand as "A", "B", "C", and "D", in which brand C is the private brand owned by grocery stores. The market share of the facial tissue market in the three most recent years (2009-2011) is shown in Figure 4.1. As seen, the largest four facial tissue brands accounts for more than 97.13% of the market share in Eau Claire. Thus, analysis of the four major brands could be seen as complete for the whole market.

In terms of the households included in our estimation, we only retain the "static" customers who satisfy the minimal reporting requirement defined in the IRI marketing data set. Also, we drop the households whose facial tissue purchase amount is 0 for all the four brands over the two periods in 2011 as no information could be derived in such cases. We also drop two extreme households whose total facial tissue purchase amount is over 140 in 2011. This leaves us 1927 households in total. The demographic information of these households is provided in Table 4.1 - 4.2 and Figure 4.2 - 4.5. The income of the households distributes relative evenly instead of in a normal shape. A spike is observed in the high income range from $75,000 to $99,999 per year. For family size, 47.47% of households have two people, followed by the case with one people and bigger families with three or more people. For race, 97.24% of the households are white, followed by black-African American and Asian. This is somewhat consistent with the population structure in Eau, Claire that 91.40% are white in the 2010 census. In terms of the house type, most households (86.16%) rent their house instead of own a house. For age, households head whose age is 65+ accounts for 40.05 percent in our sample. For education, we have a nearly normal distribution with the mode in "some high school" that accounts for 32.69%, followed by "graduated high school" (27.13%) and "technical school" (20.64%). Also, the occupation of 31.25% of the households is "private household worker", followed by "professional or technical" (21.20%). In addition, most of the households (83.03%) do not have children under 18 years old in the house. In summary, we have a sample of households with more elderly people who may have a potentially high demand for facial tissue.

For the facial tissue purchase quantity, severe over-dispersion is observed for all the four major brands. As shown in Figure 4.6, the histogram of facial tissue purchase amount (the number of uniform boxes) has a long tail for each brand in both periods. More specifically, Table 34.3 shows that, although the 95^{th} percentile of purchase quantity concentrates around only 10 uniform boxes for all the four brands, the maximum purchase amount is as high as 59 uniform boxes. Such an over-dispersion is especially severe for brand A and C.

For the prices of the four brands, the summary statistics are listed in Table 4.4. The retailing price for brand B is the highest among the four, and brand C is the cheapest one – the private brand owned by grocery store always offers a lower price in the market. Also, price variations are the biggest for brand B. This is reasonable given more price fluctuation is available for this most expensive brand to allow promotions and discounts. In contrast, standard deviations for price is the smallest for brand C, whose average price is the lowest among the four. Private brand facial tissue has the most stable pricing performance.

In summary, the structure of the data shows the need of special care for over-dispersion. With rich information in product, transaction, and household level, it is appropriate to apply the incomplete demand system with the MPLN distribution.

4.4 Empirical Results

4.4.1 Case 1. Unrestricted Incomplete Demand System

We first fit the incomplete demand system for household purchase amount of the four major brands of facial tissue without any parameter restrictions. As mentioned, we have 1927 households and two purchase periods in 2011. Also, since the household income in the IRI marketing data set is reported in category groups instead of continuous numbers, we use the household total expenditure on all the 30 categories of food and daily necessities in the data set as an indicator for household income. Gauss Hermite quadrature with 20^{th} degree of polynomial and adjusted weight is used to numerically approximate the likelihood that includes a four-dimensional integral. The estimation is conducted in MATLAB and the results are shown in Table 4.5 to Table 4.7.

For the mean ($\boldsymbol{\mu}$) of the log-normally distributed expectations of household facial tissue demand ($\boldsymbol{\lambda}$), the estimators are shown in Table 4.5 and 4.6. The result in Table 4.5 shows that the Marshallian own price elasticity for the four brands is always negative (-0.899 to -0.926) and highly significant, except for brand C (the private label owned by grocery stores) whose own price elasticity is negative but small in absolute value (-0.251) and is not significant. This is reasonable as from the summary statistics, brand C always has the lowest price and the smallest price fluctuations. In other words, the pricing strategy of brand C is very stable and consistent. The income elasticity for all the four brands is positive and highly significant. The magnitude of income elasticity is similar for brand B, C, and D (0.421 to 0.456), but a little bit larger for brand A (0.671). This means brand A has a stronger response to income change so that when people improve their economic situation, they tend to consume more facial tissue in brand A. Furthermore, we find elderly people whose age is 55 or more have a higher demand for brand C and brand D. The kids status, which indicates whether the household has children under 18 years old in house, does not have a significant influence on household demand for facial tissue. However, this finding may not be representative given 83.03% of households in our sample have no children in house. Also, we did not find a significant effect of household size on the mean of expectation of facial tissue demand.

In terms of the cross-price elasticity, the result is combined with the own-price elasticity and listed in Table 4.6. Among the twelve pairs of cross-price elasticities, four are statistically significant and negative (β_{AB}, β_{BA}, β_{CD}, and β_{CB}), indicating that brand A and B are more likely to be complement goods to each other. It is possible that the marketing strategies of these two brands – such as promotions– are always following each other, or one of these two brands is the price leader and the other is the follower so that customers may face lower prices for both brands at the same time, resulting in higher purchase volumes simultaneously. Also, a price increase in brand B and D (especially brand D) leads to a significant decrease of Marshallian Demand in brand C. The remaining cross price elasticities are all insignificant.

For the variance-covariance matrix ($\mathbf{\Sigma}$) of the log-normally distributed expectations of household demand ($\boldsymbol{\lambda}$), Table 4.7 presents the estimation result. As mentioned in the methodology section, diagonal elements in the matrix are defined for each household i as

$$\sigma_{ijj}^2 = \exp(\eta_{1j} HouseholdSize_i + \eta_{2j} Income_i).$$

In the result, income has a significant positive effect on the variations of expected demand for all of the four brands. Further, household size has a negative but small (<0.1) effect on the variance of expected demand for facial tissue of brand C. The off-diagonal elements are all significant, validating the necessity of controlling for correlations and heteroskedasticities. For the two pairs of brands, brand A and D, brand C and D, the correlations of the expected demand are positive, meaning the two pairs of brands have the same direction of changes in expected demand. For the remaining brand pairs, the correlations of expected demand changes are negative.

4.4.2 Case 2. Restricted Incomplete Demand System

To satisfy the integrability conditions of the log-linear demand specification, we estimated the restricted incomplete demand system with the results presented in Table 4.8 to 4.10. In the restricted system, the income elasticity on expected facial tissue demand is the same among the four brands, and the cross-price elasticities are set to zero.

Table 4.8 provides the estimators for the the mean ($\boldsymbol{\mu}$) of the log-normally distributed expectations of household facial tissue demand ($\boldsymbol{\lambda}$). Similar to the unrestricted case, the own-price elasticity in the restricted system is negative and highly significant for most of the four brands, except for brand C whose own-price elasticity is negative, but not significant (-0.244). For the income elasticity, the shared income elasticity is estimated to be 0.557 and is highly significant, meaning the demand for facial tissue is significantly higher when household income is increased. Also, we find elderly people whose age is 55 or more prefer brand C and D. In contrast, the effects of household size and kids status are both very small and insignificant. Since the cross-price elasticities are restricted to be zero, the Marshallian price elasticity matrix in Table 4.9 only has diagonal elements. However, the compensated (Hicksian) price elasticities can still be derived given the integrability conditions are satisfied in the restricted system. From the Slutsky equation with elasticity, the Hicksian cross-price elasticity is defined as

$$\varepsilon^H_{p,ij} = \varepsilon^M_{p,ij} + \varepsilon_{w,i} b_j,$$

where $\varepsilon^M_{p,ij}$ is the uncompensated cross price elasticity between brand i and j, $\varepsilon_{w,i}$ is the income elasticity of brand i, and b_j is the market share of brand j. Since $\varepsilon^M_{p,ij} = 0$, we have $\varepsilon^H_{p,ij} = \varepsilon_{w,i} b_j$ so that Hicksian cross-price elasticities are very small in value given the budget share of facial tissue is very low in household total expenditures.

For the variance-covariance matrix ($\boldsymbol{\Sigma}$) of household expected demand $\boldsymbol{\lambda}$, the estimated result for the restricted system is provided in Table 4.10. Again, for the diagonal elements in the matrix, income has a positive and highly significant effect that ranges from 0.162 to

0.297, indicating higher variance of purchase amount is observed for all the four brands when household income is increased. The other factor that is influencing the diagonal element, the household size, is significant for brand A, B, and C in the restricted case and has a negative sign for all brands. This means the households with more people have a more stable demand over facial tissue. However, the magnitude to such effect is still relatively small for brand A and B (<0.1). For brand C, the effect is estimated to be -0.131. In terms of the off-diagonal elements in the matrix, all correlations are significant, with the signs exactly the same as the unrestricted case.

4.4.3 Summary

Using Gauss-Hermite integration, we estimated both the unrestricted and restricted incomplete demand system for household consumption of facial tissue with the MPLN distribution. Comparing the result in the two cases, we find a high consistency in both the sign and magnitude of the estimated parameters in brand demand.

First, all the own-price elasticities are found to be significantly negative (expect for brand C with insignificant negative own-price effect) and all the income elasticities are significantly positive for all the four brands in both cases. Second, we find brand C and D are more welcomed for the elder group whose age is 55 or more. One reason could be the low price of brand C that is more affordable for the elder group, the other reasons could related to tissue quality that is more friendly to the elderly. Further, we find the demand of brand A and B to be more sensitive to household income changes. The income elasticity of expected demand is the highest for brand A in the unrestricted case, while the the income effect affecting the variance of expected demand of brand B is the highest among the four in the restricted case. In contrast, we found brand C to be a stable brand whose demand does not respond actively toward price changes in both cases. As the cheapest brand owned by grocery stores, we believe this finding is reasonable as the effect of marketing strategies like discounts are expected to be less feasible (since the original price is already low) and less impressive.

For the relationship among the four brands, we find four pairs of significant negative cross-price elasticities in the unrestricted system. The means the expected demand for brand A and B respond negatively toward the price change in each other brand, making these two brands a complement pair. Similarly, a price increase in brand B and D leads to negative response in the Marshallian demand of brand C. For the restricted case, we set the Marshallian cross-price elasticities to be zero. But still, the compensated (Hicksian) cross-price elasticities are not zero, although they are very small in positive values given the low budget share of facial tissue in household total expenditures. Thus all brands are Hicksian substitutes. Also, from the variance-covariance matrix of the expected demand in both the unrestricted and restricted systems, we find positive and significant correlations between brand A and D, brand C and D, while other pairs have negative correlations in demand. These findings are very helpful to understand the structure of facial tissue market and consumer habits toward different brands.

At the same time, we find a significant difference in the model fit between the restricted and unrestricted systems. The likelihood ratio (LR) test shows the unrestricted system fits the data significantly better than the restrictive case (p-value <0.001). In other words, although the integrability conditions can enable a consistent Hicksian welfare measure, it brings the cost of reducing the model fit. This trade off is also observed in Herriges et al. (2008) that suggesting the high cost of exact welfare analysis in count systems. We think the choice of system specifications should be in accordance with specific research questions. If one is more interested in consistent welfare measures, then it is still worthy to apply the integrability conditions although it may reduce the model fit.

4.5 Conclusion

In this paper, we develop an innovative approach analyzing consumer brand choice through an incomplete demand system with count outcomes following the multivariate Poisson-Log normal (MPLN) distribution. Different than most existing demand models using market shares as the dependent variable, our approach targets the purchase amount – the non-negative integers. To fit the count outcomes, we choose the MPLN distribution given its merit in allowing flexible covariance structures among counts, its good performance in fitting over-dispersion, and its advantage in accommodating both zero and non-zero consumption. Specifically, we emphasize the importance of utility consistency when developing multivariate count data models for consumer demand. Thus, we specify the mean of consumer purchase counts under a log-linear demand form in the incomplete demand system that introduced by von Haefen (2002). Further, we apply the Gauss-Hermite quadrature to approximate the multidimensional integration in the likelihood constructed by the MPLN density. Different than the traditional estimation methods like maximum simulated likelihood and Bayesian approach, Gauss-Hermite integration is shown to be more effective as it uses the information of parameter distributions in the selection of evaluation nodes.

In the empirical analysis, we extend the application area of the count data demand model to the real transaction (scanner) data set at retail level. This extension helps to provide more insights on consumer preferences and market structure. Using a panel of scanner data provided by the IRI marketing data set, we analyze the brand choice of 1927 static consumers on facial tissue market in Eau Claire, Wisconsin in 2011. Our model accommodates well the over-dispersed purchase count, and provides informative insights on the market structure and consumer preferences. Among the four major brands that accounts for more than 97% of the market share, we find the demand of the cheapest brand to be quite stable and does not significantly respond to price changes. Also, households with higher incomes have higher probabilities of switching brands. Besides, our model is helpful in identifying the most promising consumers for a specific brand. For example, we find two brands that are significantly more welcomed by the elder group, while household size and kids status does not have significant influence on the expected demand of facial tissue.

In summary, our study contributes to the literature by enhancing the applicability and power of the multivariate count data model in a utility-consistent framework. It also enriches existing demand system analysis by further developing models with more types of outcomes. Besides, with the more detailed, accurate, and big transaction data available in today's market, we advocate the use of scanner data in count data models with a utility consistent framework.

For future work, we think the comparison between the demand system approach and discrete choice approach would be valuable in developing utility-consistent multivariate count data models. For discrete choice models (like the mixed logit model), scanner data provides a good source of real "choice experiment" observations in which a household choose one brand (one response) among all available brands (all alternatives) in the market in each transaction (each question). By summing up the quantities purchased by the household, the multinomial brand choice in each transaction (each choice experiment question) is aggregated into count outcomes – the household demand for different brands. Comparing these two approaches, a demand system is under the "product space" and has the advantages of 1) providing multiple (own and cross) price effects, income effects to different brands, and 2) revealing the effect of demographic factors in consumer demand. But it cannot picture how the product attributes contribute to consumer choices. In contrast, the discrete choice approach is under the "characteristic space". It has the merit in 1) enabling the individual-specific price effect (although only one same price effect for all brands) that can account for heterogeneities in household level, and 2) presenting the consumer willingness to pay toward specific product attributes. But it cannot reveal the role of demographic factors in consumer preferences. Clearly, these two approaches can complement each other by providing the information that is limited in the other approach. Comparisons and combinations of these two approaches would contribute to a comprehensive view of consumer preferences with their choices in the form of multivariate count outcomes.

References

Aitchison, J. and Ho, C. (1989). The multivariate poisson-log normal distribution. *Biometrika*, 76(4):643–653.

Allenby, G., Fennell, G., Huber, J., Eagle, T., Gilbride, T., Horsky, D., Kim, J., Lenk, P., Johnson, R., Ofek, E., et al. (2005). Adjusting choice models to better predict market behavior. *Marketing Letters*, 16(3-4):197–208.

Bach, P., Farbmacher, H., and Spindler, M. (2018). Semiparametric count data modeling with an application to health service demand. *Econometrics and Statistics*, 8:125–140.

Baron, S. and Lock, A. (1995). The challenges of scanner data. *Journal of the Operational Research Society*, 46(1):50–61.

Barten, A. P. and Turnovsky, S. J. (1966). Some aspects of the aggregation problem for composite demand equations. *International Economic Review*, 7(3):231–259.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6):485–495.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. and Pakes, A. (2002). The pure characteristics discrete choice model of differentiated products demand. *Harvard, unpublished.*

Bhat, C. R., Paleti, R., and Castro, M. (2015). A new utility-consistent econometric approach to multivariate count data modeling. *Journal of Applied Econometrics*, 30(5):806–825.

Bhat, C. R. and Srinivasan, S. (2005). A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B: Methodological*, 39(3):255–278.

Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper - the IRI marketing data set. *Marketing Science*, 27(4):745–748.

Brooks, K. and Lusk, J. L. (2010). Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *American Journal of Agricultural Economics*, 92(4):1229–1241.

Brownstone, D., Bunch, D. S., and Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5):315–338.

Chib, S. and Winkelmann, R. (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1975). Transcendental logarithmic utility functions. *The American Economic Review*, 65(3):367–383.

Cotterill, R. W. (1994). Scanner data: New opportunities for demand and competitive strategy analysis. *Agricultural and Resource Economics Review*, 23(2):125–139.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *The American Economic Review*, 70(3):312–326.

Egan, K. and Herriges, J. (2006). Multivariate count data regression models with individual panel data from an on-site sample. *Journal of Environmental Economics and Management*, 52(2):567–581.

Ellickson, P. B., Lovett, M. J., and Ranjan, B. (2017). Product launches with new attributes: A hybrid conjoint-consumer panel technique for estimating demand. *Available at SSRN 3045379*.

Englin, J. and Shonkwiler, J. S. (1995). Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation. *The Review of Economics and statistics*, pages 104–112.

Epstein, L. G. (1982). Integrability of incomplete systems of demand functions. *The Review of Economic Studies*, 49(3):411–425.

Ferdous, N., Eluru, N., Bhat, C. R., and Meloni, I. (2010). A multivariate ordered-response model system for adults weekday activity episode generation by activity purpose and social context. *Transportation Research Part B: Methodological*, 44(8-9):922–943.

Haight, F. A. (1967). Handbook of the poisson distribution. 1967. *New York: Jon Wiley & Sons*.

Haque, M. M., Chin, H. C., and Huang, H. (2010). Applying bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis & Prevention*, 42(1):203–212.

Hellerstein, D. and Mendelsohn, R. (1993). A theoretical foundation for count data models. *American Journal of Agricultural Economics*, 75(3):604–611.

Herriges, J. A., Phaneuf, D. J., and Tobias, J. L. (2008). Estimating demand systems when outcomes are correlated counts. *Journal of Econometrics*, 147(2):282–298.

Hidayat, B. and Pokhrel, S. (2010). The selection of an appropriate count data model for modelling health insurance and health care demand: case of indonesia. *International Journal of Environmental Research and Public Health*, 7(1):9–27.

Inouye, D. I., Yang, E., Allen, G. I., and Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(3):e1398.

Kruger, M. W. (2017). Research use of the IRI marketing data set: bibliography. *Available at SSRN 2342688*.

LaFrance, J. T. and Hanemann, W. M. (1989). The dual structure of incomplete demand systems. *American Journal of Agricultural Economics*, 71(2):262–274.

McKendrick, A. G. (1916). Applications of the kinetic theory of gases to vital phenomena. *India Journal of Med. Res.*, 3:667–687.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365.

Scott, D. M. and Kanaroglou, P. S. (2002). An activity-episode generation model that captures interactions between household heads: development and empirical analysis. *Transportation Research Part B: Methodological*, 36(10):875–896.

Shonkwiler, J. (1995). Systems of travel cost models of recreation demand. In *Western Regional Research Project (W-133) Meetings in Pacific Grove, CA*.

Stone, R. (1954). Linear expenditure systems and demand analysis: an application to the pattern of british demand. *The Economic Journal*, 64(255):511–527.

Swait, J. and Andrews, R. L. (2003). Enriching scanner panel models with choice experiments. *Marketing Science*, 22(4):442–460.

Theil, H. (1965). The information approach to demand analysis. *Econometrica*, 33(1):67–87.

von Haefen, R. H. (2002). A complete characterization of the linear, log-linear, and semi-log incomplete demand system models. *Journal of Agricultural and Resource Economics*, 27(2):281–319.

Whitehead, J. C., Phaneuf, D. J., Dumas, C. F., Herstine, J., Hill, J., and Buerger, B. (2010). Convergent validity of revealed and stated recreation behavior with quality change: a comparison of multiple and single site demands. *Environmental and Resource Economics*, 45(1):91–112.

Wicksell, S. D. (1916). Some theorems in the theory of probability with special reference to their importance in the theory of homograde correlation. *Svenska Aktuarie-foreningens Tidskrift*, pages 165–213.

Winkelmann, R. (2000). Seemingly unrelated negative binomial regression. *Oxford Bulletin of Economics and Statistics*, 62(4):553–560.

Table 4.1: Demographic Characteristics of the Sample Household

Indicator	**Level**	**Percentage**
Income	$00,000 to $ 9,999 per yr	6.29
	$10,000 to $11,999 per yr	3.48
	$12,000 to $14,999 per yr	2.75
	$15,000 to $19,999 per yr	6.29
	$20,000 to $24,999 per yr	7.58
	$25,000 to $34,999 per yr	14.08
	$35,000 to $44,999 per yr	13.45
	$45,000 to $54,999 per yr	6.13
	$55,000 to $64,999 per yr	10.08
	$65,000 to $74,999 per yr	5.35
	$75,000 to $99,999 per yr	14.23
	$100,000 and greater per year	10.29
Family Size	One people	21.57
	Two people	47.47
	Three people	14.54
	Four people	10.27
	Five people	5
	Six or more people	1.15
Race	White	97.24
	Black-African American	1.51
	Asian	0.88
	hispanic	0.36
House Type	Ownder	13.84
	Renter	86.16
Age	18-24	0.16
	25-34	1.61
	35-44	7.97
	45-54	22.81
	55-64	27.4
	65+	40.05
Education	Some grade school or less	0.58
	Completed grade school	3.46
	Some high school	32.69
	Graduated high school	27.13
	Technical school	20.64
	Some college	10.06
	Graduated from college	4.09
	Post graduate work	1.36

Table 4.2: Demographic Characteristics of the Sample Household (Continued)

Indicator	**Level**	**Percentage**
Occupation	Cleaning, food, health service worker	9.08
	Clerical	5.38
	Craftsman	1.36
	Laborer	1.58
	Manager or administrator	8.91
	Operative (machine operator)	2.23
	Private household worker	31.25
	Professional or technical	21.2
	Retired	8.15
	Sales	10.87
Children	Child in [0-5)	1.04
	Child in [6-11)	2.59
	Child in [12-17)	8.93
	Children in [0-5) & [6-11)	0.78
	Children in [0-5) & [12-17)	0.16
	Children in [6-11) & [12-17)	3.22
	Children in [0-5),[6-11) & [12-17)	0.26
	Family size>0 yet no children	83.03

Table 4.3: Summary Statistics of Purchase Quantity by Brand

Period	Brand	N_Purchaser	Mean	Std Dev	Min	Percentile					Max
						25th	75th	90th	95th	99th	
Period 1	A	704	3.439	4.335	1	1	4	7	11	23	55
	B	450	3.642	4.064	1	1	4	8	12	23	32
	C	668	3.377	5.07	1	1	3	7	11	28	52
	D	637	2.664	2.502	1	1	3	6	8	12	19
Period 2	A	540	3.198	3.949	1	1	4	7	11.5	20	37
	B	496	3.002	3.317	1	1	4	6	8	22	24
	C	723	3.231	4.779	1	1	3	6	9	24	59
	D	566	2.466	2.303	1	1	3	5	7	13	21

Table 4.4: Summary Statistics of Purchase Price by Brand

Period	Brand	N_Purchaser	Mean	Median	Std Dev	Minimum	Maximum
Period 1	A	704	0.928	0.944	0.339	0.445	4.271
	B	450	1.629	1.605	1.083	0.368	3.729
	C	668	0.713	0.675	0.300	0.491	2.355
	D	637	0.951	0.850	0.587	0.500	2.837
Period 2	A	540	1.012	1.000	0.279	0.571	4.271
	B	496	1.646	1.622	1.081	0.478	3.554
	C	723	0.714	0.655	0.287	0.537	2.500
	D	566	1.029	0.864	0.725	0.561	3.595

Table 4.5: Estimated Mean of the Log-normally Distributed Household Demand in the Incomplete Demand System (Unrestricted)

Brand	Parameter	Estimate	Standard Error (R)	Standard Error (H)	Z-value
Brand A	Constant	-4.674***	0.287	0.269	-17.371
	Price	-0.899***	0.236	0.161	-5.572
	Income	0.671***	0.05	0.047	14.26
	HH Size	-0.038	0.056	0.05	-0.749
	Kids	-0.09	0.139	0.117	-0.775
	>54	0.073	0.08	0.064	1.138
Brand B	Constant	-3.682***	0.674	0.476	-7.742
	Price	-0.926***	0.179	0.099	-9.356
	Income	0.421***	0.158	0.096	4.375
	HH Size	0.064	0.148	0.09	0.707
	Kids	-0.105	0.384	0.224	-0.47
	>54	-0.068	0.16	0.124	-0.55
Brand C	Constant	-3.452***	0.318	0.336	-10.289
	Price	-0.251	0.213	0.163	-1.545
	Income	0.418***	0.062	0.063	6.659
	HH Size	-0.006	0.048	0.051	-0.108
	Kids	0.155	0.125	0.138	1.122
	>54	0.238**	0.101	0.106	2.245
Brand D	Constant	-4.015***	0.369	0.366	-10.977
	Price	-0.904***	0.214	0.119	-7.59
	Income	0.456***	0.07	0.067	6.802
	HH Size	0.022	0.054	0.054	0.409
	Kids	-0.155	0.132	0.147	-1.059
	>54	0.306***	0.103	0.103	2.975

* Log likelihood at converge: -17672.9

Table 4.6: Price Effects among the Four Brands (Unrestricted)

	Brand A	Brand B	Brand C	Brand D
Brand A	-0.899***	-0.3321***	0.131	-0.125
	(0.161)	(0.113)	(0.205)	(0.108)
Brand B	-0.483*	-0.926***	0.254	0.026
	(0.279)	(0.099)	(0.214)	(0.195)
Brand C	-0.086	-0.197*	-0.251	-0.385***
	(0.190)	(0.118)	(0.163)	(0.108)
Branc D	0.051	-0.082	-0.152	-0.904***
	(0.184)	(0.130)	(0.187)	(0.119)

Table 4.7: Estimated Variance-Covariance of the Log-normally Distributed Household Demand in the Incomplete Demand System (Unrestricted)

Parameter	Estimate	Standard Error (R)	Standard Error (H)	Z-value
Household Size A	-0.04	0.035	0.035	-1.165
Household Size B	0.034	0.061	0.045	0.765
Household Size C	-0.089*	0.042	0.046	-1.946
Household Size D	-0.066	0.043	0.049	-1.338
Income A	0.152***	0.015	0.015	10.342
Income B	0.211***	0.020	0.020	10.741
Income C	0.223***	0.018	0.020	10.952
Income D	0.167***	0.021	0.023	7.192
COR AB	-0.221***	0.035	0.025	-8.965
COR AC	-0.084***	0.023	0.026	-3.253
COR AD	0.177***	0.041	0.035	5.004
COR BC	-0.274***	0.020	0.022	-12.526
COR BD	-0.159***	0.038	0.035	-4.527
COR CD	0.339***	0.043	0.037	9.185

Table 4.8: Estimated Mean of the Log-normally Distribute Household Demand in the Incompleted Demand System (Restricted)

Brand	Parameter	Estimate	Standard Error (R)	Standard Error (H)	Z-value
Shared	Income	0.557***	0.036	0.031	17.745
Brand A	Constant	-4.312***	0.226	0.201	-21.489
	Price	-0.914***	0.257	0.168	-5.430
	HH Size	-0.006	0.049	0.046	-0.134
	Kids	-0.132	0.120	0.108	-1.215
	Age > 54	0.050	0.085	0.068	0.729
Brand B	Constant	-4.582***	0.304	0.260	-17.619
	Price	-0.859***	0.176	0.098	-8.786
	HH Size	0.073	0.074	0.069	1.058
	Kids	-0.008	0.169	0.162	-0.047
	Age > 54	-0.097	0.116	0.115	-0.850
Brand C	Constant	-4.280***	0.287	0.240	-17.866
	Price	-0.244	0.215	0.163	-1.502
	HH Size	-0.008	0.059	0.054	-0.154
	Kids	0.088	0.122	0.137	0.644
	Age > 54	0.214**	0.109	0.108	1.977
Brand D	Constant	-4.508***	0.236	0.222	-20.339
	Price	-0.873***	0.177	0.109	-8.035
	HH Size	0.005	0.052	0.053	0.090
	Kids	-0.154	0.127	0.145	-1.062
	Age > 54	0.309***	0.097	0.101	3.050

* Log likelihood at converge: -17700.5

Table 4.9: Price Effects among the Four Brands (Restricted)

	Brand A	Brand B	Brand C	Brand D
Brand A	-0.914*** (0.168)			
Brand B		-0.859*** (0.098)		
Brand C			-0.244 (0.163)	
Branc D				-0.873*** (0.109)

Table 4.10: Estimated Variance-Covariance of the Log-normally Distributed Household Demand in the Incomplete Demand System (Restricted)

Parameter	Estimate	Standard Error (R)	Standard Error (H)	Z-value
Household Size A	-0.055*	0.031	0.032	-1.749
Household Size B	-0.088**	0.041	0.042	-2.114
Household Size C	-0.131***	0.049	0.045	-2.902
Household Size D	-0.071	0.045	0.050	-1.427
Income A	0.162***	0.014	0.013	12.144
Income B	0.297***	0.021	0.020	15.065
Income C	0.249***	0.028	0.021	11.634
Income D	0.169***	0.021	0.023	7.258
COR AB	-0.227***	0.020	0.018	-12.440
COR AC	-0.129***	0.024	0.022	-5.739
COR AD	0.157***	0.037	0.033	4.715
COR BC	-0.249***	0.033	0.028	-8.907
COR BD	-0.148***	0.035	0.035	-4.278
COR CD	0.347***	0.042	0.036	9.576

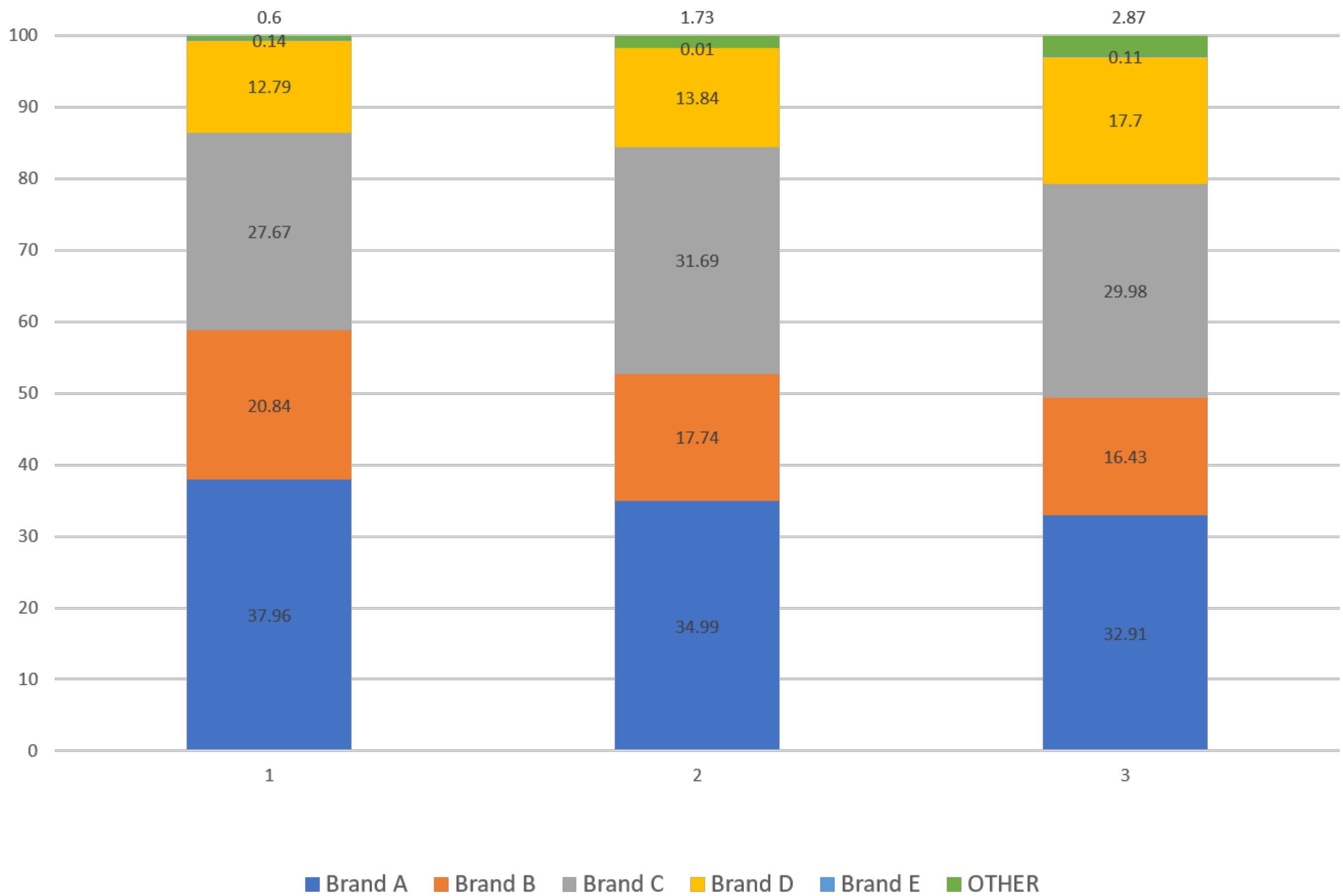


Figure 4.1: Market Share of Facial Tissue Market at Grocery Stores in Eau Claire, Wisconsin (2009 - 2011)

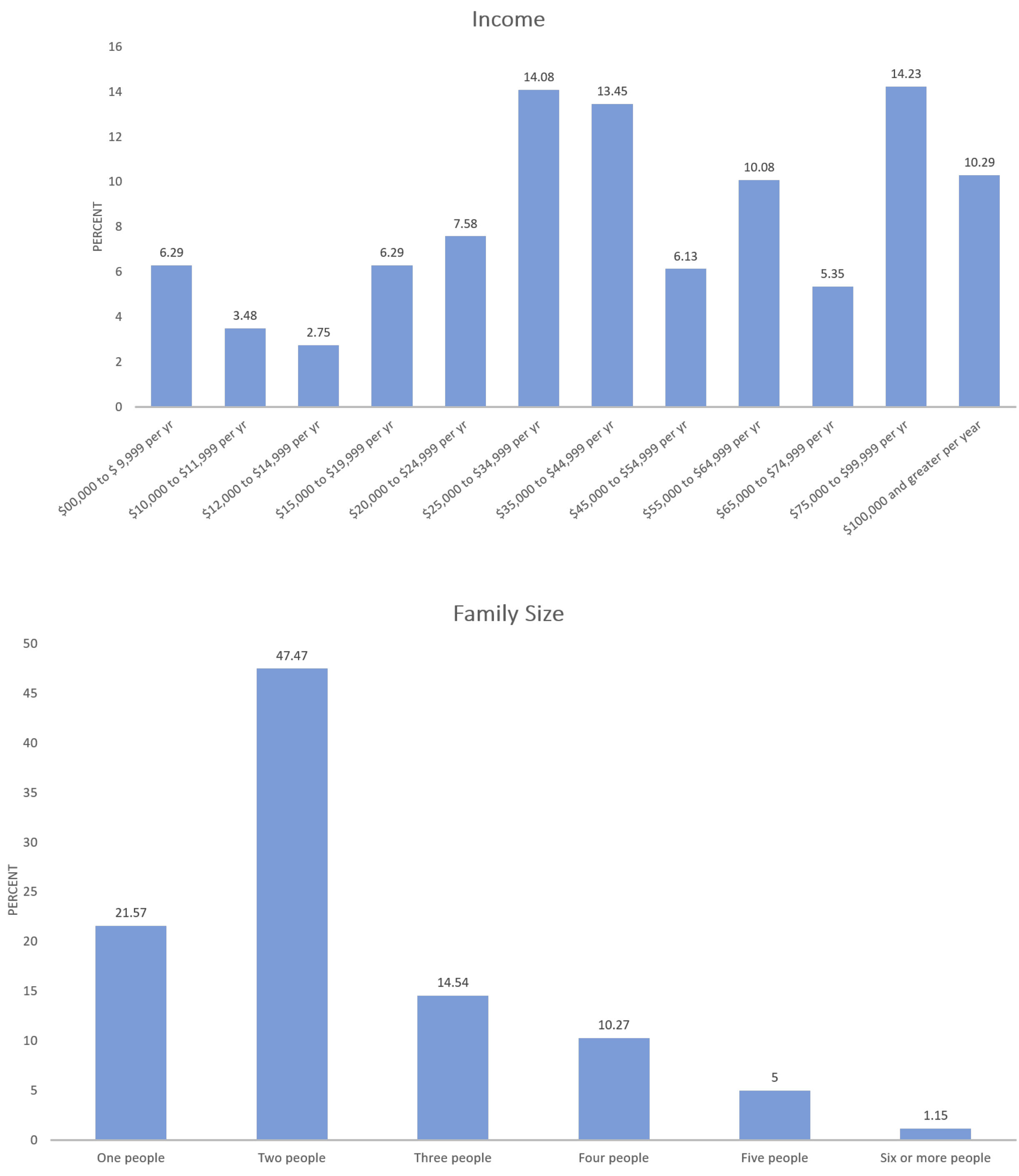


Figure 4.2: Household Demographic Factors

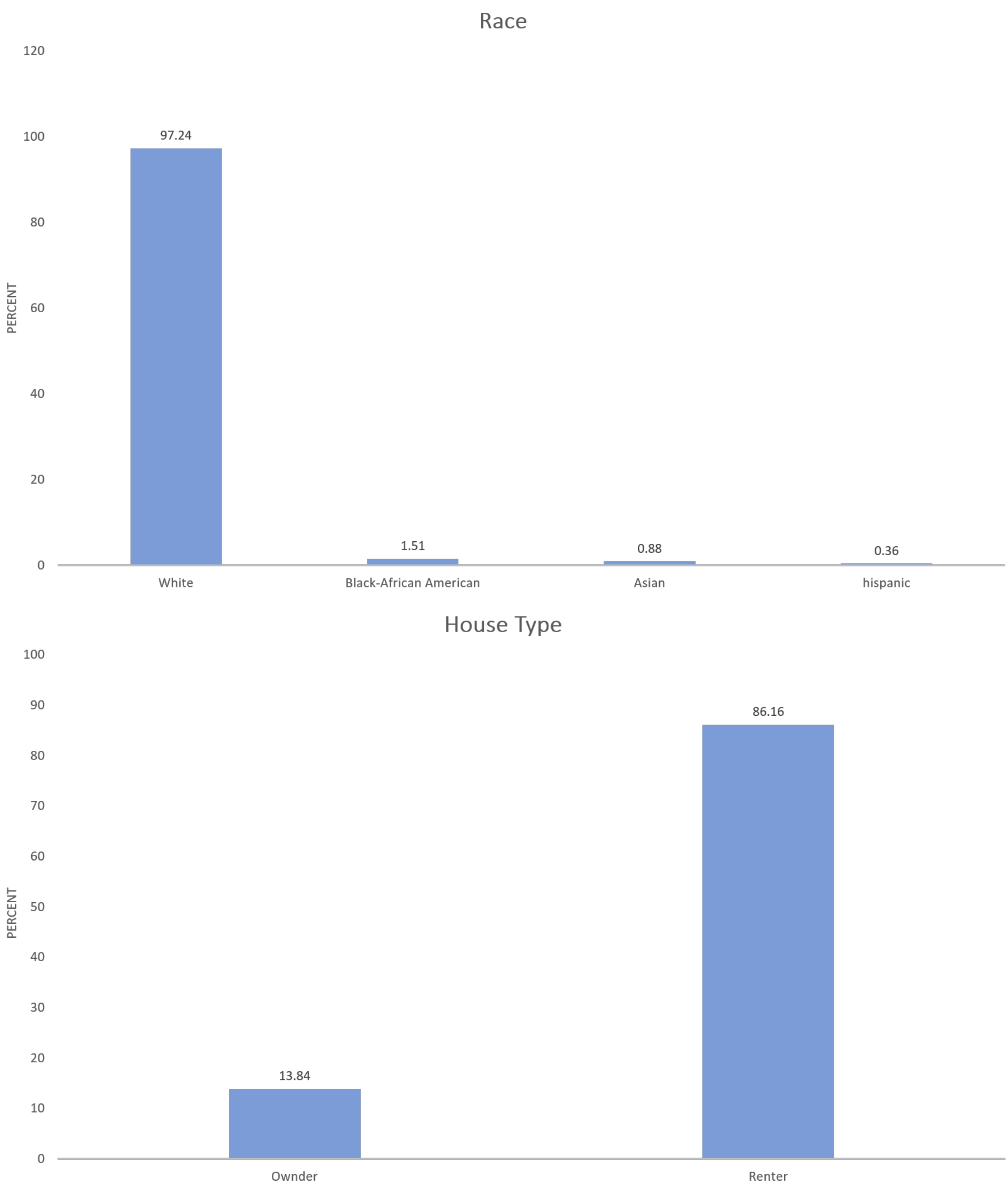


Figure 4.3: Household Demographic Factors (continued)

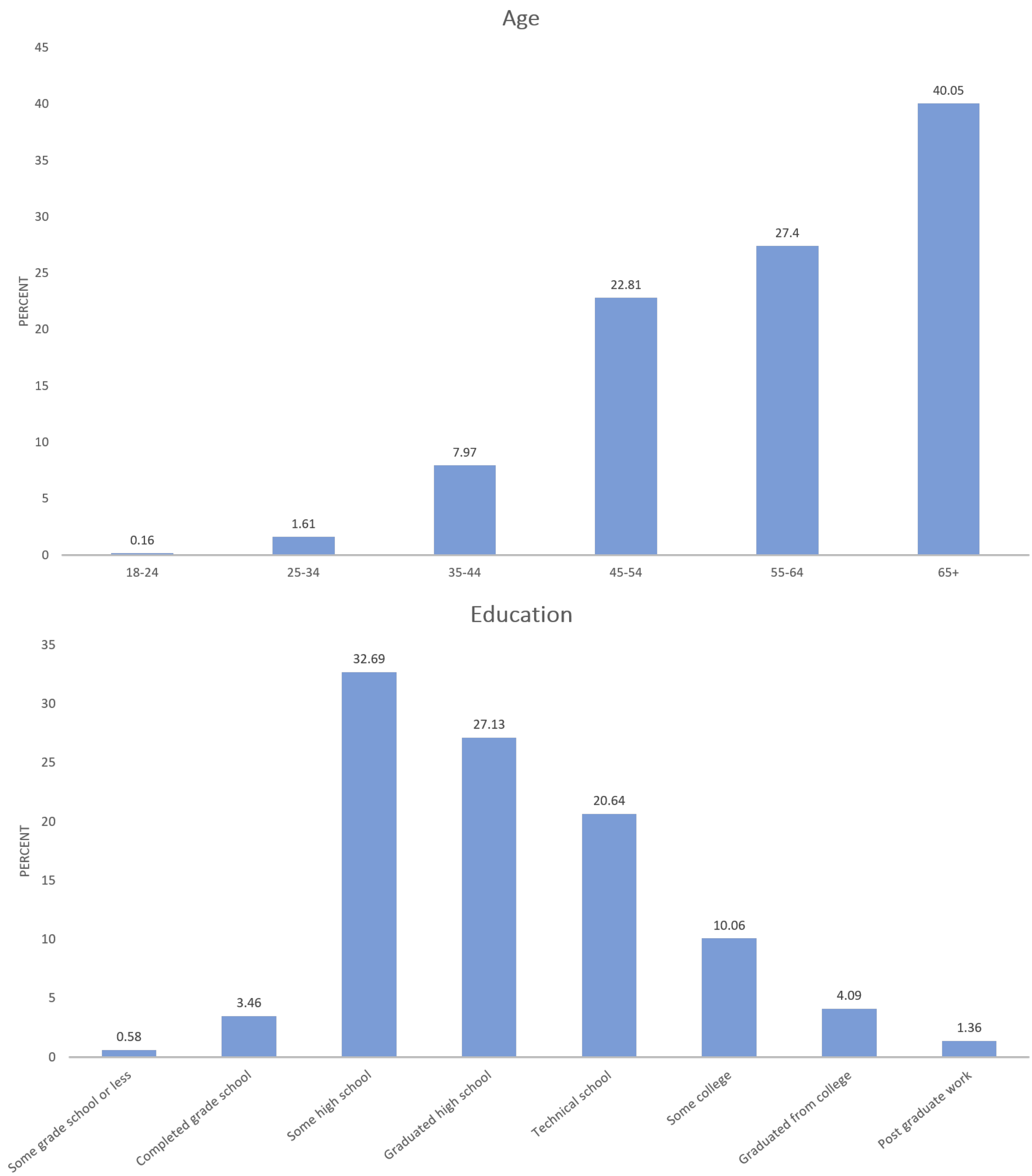
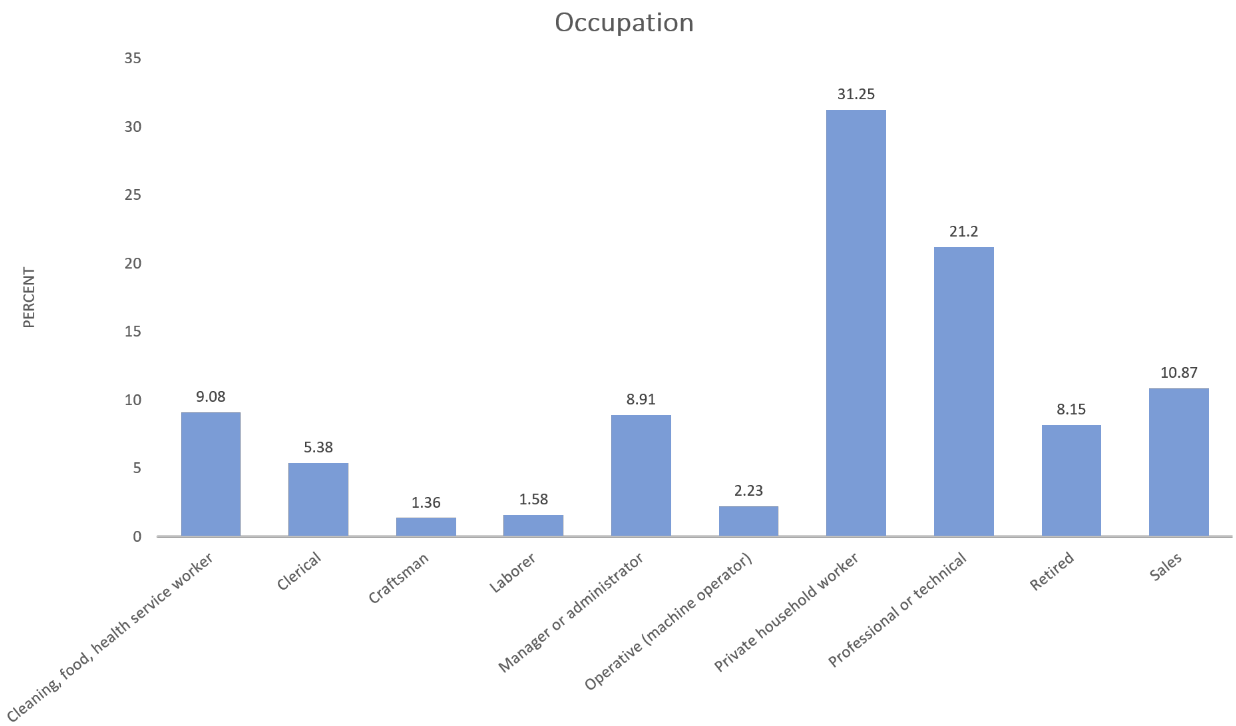


Figure 4.4: Household Demographic Factors (continued)



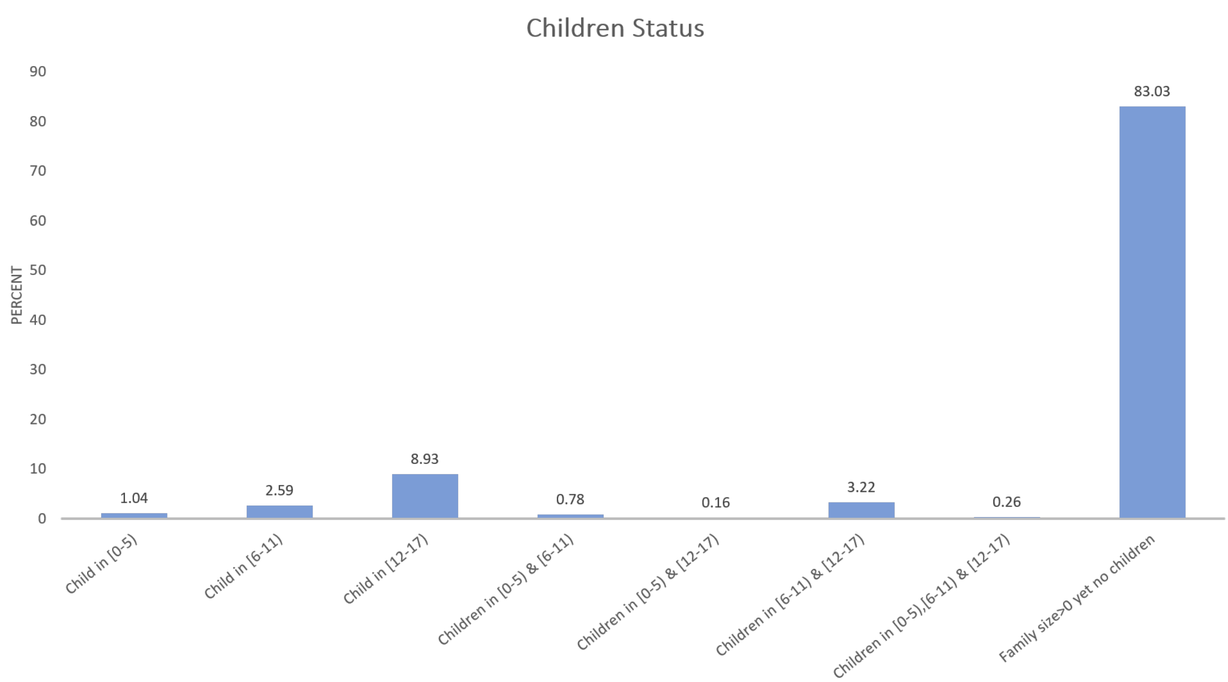


Figure 4.5: Household Demographic Factors (continued)

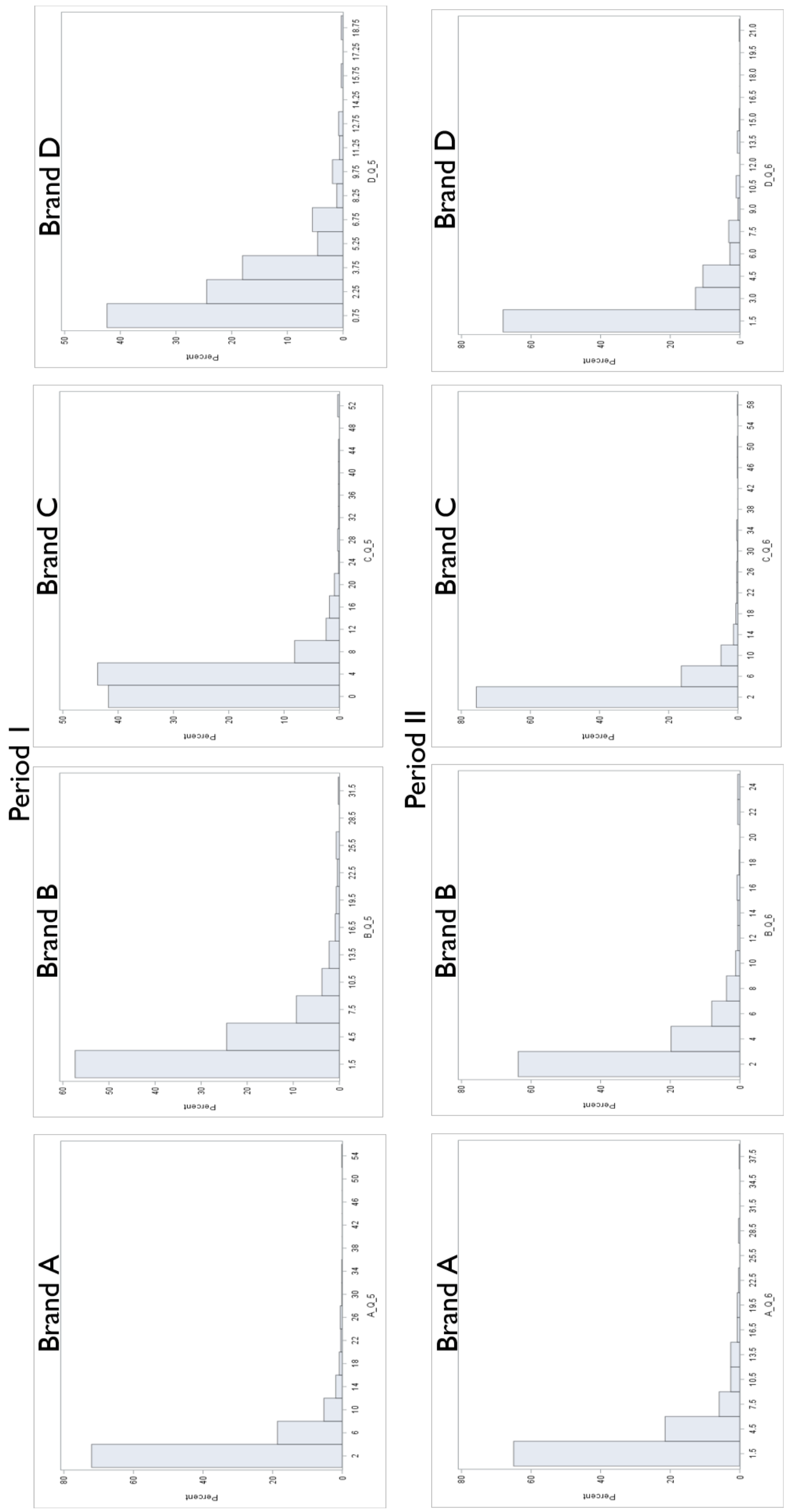


Figure 4.6: Histogram of Facial Tissue Purchase Quantity by Brand (2011)

Chapter 5

Conclusion

This dissertation investigates several innovative methods to improve the estimation and explanation of discrete choice models from the view of multivariate probability distributions. For consumers' decisions in a single discrete choice observation, the first two chapters explore the approximation of the choice probabilities and the distribution of mean willingness to pay in one of the most flexible models – the mixed logit model. For the aggregation of consumer choice in repeated discrete choices, the third chapter builds up an incomplete demand system with count outcome to analyze consumer brand choice in a utility consistent framework.

In the first chapter, we explore the potential of the quadrature approach in the estimation of mixed logit models, validates the feasibility and the satisfactory performance of Gauss-Hermite quadrature in approximating the choice probabilities especially when the number of random parameters is relatively small (≤ 6). We draw our conclusions from empirically comparing the estimation accuracy and efficiency among three candidate methods: quasi-Monte Carlo simulation, Gauss-Hermite quadrature, and spare grid integration. For the most commonly adopted estimation method, quasi-Monte Carlo simulation, we find that the problems of slow convergence rates and unstable accuracy are not negligible – the ultimate question becomes how much simulation error is acceptable, and which (quasi-) random sequence will work the best. Further, replicability is compromised for quasi-Monte Carlo unless the analyst has the identical, scrambled draws. These issues can be avoided completely by instead adopting almost exact ML estimation using Gauss-Hermite quadrature. Given its ease of application and replication, it is a powerful alternative to maximum simulated likelihood as it avoids simulation bias and simulation noise and only incurs controllable approximation error – certainly this is a desirable trade off. In terms of the other numerical integration method, sparse grid, estimation accuracy appears questionable plus the integrated choice probability is not always positive. In conclusion, for typical problems in environmental and natural resources economics with a relatively small number of random coefficients and modest sample sizes, we recommend Gauss-Hermite quadrature as worthy of consideration.

In the second chapter, we discuss the estimation of mean Willingness to Pay (WTP) in the mixed logit model under various distributional assumptions and calculation approaches. Different than most former studies which mainly focus on the statistical features of WTP, we advocate carefully examining the WTP ratio with a full consideration of the underneath economic assumptions. By narrowing down the scope to the normal goods and ruling out the possibility of zero consumption for all discrete choice questions, we first validate the rationality of a positive marginal utility of income in an economic sense – the condition to ensure stable moment approximations of the WTP as a ratio of two normally distributed variables. Further, we apply both the normal and log-normal distributions to the cost coefficient in our estimation of the mixed logit model.

The empirical analysis in the second chapter first shows the limitation of the Delta method in determining the standard error of mean WTP given it is equivalent to infinitesimal jackknife, a limiting form of the delete-1 jackknife, which assumes the mutual independence between all observations. We argue that the block delete jackknife has a more reasonable assumption of independence among respondents rather than among questions (observations). Further, our results reveal the skewness caused by applying the log-normal distribution to the cost coefficient. Although the sign-restricted distributions like the log-normal distribution can avoid negative or zero cost coefficient in statistical theory, the strong assumption on the shape of marginal utility of income may not fit the empirical data well and thus should be carefully considered. Finally, we validate the performance of the Bayesian individual-level WTP approach in capturing the distribution of mean WTP.

In addition to the empirical analysis with real survey data sets, a synthetic data set with 2000 respondent is simulated and further validates the feasibility of a normally distributed cost coefficient. Although the problem of non-existence of moments may still exist in the view of statistics, we found both theoretical and empirical evidence that the approximation distribution of the ratio performs well and follows a normal distribution.

In the third chapter, we analyze consumer brand choice through an incomplete demand system with count outcomes following the multivariate Poisson-Log normal (MPLN) distribution. Different than most existing demand models using market shares as the dependent variable, our approach targets the purchase amount – the non-negative integers – with the MPLN distribution given 1) its merit in allowing flexible covariance structures among counts, 2) the good performance in fitting over-dispersion, and 3) its advantage in accommodating both zero and non-zero consumption. Specifically, we emphasize the importance of utility consistency by specifying the mean of purchase counts under a log-linear demand form in the incomplete demand system. Further, Gauss-Hermite quadrature is applied to approximate the likelihood with multidimensional integration.

In the empirical analysis, we extend the application area of the count data demand model to the real transaction (scanner) data set at retail level. This extension helps to provide more insights on consumer preferences and market structure. Using a panel of scanner data provided by the IRI marketing data set, we analyze the brand choice of 1927 static consumers on facial tissue market in Eau Claire, Wisconsin in 2011. Our model accommodates well the over-dispersed purchase count, and provides informative insights on the market structure and consumer preferences. Among the four major brands that accounts for more than 97% of the market share, we find the demand of the cheapest brand to be quite stable and does not significantly respond to price changes. Also, households with higher incomes have higher probabilities of switching brands. Besides, our model is helpful in identifying the most promising consumers for a specific brand. For example, we find two brands that are significantly more welcomed by the elder group, while household size and kids status does not have significant influence on the expected demand of facial tissue. In summary, the third chapter contributes to the literature by enhancing the applicability of the multivariate count data model in a utility-consistent framework. It also enriches existing demand system analysis by further developing models with more types of outcomes.

Appendix A

Details in Sparse Grid Integration

The key idea of Sparse Grid integration is to carefully choose and re-weight the evaluation nodes to avoid the "curse of dimensionality" in multivariate quadrature. First, the most common approach to approximate a multi-dimensional integral is to combine the univariate quadrature rules in a tensor product approach. For a D dimensional integration with variable $\mathbf{x} = [x_1, x_2, \dots, x_D]$,

$$Q(g) := Q_1 \otimes Q_2 \otimes \cdots \otimes Q_D[g] \quad = \sum_{x_1 \in \mathbb{X}_{i_1}} \sum_{x_2 \in \mathbb{X}_{i_2}} \cdots \sum_{x_D \in \mathbb{X}_{i_D}} g(x_1 \cdot x_2 \cdot x_D)(w_1 w_2 \dots w_D)$$

With this approach, the total number of nodes is the product of the number of nodes in each dimension. For example, for a 10-dimensional integral with 20-quadrature nodes in each dimension, $20^{10} = 1.024e + 13$ nodes will be needed under the tensor production rule. That is, the total number of grid-points would be D^n for a D-dimensional integration with n nodes in each dimension. This is known as the "curse of dimension" where the computational complexity grows exponentially with the dimension. Researchers (Mysovskikh, 1968) have shown that not all the nodes in the tensor rule are necessary for the numerical integration. One method to identify the essential evaluations nodes is called sparse grid integration.

We consider a D-dimensional polynomial with total degree d as the sum of all exponents: $p_D^{(d)} \in \mathbb{P}_D^d := span\{x_1^{j_1} \cdot x_2^{j_2} \cdot \dots \cdot x_D^{j_D} | j_1 + j_2 + \cdots + j_D = d\}$. Also, it is defined that the cubature rule $Q[\cdot]$ is exact for the polynomial p_D^d if

$$
\begin{aligned}
Q[p_D^{(d)}] &= \int_{\Omega_1} \cdots \int_{\Omega_D} p_n^d(x_1, \ldots, x_D)\phi(\mathbf{x})d\mathbf{x} \\
&= \sum_{x_1 \in \mathbb{X}_{i_1}} \sum_{x_1 \in \mathbb{X}_{i_2}} \cdots \sum_{x_D \in \mathbb{X}_{i_D}} (w_1 \cdot w_2 \cdot \cdots \cdot w_D) \times p_n^{(d)}(x_1 \cdot x_2 \cdot \cdots \cdot x_D) \\
&\Rightarrow Q[p_D^{(d)}] = I[p_D^{(d)}]
\end{aligned}
$$

Further, define that quadrature or cubature formula has polynomial exactness (or degree of precision) of d if it is exact for all polynomials whose (total) degree are less or equal to d. Mysovskikh (1968) found that the number of grid-points (nodes) required to achieve the polynomial exactness of d for a cubature formula $Q[\cdot]$ is $[\binom{D+[d/2]}{[d/2]}, \binom{D+d}{d}]$. More specifically, the tensor product rule with D^n nodes would achieve the degree of exactness of $\max_{1 \leq i \leq n}\{2N_i - 1\}$, which is much higher than the true degree of the integrand we have. For example, consider a 2 dimensional integration with the degree of the polynomial equal to 3 (like $x_1 x_2^2$), the tensor product rule is not only exact for polynomials whose degree less or equal to 3 ($x_1, x_2, x_1^2 x_2, x_1 x_2^2, x_1^3, x_2^3,$), but also exact for those with higher degress of exactness ($x_1^3 x_2, x_1^3 x_2, x_1^2 x_2^2, x_{1 x_2}^{2\ 3}, x_1^3 x_2^2, x_1^3 x_2^3$). If we have node $n = 10$ for each dimension, the tensor produce rule has a polynomial exactness $2 \times 10 - 1$=19 with the $10^2 = 100$ nodes, while the minimum number of nodes to achieve the same exactness is $\binom{2+[19/2]}{[19/2]}$=55. Expand the dimension to 3, the minimum number of nodes is $\binom{3+[19/2]}{[19/2]}$=220 so that $10^3 - 220 = 780$ nodes are not necessary. With the dimension increase, tensor product rules generate exponentially unnecessary nodes.

To "smartly" select the nodes, Smolyak (1963) provides a general method for the multivariate extensions of univariate operators. Wasilkowski and Wozniakowski (1995) further develop the integration rule as below. With a given accuracy level $k \in \mathbb{N}$ for a $D-$dimensional integration, the sparse-grid integration rule can be expressed as:

$$
\begin{aligned}
S_{D,k}[g] &= \sum_{q=k-D}^{k-1} (-1)^{k-1-q} \binom{D-1}{k-1-q} \sum_{\mathbf{i}\in\mathbb{N}_q^D} (Q_{i1} \otimes \cdots \otimes Q_{iD})[g] \\
&= \sum_{q=k-D}^{k-1} \sum_{\mathbf{i}\in\mathbb{N}_q^D} \sum_{x_1\in\mathbb{X}_{i_1}} \cdots \sum_{x_D\in\mathbb{X}_{i_D}} g(x_1, \ldots, x_D)(-1)^{k-1-q} \binom{D-1}{k-1-q} \Pi_{d=1}^{D} w_{i_d}(x_d)
\end{aligned}
$$

where

$$
\mathbb{N}_q^D = \{\mathbf{x} \in \mathbb{N}^D : \sum_{d=1}^{D} i_d = D + q\}
$$

which means that we allocate the accuracy level of the quadrature to the exponent of each variable to select the nodes. For example, let's set accuracy level $k = 5$ for an integration with dimension $D = 2$, we have $q \in \{3, 4\}$. Then for $q = 3$, $\mathbb{N}_3^2 = \{(i_1 = 1, i_2 = 4), (i_1 = 2, i_2 = 3), (i_1 = 3, i_2 = 2), (i_1 = 4, i_2 = 1)\}$, for $q = 4$, $\mathbb{N}_4^2 = \{(i_1 = 1, i_2 = 5), (i_1 = 2, i_2 = 4), (i_1 = 3, i_2 = 3), (i_1 = 4, i_2 = 2), (i_1 = 5, i_2 = 1)\}$. Thus, the nodes with the accuracy level i_d for each variable x_d was adopted and combined to each other as nodes for the integration. Also, the part $(-1)^{k-1-q}\binom{D-1}{k-1-q}\Pi_{d=1}^{D} w_{i_d}(x_d)$ is the weight assigned to each combination of nodes. Note that the weight could be negative, so that theoretically it is possible to have a negative approximate integral although the integrand is positive everywhere (Heiss and Winschel, 2008). We address this question in the analysis of mixed logit models with real survey data.

Appendix B

Application of Gauss-Hermite Integration

With the Gauss-Hermite integration, we can approximate any integral with the form

$$\int_{-\infty}^{+\infty} g(\varepsilon)d\varepsilon_h = \int_{-\infty}^{+\infty} e^{-\varepsilon^2}\ h(\varepsilon)d\varepsilon$$

as the weighted average of the evaluation point w_h, that is

$$\int_{-\infty}^{+\infty} e^{-\varepsilon^2} f\left(\varepsilon\right) d\varepsilon \approx \sum_{h=1}^{d} w_h f(\varepsilon_h).$$

Here the approximation is defined by a Hermite orthogonal polynomial of degree d, $H_d(\varepsilon)$, with associated weights w_h $(h = 1, 2, \ldots, d)$. For a standard normal random variable, a change of variable results in

$$(2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{-\varepsilon^2/2} f\left(\varepsilon\right) d\varepsilon \approx \sum_{h=1}^{d} w_h^* f(\varepsilon_h^*)$$

where $\varepsilon_h^* = \sqrt{2}\varepsilon_h$ and $w_h^* = w_h/\sqrt{\pi}$. Note that $\Sigma w_h^* = 1$.

For the mixed logit model defined in the text, we have four possibly corelated parameters: $\beta_{price}, \beta_{range}, \beta_{electric},$ and β_{hybrid} within the indirect utility function:

$$\begin{aligned} V_{ijt} =& \beta_{price} Price_{ijt} + \beta_{Range} Range_{ijt} + \beta_{Electric} I_{Electric_{ijt}} + \beta_{Hybrid} I_{Hybrid_{ijt}} + \beta_{Perf1} I_{Perf1_{ijt}} + \epsilon_{ijt} \\ =& f_1(\mu_{price} + \sigma_{price}\varepsilon_{price}) Price_{ijt} + f_2(\mu_{Range} + \sigma_{Range}\varepsilon_{Range}) Range_{ijt} \\ &+ f_3(\mu_{Electric} + \sigma_{Electric}\varepsilon_{Electric}) I_{Electric_{ijt}} + f_4(\mu_{Hybrid} + \sigma_{Hybrid}\varepsilon_{Hybrid}) I_{Hybrid_{ijt}} \\ &+ \beta_{Perf1} I_{Perf1_{ijt}} + \epsilon_{ijt} \end{aligned}$$

Given the $\epsilon_{ijt} \sim EV(0, 1)$, the probability for respondent i to choose alternative T in question j is a 4-dimensional integration towards the logit form probability

$$P_{ij(T)} = \int_{R^4} \frac{\exp(V_{ij(T)})}{\sum_{t=1}^{3} \exp(V_{ijt})} \frac{exp(-1/2E\Sigma^{-1}E')}{(2\pi)^{4/2}|\Sigma|^{1/2}} d_{price} d_{Range} d_{Electric} d_{Hybrid}$$

where $E = [\varepsilon_{price} \;\; \varepsilon_{Range} \;\; \varepsilon_{Electric} \;\; \varepsilon_{Hybrid}]$ and Σ is the variance-covariance matrix of E.

To extend the Gauss-Hermite integration for the mixed logit model, we first define the set H as the Cartesian product $H = \varepsilon^*_{price} \times \varepsilon^*_{Range} \times \varepsilon^*_{Electric} \times \varepsilon^*_{Hybrid}$ (H is of dimension d^4 by 4), and define the set $W = w^*_h \times \; w^*_h \times \; w^*_h \times \; w^*_h$. Also, we let w_4 be the product of the columns of W such that it is now a d_4 by 1 vector whose sum is one. Finally, we define $E^* = HS$ where S is the (upper triangular) Cholesky decomposition of Σ.

We now define

$$\begin{aligned} V^*_{ijt} = & f_1(\mu_{price} + E^*_{price}) Price_{ijt} + f_2(\mu_{Range} + E^*_{Range}) Range_{ijt} + f_3(\mu_{Electric} + E^*_{Electric}) I_{Electric} \\ & + f_4(\mu_{Hybrid} + E^*_{Hybrid}) I_{Hybrid} + \beta_{Perf1} I_{Perf1} + \epsilon_{ijt} \end{aligned}$$

where E^*_{price}, E^*_{Range},$E^*_{Electric}$, E^*_{Hybrid} are the corresponding columns of E^*. Then the probability that respondent i obtained from choosing alternative T in question j could be rewrote as:

$$P_{ij(T)} \approx w_4' \frac{exp\left(V^*_{ij(T)}\right)}{\sum_{t=1}^{3} exp\left(V^*_{ijt}\right)}$$

By assuming the ten questions answered by the same respondent are independent from each other, the joint probability of respondent i's choices for the ten questions is the product of the probability for each questions chosen alternative.

$$P_{ij(T)} \approx w_4' \left[\prod_{j=1}^{10} \frac{exp\left(V^*_{ij(T)}\right)}{\sum_{t=1}^{3} exp\left(V^*_{ijt}\right)} \right]$$

Under the assumption that each respondents choices are independent of each other, the joint probability of all respondents choice sets is the product of all the 100 respondents probability on the ten questions, and thus the maximum likelihood estimation could be applied.

For the estimation, we choose Hermite orthogonal polynomial of degree 24, with the corresponding abscissas and weights listed in Table B.1 of the appendix. The largest weight

$(w_i^*)^4$ for abscissas is 0.033 while the smallest weight is only $7.67E - 64 (< 0.000001)$ and quite close to 0. Note that the weight w_i^* decreases rapidly as the absolute value of abscissas increasing. Considering the probability in mixed logit model would always be less than 1 for any given abscissas, we trimmed those abscissas whose weight is less than one-tenth of the mean weight of all evaluation points. That is, abscissas with weight is less than $3.01408E - 07 \ \ (= 1/(24^4 \times 10))$ are dropped. In this way, we greatly reduce the estimation point from 331776 to 10416, and only paying a very small cost in losing a total weight that is smaller than $0.09686 \ \ (= (331776 - 10416) \times 3.01408E - 07))$, which is a negligible proportion compared to the sum of all weight as 1. At the same time, we rescale the weights for remaining abscissas to assure they sum to one. Such an approach contributes significantly to reduce evaluation points and to increase the estimation efficiency. At the same time, we are able to ensure the 10416 evaluation points is a good representative of the shape of the multivariate normal/lognormal distribution.

Table B.1: 24^{th} Degree Abscissas and Weights

i	$\pm\varepsilon_i^*$	w_i^*	$(w_i^*)^4$	
1	-6.0159	1.66E-16	7.67E-64	(<0.000001)
2	-5.2594	6.58E-13	1.88E-49	(<0.000001)
3	-4.6257	3.05E-10	8.61E-39	(<0.000001)
4	-4.0537	4.02E-08	2.61E-30	(<0.000001)
5	-3.5200	2.16E-06	2.17E-23	(<0.000001)
6	-3.0125	5.69E-05	1.05E-17	(<0.000001)
7	-2.5239	8.24E-04	4.60E-13	(<0.000001)
8	-2.0490	0.0070	2.47E-09	(<0.000001)
9	-1.5843	0.0374	1.97E-06	(0.000002)
10	-1.1268	0.1277	2.66E-04	(0.00027
11	-0.6742	0.2862	6.71E-03	(0.0067)
12	-0.2244	0.4269	3.32E-02	(0.033)
13	0.2244	0.4269	3.32E-02	(0.033)
14	0.6742	0.2862	6.71E-03	(0.0067)
15	1.1268	0.1277	2.66E-04	(0.00027
16	1.5843	0.0374	1.97E-06	(0.000002)
17	2.0490	0.0070	2.47E-09	(<0.000001)
18	2.5239	8.24E-04	4.60E-13	(<0.000001)
19	3.0125	5.69E-05	1.05E-17	(<0.000001)
20	3.5200	2.16E-06	2.17E-23	(<0.000001)
21	4.0537	4.02E-08	2.61E-30	(<0.000001)
22	4.6257	3.05E-10	8.61E-39	(<0.000001)
23	5.2594	6.58E-13	1.88E-49	(<0.000001)
24	6.0159	1.66E-16	7.67E-64	(<0.000001)