Aggregation Bias in Discrete Choice Models with an Application to Household Vehicle Choice

Timothy Wong*, David Brownstone* and David Bunch†

May 20, 2015

Abstract

This paper studies the practice of aggregating choices within discrete choice models. Researchers often do not observe choices at the exact level they are made, and hence aggregate choices to the level that is observed. Modeling choices at a fine level of detail can also lead to large choice sets that exceed the practical capabilities for model estimation. However, the practice of aggregation misspecifies the true choice set of interest. We investigate this concern within the context of the Berry, Levinsohn, and Pakes (BLP) choice model for micro- and macro-level data. We compare the practice of aggregating choices to specifications from two papers that address these concerns (McFadden, 1974; Brownstone and Li, 2014), with application to vehicle choice data. We find that aggregation affects both the point estimates and standard errors obtained from the model. In particular, standard errors are smaller with aggregation. This result has significant empirical implications. Discrete choice models are widely used to estimate consumer valuation of fuel efficiency, a quantity that is relevant to energy analysts concerned that consumers undervalue fuel efficiency technologies (the "energy paradox"). If so, then there is space for policies that increase adoption of such technologies. However, estimates of consumer valuation across vehicle choice studies are inconclusive. The findings of this paper suggest that this disparity may be partly explained by the practice of aggregating choices. In addition, the BLP model applied here is usually estimated sequentially, and the standard errors derived from this process are inconsistent. Thus, this paper also derives consistent standard errors for the model and examines their performance compared to the sequential standard errors that are commonly used.

Keywords: Choice-set aggregation; Discrete choice; Energy paradox; Fuel efficiency; Vehicle choice.

---

* Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100; e-mail: tcwong1@uci.edu, email: dbrownst@uci.edu. † Graduate School of Management, University of California, Davis, Gallagher Hall, One Shields Avenue, Davis, California 95616; e-mail: dsbunch@ucd.edu. We thank Jiawei Chen, Michael Guggisberg, Phillip Li, Alicia Lloro, Jennifer Muz, Kevin Roth, Kenneth Small and Angela Vossmeyer for their helpful comments. Feedback from the California Econometrics Conference 2014, the University of California Irvine Econometrics Seminar and the International Choice Modelling Conference, 2015are also appreciated. Funding from the UC Center for Energy and Environmental Economics; the UC Multi-Campus Research Program in Sustainable Transport: Technology, Mobility, and Infrastructure, the UC Center on Economic Competitiveness in Transportation, and the University of California, Irvine, Department of Economics is greatly acknowledged.

1. Introduction

Multinomial choice models have become popular in demand estimation because, unlike systems of demand equations, the number of parameters to be estimated is not a function of the number of products, removing the obstacle of estimating markets with many differentiated products. One challenge of choice modeling in applications is determining the level of detail at which the choice set is defined. Modeling choices at their finest level can quickly cause the resulting choice set to grow so large that it exceeds the practical capabilities of estimation. In addition, it is common in micro-level datasets that household choices are not observed at the finest level possible. In such cases, researchers often aggregate choices to the level of detail that is observed. For example, when modeling a household's vehicle choice decisions, the exact choice set may contain vehicles at the make-model-trim level. Assume households decide between four vehicles: the Honda Civic LX, Honda Civic EX, Toyota Corolla G, and Toyota Corolla X; however, the researcher only observes the household's choice between two broad make-model groups, the Honda Civic and the Toyota Corolla. Hence, the researcher aggregates the exact choice set to the make-model level. In this case, the researcher creates an "average" Honda Civic whose characteristics consist of the average attributes of the Civic LX and Civic EX, such as horsepower and curb weight, and likewise for the Corolla G and Corolla X.

Despite these practical considerations, aggregation of the choice set to broader levels removes useful variation from the choice set and misspecifies the choices faced by households, which can diminish researchers' ability to understand key choice determinants. This paper studies the effect of choice set aggregation on estimates derived from choice models, applying this practice to a common policy concern regarding vehicle choice. These effects are studied within the context of a Berry, Levinsohn and Pakes (henceforth, BLP) model for micro- and macro-level data estimated with a maximum likelihood approach. This model bears strong similarities to the models applied in Goolsbee and Petrin, 2004; Chintagunta and Dubé, 2005; Train and Winston, 2007; Langer, 2014 and Whitefoot et al., 2014. Specifically, we discuss two model specifications within this BLP framework that help overcome this concern of aggregation: an aggregation correction by McFadden, 1978, and a model for broad choice data by Brownstone and Li, 2014. McFadden's aggregation correction places a distributional assumption on the elements within each aggregated choice and uses the higher moments of the distribution in the utility specification. Brownstone and Li, 2014, define the choice probability of a broad group as the sum of the choice probabilities of the elements within that group, leveraging on the existence of aggregate market share data at the exact choice level for identification.

These results are particularly meaningful for evaluating the literature on vehicle choice. Multinomial choice models are commonly applied to vehicle choice data with interest in estimating households' willingness to pay for fuel efficiency. These estimates are important because they help measure the extent that households do not value energy saving investments like vehicle fuel efficiency as standard economic theory on rational behavior would predict. This energy paradox has significant consequences. If households consistently undervalue energy saving investments, this implies that households are incurring excess environmental costs, even if proper mechanisms (e.g. carbon taxes,) are in place to control for environmental externalities. Therefore, implementing government policies that increase the energy efficiency of products, (e.g. energy labelling programs, minimum energy efficiency standards) may be in the best interest of both households and the environment. Despite the important implications of understanding the energy paradox, the existing literature on willingness to pay for fuel efficiency is inconclusive (Greene, 2010; Helfand and Wolverton, 2011). One potential source of the disparity in estimates could be the common practice of aggregating the exact vehicle choice set to broader groups based on vehicle class (Goldberg, 1998; Bhat and Sen, 2005; Bento et al., 2009; Jacobsen, 2013).

I find that aggregation affects both the point estimates and standard errors of parameters from the model. Estimates of household willingness to pay for vehicle fuel efficiency are twice as large when aggregation is modeled using McFadden's method, and four times smaller when the fully disaggregated alternatives are modeled using the Brownstone and Li approach. The Brownstone and Li approach better suits the current application because it is harder to justify the necessary distributional assumptions of McFadden's approach in the current setting. Perhaps more importantly, model estimates are more likely to appear statistically significant when choices are aggregated without correction because doing so ignores the measurement error it introduces to the choice attributes. Estimates of household willingness to pay for vehicle fuel efficiency are significant in the model that aggregates choices but not significant in the other two models. If these results extend to existing vehicle choice studies that aggregate choices, it suggests that current disparities on the energy paradox may in part stem from overconfidence in model estimates because of a failure to account for noise induced by choice set aggregation.

A series of papers study the concern of choice set aggregation as it applies to choice of recreation sites (Parsons and Needelman, 1992 Feather, 1994; Haener et al., 2004). Feather, 1994, implements a variant of the McFadden approach to modeling aggregation and finds that welfare estimates are much closer to those obtained from an exact choice model when aggregation is accounted for. Haener et al., 2004, find that only accounting for the number of exact choices within each broad group improves model fit and alleviates aggregation bias. Accounting for the heterogeneity of choice attributes within the broad group in addition to the number of exact choices within the broad group did not provide significant improvements in their application.

Spiller, 2012, studies aggregation within a discrete-continuous model of household vehicle fleet choice and utilization. She compares estimates from a model that aggregates vehicles to make-class categories to a model that defines vehicles at the make-model level, finding that demand for gasoline is more inelastic when choices are aggregated to broader levels.

Three recent papers define vehicles at the finest levels of detail possible when estimating household willingness to pay for fuel efficiency, though none of these papers employ choice models to household-level data in their work. Busse et al., 2013 employ a reduced form approach to study the effect of fuel prices on vehicle transaction prices, market shares and sales for new and used vehicles of different fuel economies while Allcott and Wozny, 2014 and Grigolon et al., 2014 employ choice models to macro-level data. Busse et al., 2014 do not find evidence that households undervalue fuel efficiency. Grigolon et al., 2014, find only modest undervaluation of fuel costs, while Allcott and Wozny, 2014, find a wide range of values varying from undervaluation to rational valuation, depending on various model assumptions.

These findings on choice set aggregation are also relevant to many other empirical questions that choice models are used to study, including the effect of mergers (Nevo, 2000), welfare gains from new products (Petrin, 2002, Goolsbee and Petrin, 2004), measuring market power (Nevo, 2001), and explaining the declining market share of dominant firms (Train and Winston, 2007).

The BLP model for micro- and macro-level data is often estimated sequentially. In the first stage, a choice model is fit where product specific constants are estimated. These constants are then used as dependent variables in an instrumental variables (IV) regression in the second stage. The standard errors from this two-step process are inconsistent because they ignore 1) the constraint that estimated shares equal macro-level market shares and 2) the correlations that exist between estimates across the two stages. Though the model has been applied to data for more than a decade, consistent standard errors for both stages of the model have never been formally derived. Previous studies that implement this model either correct only the standard errors of first stage parameters (BLP,2004; Chintagunta and Dubé, 2005), use inconsistent standard errors, or never formally state how they obtain the standard errors of model estimates.

Thus, in this paper, we also derive consistent standard errors for the entire model by recasting each sequence of the model jointly within a Generalized Method of Moments (GMM) framework. We examine the performance of both the inconsistent and consistent standard errors through a Monte Carlo study and the vehicle choice application. We show that the use of standard errors derived from sequential estimation can cause errors when hypothesis testing. Second stage standard errors from sequential estimation are downward biased. We also find that the first stage standard errors are too small. In application, the standard errors derived from the GMM framework are larger than the sequential standard errors, generally by a factor of 1.3 to 13. However, the GMM standard errors for fuel costs are larger by a staggering 14 to 43 times. These findings also relate to the disparate results across existing choice models on the energy paradox. Studies that draw inference using inconsistent standard errors (BLP, 2004; Train and Winston, 2007; Whitefoot et al, 2014) may portray estimates as more significant than they really are.

Both the concern of choice set aggregation and the inconsistency of the standard errors from sequential estimation are independent of the random coefficients framework typically included in the BLP model. Therefore, we simplify the common BLP model by removing the random coefficient specification. Including random coefficients will result in slightly larger standard errors due to sampling noise from simulated maximum likelihood estimation, and, if correctly specified, also result in richer substitution patterns across products.

This paper is structured as follows. In Section 2, we present the BLP model for micro- and macro-level data. In Section 3, we detail the maximum likelihood approach to estimating the model, including the inconsistent and consistent methods of obtaining standard errors. In Section 4, we demonstrate through a Monte Carlo experiment the inconsistency of standard errors obtained through sequential estimation and show that the consistent standard errors provide the appropriate coverage probabilities of the true value in the limit. In Section 5, we present the McFadden, 1978, correction for aggregation bias and the Brownstone and Li, 2014, model for broad choice data, and adapt these models to the BLP setting at hand. In Section 6, we apply the consistent standard error formula and aggregation corrections to a vehicle choice application and discuss the results, relating them to the existing literature on the energy paradox. Section 7 summarizes and concludes.

2. The BLP Model

The seminal BLP choice model (BLP, 1995) contributes an important innovation to the choice modeling literature by accommodating the endogeneity of product attributes, since the econometrician rarely observes the full set of attributes of products that induce households to make purchases. In BLP, 1995, the model is applied using only macro-level data. BLP, 2004, extends the model to applications that combine both micro- and macro-level data. In this paper, BLP addresses an important concern in micro-level choice modeling, the issue of unrepresentative sampling. To overcome this concern, they supplement their unrepresentative choice dataset with aggregate market share data that is believed to be more representative.

The BLP model for micro- and macro-level data used here is similar to BLP, 2004. The key difference is that BLP, 2004, incorporates household information into their model by constructing a moment that captures the covariance of product attributes and household attributes. The model here incorporates household information through a standard multinomial likelihood function, as is done in Train and Winston, 2007. The model is as follows:

Let $n = 1, \dots, N$ index households which can either purchase any of J products, $j = 1, \dots, J$ in the market or not purchase any product, characterized by selecting the "outside good", $j = 0$. The indirect utility of household n from the choice of product j, U_{nj} is assumed to follow the following linear specification:

$$U_{nj} = \delta_j + w_{nj}'\beta + \epsilon_{nj},$$

$$n = 1, \dots, N, \quad j = 0, 1, \dots, J,$$

where δ_j is a product specific constant that captures the "average" utility of product j. In other words, it is, the portion of utility from product j that is the same for all households. w_{nj} is a $(K_1 \times 1)$ vector of household attributes interacted with product attributes, β is its $(K_1 \times 1)$ vector of associated parameters, and ϵ_{nj} is an error term with mean zero that captures all remaining elements of utility provided by product j to household n. For the purpose of identification, average utility of the "outside good," δ_0, is normalized to zero. Households select the product that yields them the highest utility:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\, i \neq j \\ 0 & otherwise. \end{cases}$$

Assuming that ϵ_{nj} follows the type I extreme value distribution, the probability that household n, chooses product j, P_{nj} is:

$$P_{nj} = \frac{exp\left(\delta_j + w_{nj}'\beta\right)}{\sum_k exp(\delta_k + w_{nk}'\beta)}. \tag{1}$$

In the "traditional" conditional logit model, $\delta = \{\delta_1, \ldots, \delta_J\}$ and β, can be estimated by maximizing the following log-likelihood function:

$$L(y; \delta, \beta) = \sum_n \sum_j y_{nj}\, log\ (P_{nj}). \tag{2}$$

An interesting feature of maximum likelihood (ML) estimation of the multinomial logit is how δ is identified. The first order condition for δ is that the log likelihood function with respect to δ equals zero:

$$\frac{\partial L}{\partial \delta_j} = \sum_n \sum_i (y_{ni} - P_{ni})\, d_i^j = 0, \qquad \text{for } j = 1, \ldots, J,$$

$$where\ d_i^j = \begin{cases} 1 & if\ i = j \\ 0 & otherwise. \end{cases}$$

Rearranging and dividing both sides by N, we find that

$$\frac{1}{N} \sum_n y_{nj} = \frac{1}{N} \sum_n P_{nj}, \qquad \text{for } j = 1, \ldots, J.$$

The equation describes that the average utilities are estimated such that the predicted shares from the model, $\widehat{S_j} = \frac{1}{N} \sum_n P_{nj}$, match the in-sample shares, that is, the share of households in the sample who choose each of the products. (Train, 2009).

A key innovation of BLP, 2004, is that δ is estimated such that the predicted shares match aggregate market shares rather than in-sample shares. For $j = 1, \ldots, J$, δ_j is chosen such that

$$A_j = \widehat{S_J}$$

where A_j is the aggregate market share for product j. Berry (1994) shows that for any value of β, a unique δ exists such that the predicted shares match these aggregate market shares.

Matching predicted shares to aggregate market shares rather than in-sample shares improves estimation in the event that there is high sampling variance and hence, unrepresentative sample shares. In addition, by using aggregate market shares, average utilities can be estimated for products even if the in-sample shares for these products are zero.

Finally, it is assumed that the average utilities are a linear function of product attributes:

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$

where x_j is a $(K_2 \times 1)$ vector of exogenous product attributes, α_1 is a $(K_2 \times 1)$ vector of associated parameters, p_j is a $(K_3 \times 1)$ vector of product attributes that are endogenous with respect to average utility, α_2 is a $(K_3 \times 1)$ vector of associated parameters, and ξ_{1j} captures the average utility associated with attributes unobserved to the econometrician. Because p_j is endogenous with respect to average utility, it is correlated with unobserved attributes contained in ξ_{1j}, such that $E(\xi_{1j}|p_j) \neq 0$. There exists a set of instruments, z_j, that are correlated with p_j and uncorrelated with ξ_{1j}:

$$p_j = z_j{}'\gamma + \xi_{2j}$$

$$where\ E(\xi_{1j}|z_j) = 0.$$

For simplicity, we assume here that z_j is a $(K_3 \times 1)$ vector, (and therefore, γ is a $K_3 \times 1$ vector of associated parameters), that is, there are as many instruments as there are endogenous regressors, making the model just-identified. Although over-identification does not tremendously complicate estimation, with a just-identified model, optimal GMM methods are not necessary, which simplifies estimation.

In summary, the BLP model consists of the following equations:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\ i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + \mathrm{w_{nj}}'\beta + \epsilon_{\mathrm{nj}}, \quad \epsilon_{\mathrm{nj}} \sim \text{type I extreme value}$$

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$
$$p_j = z_j{}'\gamma + \xi_{2j}$$
$$A_j = \frac{1}{N}\sum_n P_{nj}$$

In many empirical settings, all observed attributes of households are categorical. For example, income is often only observed in categories of income ranges, and education in categories by highest level attained. When these types of data are used, and the number of household attributes is small compared to the number of households in the sample, it is common that some households in the sample are observed to have identical attributes. As long as these households are still distributed independently, it is possible to collapse households with identical attributes into representative "household types" that make repeated, independent choices. The repeated choices are simply an aggregation of the decisions made by each household in the sample, of that type. Let m denote "household type", where $m = 1, \dots, M$. Let r_m denote the number of households belonging to type m. It must be the case that $\sum_m r_m = N$. Finally, let l_m index households within a given household type. Then,

$$y_{mj} = \sum_{l_m=1}^{r_m} y_{nj}, \forall\, j = 0, 1, \dots, J\,.$$

It is easy to see that the aggregation of households to "household types" changes the outcome variable, because $\sum_j y_{mj} = r_m$ while $\sum_j y_{nj} = 1$. Nevertheless, aggregation leaves the likelihood function of the multinomial logit unchanged. There is a computational benefit from estimating at the "household type" level, rather than the household level since, as long as $r_m > 1$ for any m, then $M < N$.

To remain consistent with the Monte Carlo experiment and empirical applications that are conducted, the remainder of this paper assumes estimation with aggregation to "household types."

3. Estimation Procedure: A Maximum Likelihood Approach

Sequential estimation is conducted in two stages. The first stage involves estimating the average utilities, δ, and the parameters associated with the household-product interaction variables, β.

The first stage estimates of δ are then used in the second stage estimation of the parameters associated with the product attributes, $\alpha = [\alpha_1\ \alpha_2]$.

δ and β are estimated through an iterative process. Conditional on some initial value of δ, maximize equation (2) with respect to β to obtain conditional maximum likelihood estimates, $\hat{\beta}$. Conditional on $\hat{\beta}$, estimate δ using the contraction-mapping algorithm developed in BLP, 1995. This algorithm is itself an iterative process which yields the estimate, $\hat{\delta}$, when the following equation is iterated on until convergence:

$$\delta_{j,t+1} = \delta_{j,t} + \ln(A_j) - \ln(\widehat{S_j}), \quad \forall\, j = 1, \dots, J$$

The algorithm enforces the constraint that the predicted shares equal the aggregate market shares. The maximum likelihood and contraction mapping processes are repeated iteratively until convergence.

The second stage of the sequential process estimates the parameters associated with the product attributes, α. To do this, standard IV estimation is applied, substituting the converged values, $\hat{\delta}$, from the first stage, for the "true values" of the average utilities, δ. Let $X = [x_1'\ x_2'\ \dots\ x_J']'$. Let Z and P be similarly defined matrices containing the element, z_j and p_j respectively. Then the IV estimates are given by the familiar solution:

$$\hat{\alpha} = \left(\tilde{X}'Z(Z'Z)^{-1}Z'\tilde{X}\right)^{-1}\tilde{X}'Z\left(Z'Z\right)^{-1}Z'\hat{\delta}$$

where $\tilde{X} = [X\ \ P]$.

The standard errors for $\hat{\alpha}$ from IV estimation are downward biased. Standard IV estimation assumes that the dependent variable (in this case, δ) is known; however, in the BLP model, the dependent variable, $\hat{\delta}$, is an estimate. The standard errors for $\hat{\alpha}$ from IV estimation do not account for the variance inherent in $\hat{\delta}$ causing the standard errors of $\hat{\alpha}$ to be underestimated.

Additionally, in application, some researchers obtain the standard errors for $\hat{\beta}$ and $\hat{\delta}$ from the inverse of the Hessian of the logit likelihood function. However, $\hat{\beta}$ and $\hat{\delta}$ are not maximum likelihood estimates, since first stage estimation constrains predicted shares to match aggregate market shares, not in-sample shares. Unless the aggregate market shares equal in-sample shares, standard errors for $\hat{\beta}$ and $\hat{\delta}$ obtained in this manner are also inconsistent.

To obtain estimates of the correct standard errors for $\hat{\delta}, \hat{\beta}$ and $\hat{\alpha}$, recast each sequence of the estimation process as moments within a GMM framework. The GMM analogue to the sequential process just described involves three sample moments, one for each of the vectors of parameters δ, β and α. The first moment condition is formed from the first derivative of the logit log likelihood function, with respect to β[1]:

$$G_1(\beta, \delta) = \frac{1}{N} \sum_m H_{1m}(\beta, \delta).$$

$$= \frac{1}{N} \sum_m \sum_j y_{mj}(w_{mj} - \sum_i P_{mi} w_{mi}) .$$

The second moment condition, that identifies $\hat{\delta}$, is formed from a vector that constrains the predicted shares of the products in the model to match the aggregate market shares when minimized[2]:

$$G_2(\beta, \delta) = \frac{1}{N} \sum_m H_{2m}(\beta, \delta)$$

$$= \frac{1}{N} \sum_m A - P_m.$$

where $A = [A_1\ A_2 \ldots A_J]'$ and $P_m = [P_{m1}\ P_{m2} \ldots P_{mJ}]'$

The third moment condition estimates $\hat{\alpha}$. It is formed from a vector that when minimized, stipulates that in expectation, the instruments for the attributes of the products are uncorrelated with the error term:

[1] To obtain the BLP, 2004 estimator of the model, replace this moment with one that interacts the average attributes of households with the attributes of vehicles they purchase, then averages over all vehicles in the choice set.
[2] Note that although households are aggregated to "household types" (as indicated by the summations across m, rather than n), expectations are still taken at the household level and not at the household-type level, that is, the first and second moments are averaged over N and not M.

$$G_3(\delta,\alpha) = \frac{1}{J}\sum_j H_{3j}(\delta,\alpha)$$

$$= \frac{1}{J}\sum_j z_j\left(\delta_j - x_j\alpha_1 - p_j\alpha_2\right).$$

One can also obtain point estimates of the BLP model by minimizing the objective function $Q(\theta) = G'W_0G$ where $G = [G_1\ G_2\ G_3]'$, $\theta = [\beta\ \delta\ \alpha]'$, and W_0 is a weight matrix. When the model is just-identified, this solution, $\hat{\theta}_{GMM}$, is equivalent to the solution from sequential ML estimation,$\{\hat{\beta}, \hat{\delta}, \hat{\alpha}\}$ because it satisfies the same estimation conditions.

In line with derivations by Hansen, 1982, the variance of this GMM estimator, θ_{GMM}, is given by the following formula:

$$Var(\theta_{GMM}) = \left(M_0'W_0M_0\right)^{-1}\left(M_0'W_0S_0W_0M_0\right)\left(M_0'W_0M_0\right)^{-1}, \quad (3)$$

where

$$M_0 = \begin{bmatrix} \frac{1}{N}\frac{\partial^2 L}{\partial \beta^2} & \frac{1}{N}\frac{\partial^2 L}{\partial \beta \partial \delta} & 0_{K_1\times K_2} \\ \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \beta} & \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \delta} & 0_{(J-1)\times K_2} \\ 0_{K_2\times K_1} & \frac{1}{\sqrt{NJ}}\sum_j z_j{}' & \frac{1}{J}\sum_j -z_j{}'x_j \end{bmatrix}$$

$$S_0 = \begin{bmatrix} \frac{1}{N}\sum_m H_{1m}H_{1m}{}' & \frac{1}{N}\sum_m H_{1m}H_{2m}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m\sum_j H_{1m}H_{3j}{}' \\ \frac{1}{N}\sum_m H_{2m}H_{1m}{}' & \frac{1}{N}\sum_m H_{2m}H_{2m}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m\sum_j H_{2m}H_{3j}{}' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m\sum_j H_{3j}H_{1nm}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m\sum_j H_{3j}H_{2m}{}' & \frac{1}{J}\sum_j H_{3j}H_{3j}{}' \end{bmatrix}.$$

Since the current model is just-identified, it is efficient to set W_0 as the identity matrix. In over-identified cases ($K_3 > K_2$), the two-step optimal GMM method provides more efficient estimates. Appendix A provides a more detailed derivation and explanation of these standard errors.

Since $\frac{\partial H1}{\partial \alpha}, \frac{\partial H2}{\partial \alpha} = 0$, this GMM model is a sequential two-step estimator (Newey, 1984). Using the derivations by Murphy and Topel (1985), the following equation provides the correction for the downward bias present in the standard errors of IV estimates from the second stage of estimation:

$$Var(\hat{\alpha}) = s^2(X'Z(Z'Z)^{-1}Z'X) + G_{22}{}^{-1}\{G_{21}G_{11}^{-1}S_{11}G_{11}^{-1}G_{21} - G_{21}G_{11}^{-1}S_{12} - S_{21}G_{11}^{-1}G_{21}'\}G_{22}{}^{-1}$$

where $G_{11} = \begin{bmatrix} \frac{\partial H_1}{\partial \beta} & \frac{\partial H_1}{\partial \delta} \\ \frac{\partial H_2}{\partial \beta} & \frac{\partial H_2}{\partial \delta} \end{bmatrix}$, $G_{21} = \begin{bmatrix} \frac{\partial H_3}{\partial \beta} & \frac{\partial H_3}{\partial \delta} \end{bmatrix}$, $G_{22} = \begin{bmatrix} \frac{\partial H_3}{\partial \alpha} \end{bmatrix}$, $S_{11} = \begin{bmatrix} \frac{1}{N}\sum_m H_{1m}H_{1m}' & \frac{1}{N}\sum_m H_{1m}H_{2m}' \\ \frac{1}{N}\sum_m H_{2m}H_{1m}' & \frac{1}{N}\sum_m H_{2m}H_{2m}' \end{bmatrix}$,

The first term is the standard IV formula for standard errors. The second term is an upward correction to account for the first stage. This second term has three components. The first, containing the covariance matrix of the first stage parameters, $G_{11}^{-1}S_{11}G_{11}^{-1}$, accounts for the variance of δ, while the second and third components account for the correlation between the errors across the two stages.

3.1 Speeding up the Contraction Mapping Algorithm

The contraction-mapping algorithm can incur a high time cost particularly when tight stopping criteria are used, and the dimensions of the choice set are large. To speed up the contraction-mapping algorithm, we implement the Li, 2012 modification that augments the contraction mapping with an analytic Newton-Raphson algorithm. The modification produces a 280-fold improvement in the number of iterations necessary for convergence over the unmodified contraction mapping algorithm and a six-fold improvement in estimation time when using simulated data sets. The modification is as follows:

$$\delta_{j,t+1} = \delta_{j,t} + \mathrm{H}^{-1}[\ln(A_j) - \ln(\widehat{S_j})] \ \forall j = 1,2 \ \dots J,$$

where H is the matrix of first order partial derivatives of $[-\ln(\hat{S})]$, which can be shown to equal

$$H = \begin{bmatrix} 1 - \frac{\sum_n P_{n1}^2}{\sum_n P_{n1}} & -\frac{\sum_n P_{n1}P_{n2}}{\sum_n P_{n1}} & \cdots & -\frac{\sum_n P_{n1}P_{nJ}}{\sum_n P_{n1}} \\ -\frac{\sum_n P_{n2}P_{n1}}{\sum_n P_{n2}} & 1 - \frac{\sum_n P_{n2}^2}{\sum_n P_{n2}} & \cdots & -\frac{\sum_n P_{n2}P_{nJ}}{\sum_n P_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\sum_n P_{nJ}P_{n1}}{\sum_n P_{nJ}} & -\frac{\sum_n P_{nJ}P_{n2}}{\sum_n P_{nJ}} & \cdots & 1 - \frac{\sum_n P_{nJ}^2}{\sum_n P_{nJ}} \end{bmatrix}$$

One of the drawbacks of the Newton-Raphson algorithm is that in some situations, the method fails to converge. To avoid this problem from occurring in the empirical example, we only invoke the Newton modification when $\delta_{t+1} - \delta_t \leq 10^{-3}$. Using this ad-hoc rule, the contraction mapping algorithm does converge each time it is called in the empirical example, although the time savings are now only four-fold.

Recent work shows certain approaches to estimating the BLP models on macro-level data behave poorly. Nevo, 2000, Dubé et al., 2012, and Knittel and Metaxoglou, 2014, among others, note that the model is sensitive to starting values, requires very tight convergence criteria, can have multiple local minima, may falsely stop at points that are not even local minima, and is sensitive to the choice of optimization routine used. In light of these findings, in the following chapter of this dissertation, we study these concerns within the BLP model for combined micro- and macro-level data, finding that in these data settings, the ML approach does not suffer from such numerical concerns.

4 A Monte Carlo experiment on the consistency of standard errors

The following Monte Carlo experiment analyzes the properties of the standard errors from sequential ML estimation and the analytic GMM standard errors. The experiment proceeds as follows: Exogenous variables, $\{w_{mj}, x_j, z_j\}$ are generated for a large population of household types, N_{pop}, and a fixed number of products, J, For each generated household type, we create r_m households. For simplicity, an equal number of households are generated for each type, $r_m = r$, $\forall m$. More detailed information on the data generation process is available in Appendix B. Each iteration of the experiment then follows these steps:

1. New draws of the error terms, $\{\epsilon_{mj}, \xi_{1j}, \xi_{2j}\}$ are generated and with that, the endogenous variables, $\{y_{mj}, \delta_j, p_j\}$ are created as functions of the exogenous variables, error terms and the predetermined values of parameters, $\{\alpha, \beta, \gamma\}$.
2. Using the vector of selected products, $\{y_{mj}\}$, population market shares, A_j, are created.
3. A random subsample of household types, M, is selected from N_{pop}. The number of households in a given subsample, $N = \sum_m r_m = M \times r$
4. Estimation proceeds as outlined in the previous section; the model is estimated sequentially using the iterative ML and contraction mapping algorithm process in the first stage, and IV estimation in the second stage.

5. To obtain standard errors from the sequential process, the standard errors of β are calculated as the inverse of the negative Hessian matrix of the log likelihood function, and the standard errors of α follow the standard IV variance formula. To obtain the corrected standard errors, we use the GMM variance formula in equation (3).

Table 1 displays the coverage probabilities of 80% confidence intervals constructed for β and α. These coverage probabilities were obtained from running 1000 Monte Carlo repetitions of the experiment for various sizes of N, where $J = 30, N_{pop} = 100{,}000$, and $r = 100$. The choice of $J = 30$ is sufficient to obtain asymptotic results in that dimension.

Parameter	$N = 2{,}500$		$N = 10{,}000$		$N = 60{,}000$	
	Sequential	Joint	Sequential	Joint	Sequential	Joint
$\widehat{\beta_1}$	0.390	0.907	0.371	0.839	0.382	0.807
$\widehat{\beta_2}$	0.606	0.883	0.672	0.806	0.700	0.805
$\widehat{\alpha_0}$	0.789	0.813	0.791	0.796	0.810	0.810
$\widehat{\alpha_{11}}$	0.747	0.797	0.794	0.806	0.806	0.806
$\widehat{\alpha_{12}}$	0.597	0.858	0.746	0.805	0.781	0.797
$\widehat{\alpha_2}$	0.807	0.809	0.829	0.827	0.802	0.802

Table 1: Coverage probabilities of 80% confidence intervals for β and α, for $N = 2500, 10000$ and 60000, where $J = 30, N_{pop} = 100,000$, and $r = 100$.

For each value of N, we report two sets of coverage probabilities, the "Sequential" column reports coverage probabilities calculated using standard errors from sequential estimation of the model. The "Joint" column reports coverage probabilities calculated using the analytic standard errors derived from equation (3).

As expected, Table 1 shows that the coverage probabilities for β from the sequential model converge to the wrong values. Confidence intervals for $\beta_1$and β_2 that should capture the true value of the parameter 80% of the time across repeated samples, only capture the true value 37-39% and 60-70% of the time respectively. While the sequential standard errors of β are too small in this study, and more severely so for β_1 than β_2, this result is specific to this Monte Carlo study. The sign and severity of the inconsistency is indeterminate across applications. As previously mentioned, the inconsistency of the sequential estimator stems from the fact that standard errors are derived from a likelihood function that does not contain the same information as was used to obtain the point estimates. Since the average utilities are estimated by matching predicted probabilities to aggregate shares, it is incorrect to obtain standard errors as if the predicted probabilities were matched to in-sample shares.

Confidence intervals for $\hat{\beta}$ constructed from the corrected standard errors perform much better. When $N = 2{,}500$, it can be inferred from the coverage probabilities that there is upward bias in the size of confidence intervals (and in turn, the standard errors) in small household samples. Nevertheless, as N increases to $60{,}000$, these coverage probabilities converge to 80%.

As theory predicts, the uncorrected standard errors of $\hat{\alpha}$ from sequential estimation are downward biased. This bias is more pronounced when $N = 2{,}500$, than when $N = 60{,}000$. In fact, when $N = 60{,}000$, the improvements from joint estimation are meager. The downward bias occurs because the standard IV standard errors do not account for the variance present in $\hat{\delta}$. Therefore, the larger the variances of the average utilities, $\hat{\delta}$, the larger the bias in the standard errors from sequential estimation. Since estimates of the average utilities are less precise when the household sample size is small, there are larger downward biases when N is small. As N increases, the variance of the estimates of $\hat{\delta}$ shrink, causing the bias of the uncorrected standard errors of $\hat{\alpha}$ to shrink as well. The standard errors from the joint GMM model, which do account for the variances of $\hat{\delta}$, have the coverage probabilities of approximately 80% across all three values of N considered.

Figure 1 displays box plots of the coverage probabilities of the 80% confidence intervals for the 29 δ_j's from the same Monte Carlo experiment. The left side presents the coverage probabilities using standard errors from the sequential model, while the right side presents the coverage probabilities using the corrected standard errors from the joint model.

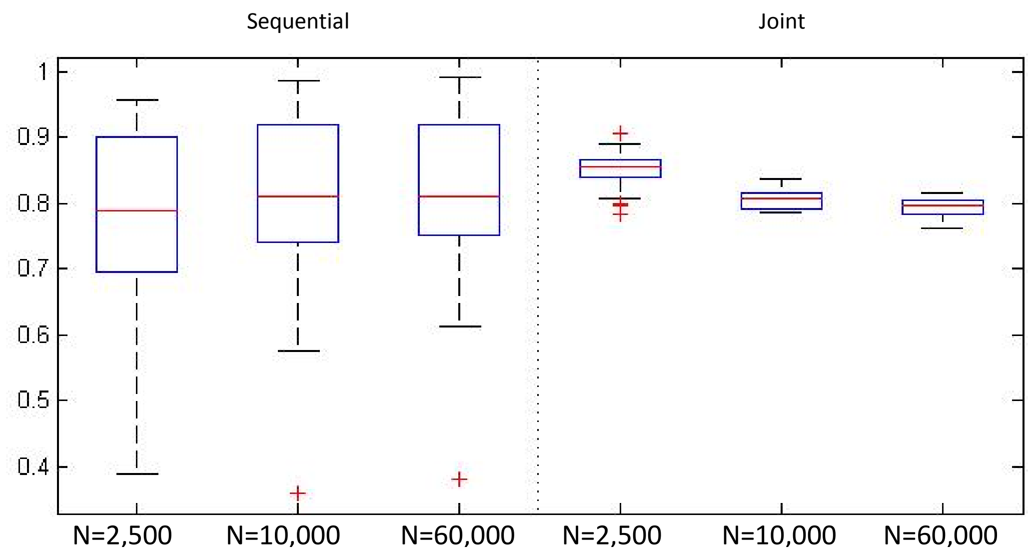

Figure 1: Box plots of the coverage probabilities of 80% confidence for the "average" utilities estimated when $M = 25, 100$ and 600, where $J = 30, N_{pop} = 100,000$, and $r = 100$.

The three box plots on the left side of the figure show there is large variation in the coverage probabilities of the $\delta_j' s$, though this variation narrows as N increases. The inability to obtain coverage probabilities of the $\delta_j' s$ that center around 0.8 reflect the inconsistency of using maximum likelihood derivations to calculate the standard errors of estimates that are not obtained from maximum likelihood estimation.

The three box plots on the right side of the figure suggest that there is an upward bias in the standard errors for many of the δ_j's in small samples. When $N = 2{,}500$, even the first quartile value is greater than 80%, and the distribution shows several outliers. Nevertheless, this bias diminishes quickly, as sample sizes increases. As shown in the box plots for $N = 10{,}000$ and $N = 60{,}000$, the coverage probabilities for the 80% confidence intervals for all of the δ_j's center tightly around 80%.

To increase efficiency of the second stage estimates, Goolsbee and Petrin, 2004, and Chintagunta and Dubé, 2005, weigh the second stage observations by the inverse of the variance of δ but use the sequential approach to calculate the standard errors.[3] Running a Monte Carlo study on this

[3] Goolsbee and Petrin, 2004, do not incorporate macro-level share data in their model, hence their standard errors of δ obtained from the Hessian of the log likelihood function are consistent. Chintagunta and Dubé, 2005, obtain standard errors of δ through a parametric bootstrap.

method, we find that the second stage standard errors behave similarly to the sequential standard errors in Table 1 at best. In many Monte Carlo runs, second stage standard errors were more downward biased when weights were used.

5 Methods of Aggregation

In this section, we introduce two methods that address the concerns related to aggregation of products to broad levels. The first model, introduced in McFadden, 1978, proposes that the covariance matrix of the attributes within each broad group and the number of products within each broad group be included in the likelihood function of the choice model. The second model is a model for broad choice data, introduced in Brownstone and Li, 2014. In their model, equation (2) is defined in terms of the broad choice sets from which household choices are observed and the broad choice probabilities are defined as the sum of the probabilities of the exact choices contained within each broad choice. Before introducing the two models, it is necessary to formally define broad choice data.

Define C as the exact choice set that contains all products, $j = 1, 2, \dots, J$. C is decomposed into B groups, denoted $C_b, b = 1, 2, \dots, B$. Each product, j, belongs to only one choice group such that, $C = \bigcup_{b=1}^{B} C_b$ and $\bigcap_{b=1}^{B} C_j = \emptyset$. Individuals' exact choices, y_{nj}, are not observed. Instead, what is observed are individuals' choices among the broad choice groups, Y_{nb}:

$$Y_{nb} = \begin{cases} 1 \; if \; y_{nj} \in C_b \\ 0 \; otherwise. \end{cases}$$

5.1 McFadden's Method for Aggregation

The empirical application concerned in McFadden, 1978, is the modeling of household choice of residential location. Here, the broad choice groups are communities where households are known to reside while the exact choice set contains the dwellings within these communities.

Let w_{nj} denote the observed attributes of household n interacted with the attributes of dwelling j. When the number of dwellings within a community is large, and w_{nj} behaves as if it is independently identically normally distributed with mean, w_{nj}^{*}, and variance Ω_{nb}, then, McFadden, 1978, shows that for the conditional logit with linear utility specification, the probability that household n chooses community b converges to:

(4)

$$\bar{P}_{nb} = \frac{\exp(\delta_b + w_{nb}^{*}{}'\beta + \frac{1}{2}\beta'\Omega_{nb}\beta + \log(D_b))}{\sum_k \exp(\delta_k + w_{nk}^{*}{}'\beta + \frac{1}{2}\beta'\Omega_{nk}\beta + \log(D_k))}$$

where D_b is the number of dwellings in community b.

The presence of the term, $\frac{1}{2}\beta'\Omega_{nb}\beta$, in (4) comes from the fact that the sample mean and sample sum of squared errors are sufficient statistics for the normal distribution with unknown mean and variance. To account for the distribution of characteristics of products within group b, it is necessary to condition on both quantities. The intuition here is that community attributes with larger variances should have a greater impact on the probability that the community is selected. A simpler approach to incorporate Ω_{nb}, is to relax the constraint that its associated parameter, β, is equal to the parameter on w_{nb}^{*}. This approach yields consistent estimates without the complexity of non-linear constraints.

Lerman, 1977, explains that the $\log(D_b)$ term is a measure of community size. "Other conditions being equal, a large tract (i.e., one with a large number of housing units) would have a higher probability of being selected than a very small one, since the number of disaggregate opportunities is greater in the former than the latter." Here, the coefficient associated with $\log(D_b)$ is assumed to be one, because it is assumed that the logit model applies to each product in the exact choice set. Should this assumption not hold, then the coefficient on $\log(D_b)$ will differ from one.

The simplest way to apply McFadden's method to address aggregation concerns within the BLP setting, is to apply the likelihood function as in equation (2), to household choice of communities, replacing the choice probability defined in equation (1) with equation (4).

One limitation of this method is that it does not controls for the aggregation present in the product specific constants, δ_b. Recall that in the BLP model, the product specific constants are a function of product attributes. Using McFadden's method in this model, it is assumed that $\delta_b$is a linear combination of x_b^* and p_b^*, the means of x_j and p_j. To apply McFadden's method to account for the aggregation in x_b^* and p_b^*, one would have to include the covariance matrix of x_b and p_b in P_{nb}. Since P_{nb} would then be a function of both β and α, the sequential approach to estimating the BLP model would no longer be valid. Because of the estimation complexities this introduces, in this paper, we do not control for aggregation in x_b^* and p_b^*. Additionally, while McFadden's method applies easily to the conditional logit and nested logit, it is not immediately intuitive how the method can be applied to frameworks with more flexible substitution patterns, such as the mixed logit.

Nevertheless, the BLP model with McFadden's method has its advantages. The model is relatively easy to compute, particularly if the non-linear constraints are ignored, and unlike the proceeding model, does not require aggregate market share data at the exact choice level for identification.

5.2 A BLP Model for Broad Choice Data

When the researcher observes individuals' choices among broad choice groups, but macro-level market share data is available at the exact choice level, Brownstone and Li propose estimating household choice at the broad group level, defining the probability of choosing a broad group as the sum of the probabilities of the exact choices contained within the group. This involves replacing the likelihood function in equation (2) with the following:

$$L(y; \delta, \beta) = \sum_{n} \sum_{b} Y_{nb}\, log\ (\tilde{P}_{nb}) \tag{5}$$

where $\tilde{P}_{nb} = \sum_{j \in C_b} P_{nj}$ and P_{nj} is defined as in equation (1).

Estimation of this model follows the maximum likelihood approach detailed in Section 3. The contraction mapping algorithm is used to estimate δ_j, while β is estimated by maximizing equation (5).

Brownstone and Li, 2014, show that equation (5) is not globally concave and generally, has less concavity than equation (2). They also show that the parameters of the model, estimated on exact choice data, are better identified and have smaller variances than estimates from the model for broad choice, given the added uncertainty in the model for broad choice, stemming from having only partial information on household decisions.

There are numerous advantages to estimating the broad choice model over McFadden's method. It avoids aggregation altogether and preserves the sequential estimation process. Also, the model for broad choice does not require that asymptotic distributional assumptions be placed on the variables of the exact choice set, X_{nd}, within each broad group as is the case with McFadden's method. There may not always be an intuitive way to partition the exact choice set into groups that are all large to best approximate the asymptotic normality assumption required for consistency when using McFadden's method. Additionally, the model for broad choice produces estimates of product specific constants for each product in the exact choice set, rather than product specific constants at the broad group level. This allows for more observations in the second stage regression, providing more variation and power to the estimates in that stage. This addresses a concern of these types of models, raised in BLP, 2004. Because these models are typically estimated on single cross-sections of data, there are often not enough observations in the second stage to obtain precise estimates of β.

However, the model for broad choice is poorly identified without macro-level market share data at the exact choice level. Additionally, the model for broad choice can be computationally burdensome to estimate because of the higher dimensions of δ. Finally, given that there are many more first stage parameters to be estimated in this model, larger household sample sizes are required to obtain significant results. Because these first stage parameters are used in the second stage, this can also affect the efficiency of the second stage estimates as well.

6 A Vehicle Choice Application: Explaining the inconclusive evidence on the energy paradox

Standard economic theory dictates that households should invest in an energy saving opportunity if the upfront cost of investment is lower than the present value of future savings from decreased energy bills. However, analysts of the energy industry have long debated the existence of an energy paradox, that households undervalue energy saving investments, hence underinvest in such technologies and over-consume energy. Explanations for this paradox include market failures such as imperfect information, principle-agent issues, credit constraints, and non-rational behavior such as loss aversion and hyperbolic discounting (Gillingham and Palmer, 2014). If the paradox exists, resulting market inefficiencies mean that many households in the state are spending more on fuel than they would if market failures did not exist. Additionally, the paradox suggests that private vehicle travel incurs excess environmental costs, even if proper mechanisms (e.g. carbon taxes) are in place to control for environmental externalities from vehicle emissions. This would mean there is a role for government intervention to correct the market failures through instruments that encourage increased household investments in fuel efficiency technologies such as rebates, taxes, fuel efficiency information programs and mandated fuel efficiency standards.

The energy paradox remains contested, in part because existing estimates of household valuations of energy saving investments are varied and inconclusive. Some of this disagreement comes from studies on vehicle choice (Greene, 2010). In this section, we apply the standard error corrections and methods for aggregation from the previous sections to a vehicle choice application in effort to explain some of this lack of consensus on whether the energy paradox exists.

I model the choices of households in the United States who purchase new model year 2008 vehicles between October 2007 and September 2008. As in Train and Winston, 2007, this model omits households that buy used vehicles or do not make any vehicle purchase over the sample period. Train and Winston, 2007, contend that "preferences among new car buyers can be estimated more accurately by estimating directly on a sample of new car buyers," and has the "practical advantage that it can include explanatory variables whose distributions are not known for the general population."

Data are obtained from Brownstone and Bunch, 2013. They compile data on households and their purchase decisions from the 2009 National Household Transportation Survey (NHTS). The NHTS sample is not a simple random sample. Households in 20 regions were oversampled because metropolitan transportation planning organizations in those regions sponsored larger samples for their own use. In addition, a single interview was conducted for each household between April 2008 and May 2009. Households who were interviewed earlier are more likely to have purchased model year 2008 vehicles after their NHTS interview, and this purchase is not reflected in the sample.

Multinomial choice models still yield consistent estimates of model parameters despite the use of stratified samples as long as all household heterogeneity is fully captured in the model specification through w_{nj} (Manski and Lerman, 1977). However, to obtain population averaged estimates from this model, one must place weights that correct sample stratification.

If household heterogeneity is not fully specified in the model, then estimates of α will not yield the average effect of vehicle attributes on utility. For example, assume that wealthy households are less sensitive to vehicle price than the rest of the population and that this heterogeneity is not accounted for in the model. Then, if wealthy households are over-represented in the sample, the estimates of the average effect of household disutility from vehicle price will be too small.

Alternatively, if the researcher has information about the distribution of characteristics in the population, he could place weights on the observations such that the weighted observations are representative of the population. One can incorporate such weights in estimation through the Weighted Exogenous Sampling Maximum Likelihood Estimator (WESML) by Manski and Lerman, 1977. WESML was developed to address choice-based sampling concerns (where the sample is stratified based on the observed choices, y_{nj}) but also applies when one is interested in estimating "average effects" without a full specification of household heterogeneity.

The results here assume that w_{nj} sufficiently captures household heterogeneity such that estimates from the model are consistent for α. To test the robustness of this assumption, we also estimate the BLP model for broad choice data using sampling weights. These results are presented in Appendix C.

Vehicle attributes are provided by the Volpe Center and supplemented with data from Polk, the American Fleet Magazine, and the National Automobile Dealers Association. Vehicle price data are adjusted adding the gas guzzler tax for some vehicles and subtracting estimated purchase subsidies for hybrid vehicles. Vehicle attribute data are available at the trim level, however, NHTS household choices are only observed at the Make/Model/Fuel-type level.

Macro-level market share information at the Make/Model/Fuel-type/trim[4] level is also obtained from the Volpe Center. They collect information on production volumes which represent all model year 2008 vehicles that are produced (and eventually sold.) These production volumes are adjusted to omit fleet vehicle purchases.

[4] "Make" refers to the manufacturer of the vehicle (e.g. Ford, General Motors, Toyota.) "Model" refers to the product name (e.g. Focus, Chevrolet Impala, Prius.) "Fuel-type" refers to the power source to move the vehicle (gasoline, natural gas, gasoline-electric hybrid). "Trim" denotes different configurations of standard equipment and amenities for a given vehicle make and model, such as manual or automatic transmission, fabric of leather seats, and number of engine cylinders (e.g. Honda Civic DX, Honda Civic LX.)

There are 10,500 NHTS households in the dataset who purchase at least one new model year 2008 vehicle during the sample period. All household characteristic variables are categorical in nature. Because of this, we aggregate to 4157 unique "household types," with between one and forty-one households within each type. There are 235 broad groups of vehicles that households choose from, and 1120 vehicles in the exact choice set.

Table 2 provides some descriptive statistics about the NHTS household sample as well as a comparison of NHTS in-sample vehicle shares to the Volpe aggregate market shares. Table 3 summarizes the utility specification that is used in the models.

NHTS Socioeconomic Attribute Variables	Sample Value (%)	
Percent retired with no children	34.17	
Percent whose children is under the age of 15	26.89	
Percent living in urban areas	68.07	
Percent of household respondents with college degree	48.12	
Average gasoline price at time of vehicle purchase ($)	3.46	
Household Income Distribution†:		
Less than $25,000	5.98	
$25,000 - $75,000	35.36	
$75,000 - $100,000	16.62	
Greater than $100,000	35.28	
Income Missing	6.76	
Household Size Distribution		
1	10.96	
2	49.31	
3	17.14	
4+	22.58	
Market share of MY2008 vehicle purchases by Manufacturer (NHTS household in-sample shares vs. Volpe aggregate market shares)	Share (%) NHTS	Share (%) Volpe
General Motors	21.76	20.76
Toyota	18.82	18.98
Honda	15.53	13.45
Ford	13.87	12.05
Other Japanese	8.87	7.58
Chrysler	8.53	9.79
European	6.46	7.59
Korean	4.30	5.06

Table 2: Descriptive Statistics of the NHTS sample and market shares

†Although five household income categories are observed, we use only four in the empirical application. We combine the lowest two categories into one for purposes of identification as we find the results for the two categories are very similar.

x_j	w_{nj}
Price	(Price) × (75,000<Income<100,000)
Horsepower/Curb Weight	(Price) × (Income>100,000)
Hybrid	(Price) × (Income Missing)
Curb Weight	(Prestige) × (Urban)
Wagon	(Prestige) × (Income>100,000)
Mid-Large Car	(Performance Car) × (Income>100,000)
Performance Car	(Japan) × (Urban)
Small-Medium Pickup	(Van) × (Children under 15)
Large Pickup	(Large SUV) × (Children under 15)
Small-Medium SUV	(Small SUV) × (Children under 15)
Large SUV	(Korea) × (Rural)
	(Seats≥5) × (Household Size≥4)
	(Mid-Large Car) × (Retired)
	(Prestige) × (Retired)
	(Import) × (College)
	(Prestige) × (Japan) × (College)
	(Prestige) × (Europe) × (College)
	(Prestige) × (Japan) × (Urban)
	(Performance Car) × (College)
	Fuel Operating Cost (cents per mile)
	(Fuel Operating Cost) × (College)

Table 3: Vehicle attributes, x_j, and vehicle-household attribute interactions, w_{nj}, included in the estimated model
Note**:** Fuel operating cost is the product of gallons per mile and fuel price (in cents per mile)
"Korea," "Japan," and "Europe" are dummy variables that equal 1 if the vehicle is made in that region and 0 otherwise.
"Prestige" is a dummy variable that equals 1 if the vehicle is classified as a "prestige brand" by the American Fleet Magazine.
The following vehicle classes were adopted from the American Fleet Magazine: Mid-Large Car, Performance Car, Small-Medium Pickup, Large Pickup, Small-Medium SUV and Large SUV.

I consider three specifications of the BLP model for micro- and macro-level data: a model that aggregates to the Make-Model-Fuel type level (BLP with aggregated choices), the BLP with McFadden's method for aggregation, where data is also grouped at the Make-Model-Fuel type level but the number of choices and co-variance matrix of attributes within each group are utilized in estimation, and the Brownstone and Li, 2014, BLP model for broad choice data.

First, we present the results of the standard error corrections for the BLP model with aggregated choices in Table 4. We focus this discussion on the price and fuel operating cost estimates. Results for the full model are listed in Appendix D. Table 4 shows that the uncorrected standard errors from sequential estimation are smaller than the corrected standard errors. These findings are consistent with the findings from the Monte Carlo study in Table 1. Hence, a likely explanation for this is that the former are biased downward. Most corrected standard errors are larger by a factor of about 2 to 3. The corrected standard error for fuel operating cost, however, is 18 times larger. This is a concern given the importance of this variables in constructing estimates of how much households value fuel efficiency improvements.

Variable	BLP with Aggregated Choices			
	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error	Ratio of Corrected to Uncorrected Standard Errors
(Price) × (75,000<Income<100,000)	0.065	0.004 ***	0.014 ***	3.067
(Price) × (Income>100,000)	0.102	0.004 ***	0.015 ***	3.556
(Price) × (Income Missing)	0.094	0.005 ***	0.015 ***	3.140
Fuel Operating Cost (cents per mile)	-2.877	0.053 ***	0.953 ***	18.064
(Fuel Operating Cost) × (College)	-0.061	0.009 ***	0.020 ***	2.225
Price	-0.116	0.019 ***	0.026 ***	1.368

Table 4: BLP with Aggregated Choices: The sequential and GMM standard errors for select parameters.
*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table 5 presents select point estimates and standard errors, both corrected and uncorrected, for the BLP model for broad choice data. Full results are in Appendix D. The ratio of corrected to uncorrected standard errors display similar behavior as in Table 4. Again, the greatest correction occurs for the fuel operating cost variable, and the ratios are larger across the board compared to Table 4.

Variable	BLP Model for Broad Choice Data			
	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error	Ratio of Corrected to Uncorrected Standard Errors
(Price) × (75,000<Income<100,000)	0.038	0.006 ***	0.052	9.379
(Price) × (Income>100,000)	0.123	0.008 ***	0.100	13.343
(Price) × (Income Missing)	0.079	0.006 ***	0.056	9.383
Fuel Operating Cost (cents per mile)	-0.599	0.048 ***	2.044	42.908
(Fuel Operating Cost) × (College)	-0.057	0.013 ***	0.076	5.792
Price	-0.098	0.008 ***	0.097	12.686

Table 5: BLP with Broad Choice Data: The sequential and GMM standard errors for select parameters.
*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table 6 presents the point estimates and corrected standard errors for the price, fuel cost, horsepower, and curb weight variables from the BLP model with Aggregated Choices, BLP model with McFadden's method for aggregation, and the BLP for broad choice data.[5] These variables

[5] For results from the full model specifications, see Appendix D.

vary by trim and hence are aggregated in both the BLP model with aggregated choices and the BLP Model with McFadden's method. Recall that the BLP model with McFadden's method controls for aggregation in the {Price × Income} and {Fuel Operating Cost} variables but not for the aggregation in variables held constant across households (i.e. price, horsepower/curb weight, curb weight.) In the BLP model for broad choice, there is no aggregation of vehicle attributes.

There are interesting comparisons to be made across the three models regarding the significance of estimates. Table 6 shows that more variables appear statistically significant with aggregation than without. When using McFadden's method, only the coefficients on curb weight and fuel operating cost are significant at the 1% level. By controlling for aggregation in the first stage, all coefficients on price variables and the coefficient on Fuel Operating Cost × College lose significance. In the BLP model for broad choice data, all variables reported in Table 6 lose significance, as is the case for all but two variables in the entire specification (see Table D3).

There are two explanations for why variables are less significant when aggregation is either accounted for or avoided altogether. First, by treating the aggregated choice set as if it is the exact choice set, the BLP model with aggregated choices ignores the uncertainty from the fact that choices are only partially observed, and aggregated vehicle attributes contain measurement error. Hence, the model produces smaller standard errors than the two models that do account for aggregation. These findings are consistent with comparisons between the "peakedness" of the likelihood function of the exact choice model and the broad choice model, presented in Brownstone and Li, 2014. Second, with respect to the BLP model for broad choice data, because of the high dimensionality of δ, estimates of these parameters are less precise. The large standard errors associated with these estimates inflate the standard errors of the second stage coefficients making them less significant. Though the BLP model for broad data allows for more observations in the second stage, unfortunately, the gains in precision from these added observations are swamped by the lack of precision from the increase in the dimension of $\hat{\delta}$.

Variable	BLP with Aggregated Choices			BLP with McFadden's Method			BLP for Broad Choice Data	
	Estimated Parameter	Corrected Standard Error		Estimated Parameter	Corrected Standard Error		Estimated Parameter	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.065	0.014	***	0.001	0.067		0.038	0.052
(Price) × (Income>100,000)	0.102	0.015	***	0.004	0.056		0.123	0.100
(Price) × (Income Missing)	0.094	0.015	***	0.011	0.080		0.079	0.056
Fuel Operating Cost (cents/mile)	-2.877	0.953	***	-2.946	0.263	***	-0.599	2.044
(Fuel Operating Cost) × (College)	-0.061	0.020	***	-0.027	0.466		-0.057	0.076
Price	-0.116	0.026	***	-0.064	0.120		-0.098	0.097
Horsepower / Curb weight	158.582	53.803	***	144.232	93.200		20.737	111.690
Curb Weight	7.569	2.511	***	7.084	1.907	***	0.002	0.006

Table 6: Select estimates across BLP models with various methods for aggregation.
*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Given that there are more alternatives in the BLP model for broad choice data and more variables in the BLP model with McFadden's method, we cannot speak of the differences in magnitudes of coefficients across the three models. However, comparisons can be made across utility independent quantities such as the estimates of willingness to pay for improvements in fuel operating costs.

Willingness to pay for a 1 cent/mile improvement in fuel efficiency (thousands)†	Estimated Parameter	Uncorrected Standard Error		Corrected Standard Error‡		Ratio of Corrected to Uncorrected Std. Errors	Implied Discount Rate
BLP Model with Aggregated Choices	24.695	4.090	***	10.128	**	2.477	-23.675
BLP Model with McFadden's Method	46.083	14.663	***	83.105		5.667	-28.132
BLP Model for Broad Choice Data	6.123	0.683	***	22.706		33.234	-10.785

Table 7: Willingness to pay estimates across the three model specifications
*Note: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*
† willingness to pay for a 1 cent/mile reduction in fuel operating costs for households with no college education and income below $75,000 (in thousands of dollars).
‡ calculated using the delta method:

$$Var(williness\ to\ pay) = Var\left(\frac{\beta_{fuelop}}{\alpha_{price}}\right) = \frac{\beta_{fuelop}^2}{\alpha_{price}^4}\sigma_{price}^2 + \frac{1}{\alpha_{price}^2}\sigma_{fuelop}^2 - \frac{2\beta_{fuelop}}{\alpha_{price}^3}\rho_{fuelop,price}\sigma_{price}\sigma_{fuelop},$$

$$\sigma_{price}^2 = var(\alpha_{price}), \sigma_{fuelop}^2 = var\left(\beta_{fuelop}\right), \rho_{fuelop,price} = corr(\beta_{fuelop}, \alpha_{price})$$

Table 7 presents the implied willingness to pay estimates for a 1 cent/mile improvement in fuel operating cost, in thousands of dollars, for households with no college education and incomes less than $75,000. These quantities are constructed using the estimates in Table 6. This 1 cent/mile improvement in fuel cost is a 7.4% improvement over the average fuel operating cost of households in the sample. The final column of Table 7 provides the implied discount rate assuming vehicles are held for 14 years, with an annual mileage of 18,778, assumptions used in Greene, 2010. The negative discount rates in this column indicate that across all three models, households overvalue future fuel savings compared to present day investments in vehicle fuel efficiency.

The BLP model with McFadden's method has the largest willingness to pay estimates while the BLP model with broad choice data has the smallest estimates, though only the BLP model with aggregated choices presents an estimate that is significant at any of the conventional levels. An interesting point to draw from Table 7 is that the use of uncorrected standard errors leads one to believe that all three estimates of willingness to pay are significant, whereas with corrected standard errors only one of them (BLP with aggregated choices) is, and only at the 5% level. For the BLP model for broad choice data, the difference is driven largely by the fact that using the uncorrected standard errors grossly understates the uncertainty in the fuel operating cost coefficients. For the BLP model with McFadden's method, the difference is because the uncorrected standard errors understate the uncertainty in the price coefficients. The impreciseness of these estimates suggests that when aggregation is accounted for, this dataset does not speak loudly enough to provide conclusive evidence on the existence of the energy paradox within this model specification.

One will also notice from Table 6 that the estimates associated with horsepower and curb weight vary more drastically across models than estimates associated with other variables. This means that many utility invariant quantities formed from these variables such as willingness to pay for horsepower and willingness to pay for curb weight will differ across models as is the case with vehicle fuel efficiency. This suggests that choice set aggregation also affects the estimates associated with attributes other than fuel efficiency. In addition, fuel efficiency, horsepower, and curb weight are highly correlated variables. It is likely that the cost of aggregation increases in the face of collinearity because aggregation further obfuscates the ability to disentangle these competing effects. This is reflected in the instability of the estimates of these three variables across the three models.

The Make/Model/Fuel-type groups contain between one and fifty-five trims, with 47.7% of the groups containing 3 trims or less. Thus the assumption that attributes within each broad group follow a normal distribution may not hold for all Make-Model groups. This means that the estimates from the BLP model with McFadden's method for aggregation may be less plausible in this empirical application. The BLP model with McFadden's method may perform better if choices are aggregated to a higher level (for example, the Make/Class level) such that there are many more vehicle trims within each broad group. The distributional assumptions may be better approximated in such a setting.

There are a number of possible reasons why willingness to pay for fuel efficiency is not estimated with sufficient precision in the BLP Model for Broad Choice Data to provide conclusions about the existence of an energy paradox. Given the larger choice set in this model, an even larger number of households is required for asymptotic assumptions to apply. A larger household sample may also be required to obtain significant estimates. It is also possible that there is a lack of sufficient variability in vehicle fuel efficiency within this dataset to pin down household valuation of this attribute. In addition, vehicle fuel efficiency is highly correlated with horsepower and curb weight, making it difficult to disentangle the value households place on each of these attributes. Finally, it is possible that the model specification is simply too rich to be identified by the current data because of the uncertainty from only broadly observing choices. If pared down to a smaller model, it may perform better.

7. Conclusion

In this paper, we examine the implications of choice set aggregation on parameter estimates in multinomial choice models, a practice common when household choices are not fully observed and when modeling choices at the most detailed level renders the model too large for estimation in standard computing environments. We discuss two models that account for choice set aggregation. The first is a method for aggregation by McFadden, 1978, that places distributional assumptions on the elements within each aggregated alternative and uses the higher order moments of the distribution in the utility specification. The second is a model for broad choice by Brownstone and Li, 2014, that defines the choice probability of a broad group of products as the sum of the probabilities of products within that group, circumventing the need for aggregation.

Applying these models, and a model that aggregates choices, to vehicle choice data reveals that aggregation affects the point estimates of the model. In addition, standard errors are smaller when the presence of aggregation is ignored, because the measurement error from aggregation does not get accounted for in estimation.

Choice set aggregation is studied within a model that employs the innovations from a series of Berry, Levinsohn, and Pakes papers. One can incorporate these innovations into choice models by estimating the model sequentially. However, standard errors from this estimation procedure are inconsistent. We derive consistent standard errors for the model by recasting the model within a GMM framework, and find, both in a Monte Carlo study and a vehicle choice application, that for all parameters of interest, the GMM standard errors are larger than the inconsistent standard errors derived from sequential estimation. For most variables, the GMM standard errors are larger by a factor of 1.3 to 5, though the largest correction inflates the standard errors of a variable by a factor of 18.

If these findings on choice set aggregation and inconsistent sequential standard errors extend to existing vehicle choice studies that aggregate the choice set, than this may help explain some of the variation in existing estimates of households' valuations of fuel efficiency from choice models which have been a source of conflict on the presence of an energy paradox. First, the variation in estimates across studies may be driven in part by the practice of choice set aggregation. Second, the use of inconsistent standard errors from the sequential estimation process may lead researchers to be overconfident in the significance of model estimates. Finally, models that aggregate the choice set ignore the measurement error that aggregation induces also causing standard errors of their estimates to be too small. Addressing these concerns may help increase coherency across studies on how much value households place on fuel efficiency.

The findings from this paper should also serve as a call to search for better data and identification strategies in vehicle choice models so that future research can provide more precise estimates on the matter. For example, there is greater variation in vehicle fuel efficiency in the present decade than the last because of higher fuel prices, more hybrid vehicle models and tighter fuel efficiency standards. Also, data on used-vehicle purchases will introduce attractive variation in vehicle attributes since the mix of vehicle attributes has changed quite rapidly in the last decade.

Exploiting this variation may provide tighter estimates of household valuation for fuel efficiency and a clearer resolution to the energy paradox.

More generally, these findings send a cautionary message to choice model practitioners on the importance of giving due consideration to how choice sets are defined. Aggregating choices without accounting for the variation across aggregated alternatives may lead the researcher to flawed and invalid conclusions from model estimates. Given the popularity of multinomial choice models across a variety of fields, including transportation, industrial organization and marketing, such practices may have widespread consequences. This paper brings attention to two existing models that provide methods to address aggregation in hope that practitioners will adopt these methods in future work.

References

Allcott, H., and Greenstone, M., 2012. Is There an Energy Efficiency Gap? Journal of Economic Perspectives. 26(1): 3-28.

Allcot, H., and Wozny, N., 2014. Gasoline Prices, Fuel Economy, and the Energy Paradox. The Review of Economics and Statistics. 96(5): 779-795.

Bento, A. M., Goulder, L. H., Jacobsen, M. R., von Haefen, R. H., 2009. Distributional and Efficiency Impacts of Increased US Gasoline Taxes. American Economic Review. 99(3):667 – 699.

Berry, S. T., 1994. Estimating Discrete-Choice Models of Product Differentiation. The RAND Journal of Economics. 25(2), 242-262.

Berry, S.T., Levinsohn, J., Pakes, A., 1995. Automobile Prices in Market Equilibrium. Econometrica. 63(4), 841-890.

Berry, S.T., Levinsohn, J., Pakes, A., 2004. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. Journal of Political Economy. 112(1), 68-105.

Bhat, C., Sen, S., 2006. Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-continuous Extreme Value (MDCEV) Model. Transportation Research Part B: Methodological 40:35-53.

Brownstone, D., Bunch, D. 2013. A Household New Vehicle Purchase Model to Support Analysis of the Impact of CAFÉ Standards. VOLPE Report

Brownstone, D., Li, P., 2014. A Model for Broad Choice Data. Department of Economics Working Paper, UC Irvine.

Brownstone, D., Wong, T., 2014. Adapting Discrete Choice Models for Massively Parallel Computing Environments. Department of Economics Working Paper, UC Irvine.

Busse, M.R., Knittel, C.R., Zettelmeyer, F., 2013. Are Consumers Myopic? Evidence from New and Used Car Purchases. American Economic Review. 103(1): 220 – 256.

Byrd R.H., Nocedal, J., Waltz, R. A., 1999. KNITRO: An Integrated Package for Nonlinear Optimization. In G di Pillo and M. Roma (Eds.), Large-Scale Nonlinear Optimization. Springer. pp35-59.

Byrd, R. H., Hribar, M. E., Nocedal, J., 1999. An Interior Point Method for Large Scale Nonlinear Programming. SIAM Journal on Optimization. 9(4), 877-990

Chantagunta, P., Dubé, J., 2005. Storekeeping Unit Brand Choice Model that combines Household Panel Data and Store Data. Journal of Marketing Research 42(3), 368-379.

Dubé, J., Fox, J.T., Su, C., 2011. Improving the Numerical Performance of BLP Static Dynamic Discrete Choice Random Coefficients Demand Estimation. Econometrica 80(5), 2231-2267.

Feather, P. M., 1994. Sampling and Aggregation Issues in Random Utility Model Estimation. American Journal of Agricultural Economics 76(4), 772-780.

Gillingham, K., Palmer, K., 2014. Bridging the Energy Efficiency Gap: Policy Insight from Economic Theory and Empirical Evidence. Review of Environmental Economics and Policy. 8(1): 18-38.

Goldberg, P.K., 1998. The Effects of the Corporate Average Fuel Efficiency Standards in the U.S.. The Journal of Industrial Economics, XLVI(1):1-33.

Goolsbee, A., Petrin, A., 2004. The Household Gains from Direct Broadcast Satellites and the Competition with Cable TV. Econometrica. 72(2), 351-381.

Greene, D., 2010. How Households Value Fuel Economy: A Literature Review. Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Report EPA-420-R-10-008.

Grigolon, L., Reymaert, M., Verboven, F., 2014. Consumer Valuation of Fuel Costs and the Effectiveness of Tax Policy: Evidence from the European Car Market. Working Paper.

Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica. 50(4): 1029-1054.

Haener, M. K., Boxal, P.C., Adaamowicz, W. L., Kuhnke, D. H., Kuhnke., 2004. Aggregation Bias in Recreation Site Choice Models: Resolving the Resolution Problem. Land Economics 80(4): 561-574.

Helfand, G., Wolverton, A., 2011. Evaluating the Household Response to Fuel Economy: A Review of the Literature. International Review of Environmental and Resource Economics 5, 103-146.

Jacobsen, M., 2013. Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity. American Economic Journal: Economic Policy. 5(2):148-187.

Knittel, C., Metaxoglu, K., 2014. Estimation of Random Coefficient Demand Models: Two Empiricists' Perspective. Review of Economics and Statistics. 96(1), 34-59.

Lerman, S. R., 1977. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. Transportation Research Board Record. 610:6-11.

Li, P. 2012. Speeding up the Contraction Mapping Algorithm. In Essays on missing data models and MCMC estimation. Ph.D. Dissertation. UC Irvine Department of Economics.

Manski, C. F., Lerman, S. R., 1977. The Estimation of Choice Probabilities from Choice Based Samples. Econometrica. 45(8): 1977-1988.

McFadden, D., 1978. Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J.Weibull (Eds.), Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam: 75-96.

Murphy, K. M., Topel, R. H., 2002. Estimation and Inference in Two-Step Econometrics Models. Journal of Business and Statistics. 20:1 88-97.

Nevo, A., 2000. Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. Rand Journal of Economics. 31, 395-421.

Nevo, A., 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. Econometrica. 69(2), 307-342.

Newey, W. K., 1984. A Methods of Moments Interpretation of Sequential Estimators. Economics Letters. 14, 201–206.

Parsons, G.R., Needelman, M., 1992. Site Aggregation in a Random Utility Model of Recreation. Land Economics, 68(4), 418-433.

Petrin, A., 2002. Quantifying the Benefits of New Products: The Case of the Minivan. Journal of Political Economy. 110, 705-729.

Spiller, E., 2012. Household Vehicle Bundle Choice and Gasoline Demand: A Discrete-Continuous Approach. Working Paper.

Train, K. E., Winston, C., 2007. Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers. International Economic Review. 48(4), 1469-1496.

Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press: New York.

Appendix

A. Derivation of the corrected variance-covariance matrix of the sequential estimator

The variance of any GMM estimator is given by the following formula:

$$Var(\theta_{GMM}) = (M_0' W_0 M_0)^{-1}(M_0' W_0 S_0 W_0 M_0)(M_0' W_0 M_0)^{-1},$$

where

$$M_0 = \begin{bmatrix} \frac{1}{N}\frac{\partial^2 L}{\partial \beta^2} & \frac{1}{N}\frac{\partial^2 L}{\partial \beta \partial \delta} & 0_{K_1 \times K_2} \\ \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \beta} & \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \delta} & 0_{(J-1)\times K_2} \\ 0_{K_2 \times K_1} & \frac{1}{\sqrt{NJ}}\sum_j z_j' & \frac{1}{J}\sum_j -z_j' x_j \end{bmatrix}$$

where

$$\frac{\partial^2 L}{\partial \beta^2} = -\sum_m \sum_j p_{mj}(w_{mj} - \overline{w}_m)(w_{mj} - \overline{w}_m)'$$

$$\frac{\partial^2 L}{\partial \beta \partial \delta} = -\sum_n \sum_j p_{nj}(w_{nj} - \overline{w}_n)\left(d_j^i - p_{ni}\right)' \forall i = 1, \dots, J-1;$$

$$\frac{\partial P_{mj}}{\partial \beta} = P_{mj}(w_{mj} - \overline{w}_m)$$

$$\frac{\partial P_{mj}}{\partial \delta_i} = \begin{cases} P_{mj}(1 - P_{mj}) & if\ j = i \\ P_{mj} P_{m-j} & otherwise \end{cases}$$

where $\overline{w}_m = \sum_l p_{ml} w_{ml}$,

$$S_0 = \begin{bmatrix} \frac{1}{N}\sum_m H_{1m}H_{1m}' & \frac{1}{N}\sum_m H_{1m}H_{2m}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{1m}H_{3j}' \\ \frac{1}{N}\sum_m H_{2m}H_{1m}' & \frac{1}{N}\sum_m H_{2m}H_{2m}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{2m}H_{3j}' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{3j}H_{1nm}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{3j}H_{2m}' & \frac{1}{J}\sum_j H_{3j}H_{3j}' \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{N}\{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}\ \{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}' & \frac{1}{N}\left(\sum_m M_j - P_{mj}\right)\{\sum_j y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_{\boldsymbol{m}}\sum_{\boldsymbol{j}}\{\boldsymbol{z_j}(\boldsymbol{\delta} - \boldsymbol{x_j\alpha})\}\{\sum_{\mathbf{m}} \mathbf{y_{mj}}(\mathbf{w_{mj}} - \sum_{\mathbf{i}} \mathbf{P_{mi}w_{mi}})\}' \\ \frac{1}{N}\{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}\left(\sum_m M_j - P_{mj}\right)' & \frac{1}{N}\left(\sum_m M_j - P_{mj}\right)\left(\sum_m M_j - P_{mj}\right)' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_{\boldsymbol{m}}\sum_{\boldsymbol{j}}\{\boldsymbol{z_j}(\boldsymbol{\delta} - \boldsymbol{x_j\alpha})\}\left(\sum_{\boldsymbol{m}} \boldsymbol{M_j} - \boldsymbol{P_{mj}}\right)' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_{\boldsymbol{m}}\sum_{\boldsymbol{j}}\{\sum_{\mathbf{m}} \mathbf{y_{mj}}(\mathbf{w_{mj}} - \sum_{\mathbf{i}} \mathbf{P_{mi}w_{mi}})\}\{\boldsymbol{z_j}(\boldsymbol{\delta} - \boldsymbol{x_j\alpha})\}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_{\boldsymbol{m}}\sum_{\boldsymbol{j}}\left(\sum_{\boldsymbol{m}} \boldsymbol{M_j} - \boldsymbol{P_{mj}}\right)\{\boldsymbol{z_j}(\boldsymbol{\delta} - \boldsymbol{x_j\alpha})\}' & \frac{1}{J - K_2}\sum_j z_j(\delta - x_j\alpha)(\delta - x_j\alpha)'z_j' \end{bmatrix}$$

If the model is just identified (as is assumed in this application), then it is sufficient to set $W_0 = I$. In an over identified setting, the two-step optimal GMM procedure can be used to generate the appropriate estimate of W_0.

The expectations for the first two moments, $G_1$and G_2, are taken across households while the expectation for the third moment, G_3, is taken across alternatives. This means that the estimates, β and δ, obtained from minimization of $G_1$and G_2, are consistent and asymptotically efficient as $N \to \infty$ while the estimate of α obtained from G_3 is consistent and asymptotically efficient as $J \to \infty$. Berry, Linton and Pakes, 2002, note that for asymptotic normality to hold, N must increase at rate J^2.

Because of this, there is a peculiarity to the estimation of S_0 that is worth noting. $S_0$is the sum of the outer product of the contributions to the moment function. While the cross product of H_{1m} and H_{2m} is simply $H_{1m}H_{2m}'$, calculating the other cross products is not straightforward as there is a dimensionality mismatch between H_{1m} and H_{3j}, and, H_{2m} and H_{3j} because there are N observations in H_{1m} and H_{2m} and only J observations in H_{3j}.

The solution to this is to divide each of the N moment contributions in H_{1m} and H_{2m} by $\sqrt{N}$ and to divide each of the J moment contributions in H_{3m} by $\sqrt{J}$. Then, each of the N moment contributions in $H_{1m}$and H_{2m} is multiplied with each of the J moment contributions in H_{3j}, and these quantities are summed. This produces the formulas in entries (3,1), (3,2), (1,3) and (2,3) of S_0. Dividing by $\sqrt{N}$ and $\sqrt{J}$ ensures that these quantities converge to finite values as J and $N \to \infty$.

B. Data Generating Process for the Monte Carlo Study

$$y_{mj} = \sum_{l_m=1}^{r_m} y_{nj} \; \forall n \in m$$

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\, i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + w_{mj1}\beta_1 + w_{mj2}\beta_2 + \epsilon_{nj}, \quad where\ n \in m$$

$$\delta_j = \alpha_0 + x_{j1}\alpha_{11} + x_{j2}\alpha_{12} + p_j\alpha_2 + \xi_{1j},$$

$$p_j = z_j\gamma + \xi_{2j}$$

$$(\xi_{1j}, \xi_{2j}) \sim N(0_2, \Omega), \quad \Omega = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

where

x_{j1} is a $J \times 1$ vector of binary values, drawn from a Bernoulli distribution with success probability of 0.5,

x_{j2} is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2,

v_{m1} and v_{m2} are $M \times 1$ vectors of continuous values drawn from a standard normal distribution bounded between -2 and 2,

w_{mj1} is an $MJ \times 1$vector of interactions between v_{m1} and x_{j1} ,

w_{mj2} is an $MJ \times 1$vector of interactions between v_{m2} and x_{j2} , and

z_j is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2.

Also, $\beta_1 = 0.8$, $\beta_2 = -0.7$, $\alpha_0 = 0$, $\alpha_{11} = -0.5$, $\alpha_{12} = 1$, $\alpha_2 = 0.5$, and $\gamma = 1.3$.

C. Results from the BLP model with sampling weights

The household sample obtained from the 2009 National Household Transportation Survey is not a representative sample. Households in 20 regions were oversampled because metropolitan transportation planning organizations in those regions sponsored larger samples for their own use. In addition, a single interview was conducted for each household between April 2008 and May 2009. Households who were interviewed earlier are more likely to have purchased model year 2008 vehicles after their NHTS interview, and these purchases are not reflected in the sample.

Estimates from the choice model will not be consistent under this sampling scheme if household heterogeneity along the stratified dimensions are not fully accounted for in the model specification. The main results of this paper assume that heterogeneity is fully specified. To test this assumption, we estimate the BLP model for broad choice data incorporating sampling weights in estimation that allow the data to approximate a simple random sample.

Unfortunately, in estimation, a few variables no longer converge once weights are used. These variables are:

College X Prestige X Japan
College X Prestige X Europe
Fuel Operating Cost
College X Fuel Operating Cost

Because the "fuel operating cost" variables no longer converge, to capture the effect of fuel efficiency on utility, "gallons per miles" is included in the model. Since the estimates from the weighted and unweighted specifications are no longer directly comparable, in Table C1, we present the implied discount rates from both models for the average household:

	Willingness to Pay (thousands)*	Implied Discount Rate†
BLP Broad Choice without Weights (I)	-0.263	Undefined
BLP Broad Choice without Weights (II)	8.213	-13.809
BLP Broad Choice with Weights	0.526	-19.710

Table C1: Implied Discount Rates from household willingness to pay for fuel efficiency

Notes: *For the BLP Broad Choice without Weights, willingness to pay is for a 1 cent/mile improvement in fuel operating cost. For BLP Broad Choice with Weights, willingness to pay is for a gallon per mile improvement in fuel efficiency.

BLP Broad Choice without Weights (I) takes the weighted average of willingness to pay across income and college groups. BLP Broad Choice without Weights (II) takes the weighted average of willingness to pay across income and college groups assuming that the willingness to pay for households with income greater than $100,000 is zero.
† Implied interest rates are calculated assuming households drive vehicles 18,778 miles a year and hold vehicles for 14 years.

BLP Broad Choice without weights (I) yields an undefined interest rate because the average willingness to pay for a one cent/mile reduction in fuel operating cost across the sample is negative. This is driven by the fact that the coefficient on vehicle price, used to construct the willingness to pay measure, is positive for the households in the highest income category. BLP Broad Choice without weights (II) forces the negative willingness to pay values of high income households to equal zero. This results in an average willingness to pay of $8,213.95 and an implied discount rate of -14.8%.

The BLP Broad Choice with weights estimates that households are willing to pay $525.58 for a one gallon per mile improvement in fuel efficiency. This implies a discount rate of -19.71%. These findings provide some evidence that the model is robust to the use of sampling weights. All three methods in the table generate implied interest rates that suggest an overvaluation of fuel efficiency. When a zero-bound constraint is imposed in aggregating the willingness to pay estimates from the model without weights, it yields an implied interest rate that is qualitatively similar to the implied interest from the model estimated with weights (-14% vs -19%).

D. Full results from the vehicle choice application with aggregated choices

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.065	0.004 ***	0.0136 ***
(Price) × (Income>100,000)	0.102	0.004 ***	0.0153 ***
(Price) × (Income Missing)	0.094	0.005 ***	0.0153 ***
(Prestige) × (Urban)	0.609	0.095 ***	0.2165 ***
(Prestige) × (Income>100,000)	0.167	0.091 *	0.1652
(Performance Car) × (Income>100,000)	0.242	0.103 **	0.1427 *
(Japan) × (Urban)	0.367	0.048 ***	0.1040 ***
(Van) × (Children under 15)	0.849	0.084 ***	0.1238 ***
(Large SUV) × (Children under 15)	0.157	0.056 ***	0.0776 **
(Small SUV) × (Children under 15)	0.503	0.112 ***	0.1683 ***
(Korea) × (Rural)	-0.614	0.105 ***	0.1408 ***
(Seats≥5) × (Household Size≥4)	0.406	0.032 ***	0.1026 ***
(Mid-Large Car) × (Retired)	0.953	0.050 ***	0.0977 ***
(Prestige) × (Retired)	0.575	0.073 ***	0.1396 ***
(Import) × (College)	0.398	0.047 ***	0.0694 ***
(Prestige) × (Japan) × (College)	-0.127	0.177	0.2941
(Prestige) × (Europe) × (College)	-0.598	0.139 ***	0.2156 ***
(Prestige) × (Japan) × (Urban)	-0.180	0.154	0.3329
(Performance Car) × (College)	0.622	0.099 ***	0.1671 ***
Fuel Operating Cost (cents per mile)	-2.877	0.053 ***	0.9533 ***
(Fuel Operating Cost) × (College)	-0.061	0.009 ***	0.0202 ***
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.116	0.019 ***	0.026 ***
Horsepower/Curb weight	158.582	24.622 ***	53.803 ***
Hybrid	-13.222	0.815 ***	4.102 ***
Curb weight	7.569	0.309 ***	2.511 ***
Wagon	0.892	1.248	1.384
Mid-Large Car	-0.685	0.540	0.548
Performance Car	-0.354	0.850	0.889
Small-Medium Pickup	6.196	1.087 ***	2.101 ***
Large Pickup	5.231	1.155 ***	1.454 ***
Small-Mid SUV	9.292	1.668 ***	3.198 ***
Large SUV	2.275	0.472 ***	0.866 ***

Table D1: BLP model with aggregated choices: parameter estimates

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

D.1. Full results from the vehicle choice application with McFadden's method of aggregation

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.001	0.002	0.067
(Price) × (Income>100,000)	0.004	0.001 ***	0.056
(Price) × (Income Missing)	0.011	0.004 ***	0.080
(Prestige) × (Urban)	0.620	0.094 ***	2.735
(Prestige) × (Income>100,000)	1.195	0.072 ***	2.257
(Performance Car) × (Income>100,000)	0.712	0.112 ***	1.121
(Japan) × (Urban)	0.363	0.048 ***	1.625
(Van) × (Children under 15)	0.946	0.084 ***	5.639
(Large SUV) × (Children under 15)	0.251	0.055 ***	2.339
(Small SUV) × (Children under 15)	0.679	0.111 ***	4.999
(Korea) × (Rural)	-0.597	0.105 ***	3.655
(Seats≥5) × (Household Size≥4)	0.427	0.035 ***	0.222 *
(Mid-Large Car) × (Retired)	0.891	0.049 ***	3.914
(Prestige) × (Retired)	0.429	0.072 ***	1.096
(Import) × (College)	0.437	0.047 ***	1.960
(Prestige) × (Japan) × (College)	-0.096	0.176	9.624
(Prestige) × (Europe) × (College)	-0.607	0.139 ***	11.408
(Prestige) × (Japan) × (Urban)	-0.195	0.154	2.158
(Performance Car) × (College)	0.691	0.113 ***	1.676
Fuel Operating Cost (cents per mile)	-2.946	0.056 ***	0.263 ***
(Fuel Operating Cost) × (College)	-0.027	0.009 ***	0.466
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.064	0.020 ***	0.120
Horsepower/Curb weight	144.232	26.475 ***	93.200
Hybrid	-12.288	0.880 ***	3.364 ***
Curb weight	7.084	0.398 ***	1.907 ***
Wagon	-0.054	1.342	1.403
Mid-Large Car	-0.572	0.615	4.454
Performance Car	-0.517	0.928	2.251
Small-Medium Pickup	6.817	1.201 ***	2.220
Large Pickup	4.777	1.352 ***	3.047
Small-Mid SUV	2.973	0.604 ***	0.885 ***
Large SUV	2.676	1.054 **	4.440

Table D2: BLP model with McFadden's method: parameter estimates

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

D.2. Full results from the vehicle choice application with broad choice data

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.038	0.006 ***	0.052
(Price) × (Income>100,000)	0.123	0.008 ***	0.100
(Price) × (Income Missing)	0.079	0.006 ***	0.056
(Prestige) × (Urban)	-1.253	0.137 ***	0.698 *
(Prestige) × (Income>100,000)	-1.591	0.167 ***	1.381
(Performance Car) × (Income>100,000)	1.518	0.212 ***	2.159
(Japan) × (Urban)	-0.453	0.075 ***	0.662
(Van) × (Children under 15)	0.788	0.144 ***	0.934
(Large SUV) × (Children under 15)	0.707	0.166 ***	1.160
(Small SUV) × (Children under 15)	0.352	0.084 ***	0.584
(Korea) × (Rural)	0.341	0.157 **	0.917
(Seats≥5) × (Household Size≥4)	-1.664	0.351 ***	1.063
(Mid-Large Car) × (Retired)	0.108	0.080	0.287
(Prestige) × (Retired)	-0.415	0.102 ***	0.339
(Import) × (College)	0.362	0.074 ***	0.518
(Prestige) × (Japan) × (College)	-0.294	0.194	1.338
(Prestige) × (Europe) × (College)	-1.747	0.296 ***	2.043
(Prestige) × (Japan) × (Urban)	-0.420	0.177 **	1.156
(Performance Car) × (College)	2.815	0.327 ***	4.622
Fuel Operating Cost (cents per mile)	-0.599	0.048 ***	2.044
(Fuel Operating Cost) × (College)	-0.057	0.013 ***	0.076
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.098	0.008 ***	0.097
Horsepower/Curb weight	20.737	8.960 **	111.690
Hybrid	-2.193	0.727 ***	9.051
Curb weight	0.002	0.000 ***	0.006
Wagon	0.048	0.454	1.209
Mid-Large Car	2.258	0.317 ***	2.185
Performance Car	-3.280	0.392 ***	5.066
Small-Medium Pickup	-1.179	0.464 **	3.228
Large Pickup	-0.777	0.392 **	3.035
Small-Mid SUV	2.888	0.304 ***	1.030 ***
Large SUV	2.697	0.496 ***	1.760

Table D3: BLP model with broad choice data: parameter estimates

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*