Sensitivity of Vehicle Market Share Predictions to Discrete Choice Model Specification

C. Grace Haaf
Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: grace.haaf@gmail.com

Jeremy J. Michalek
Professor
Mechanical Engineering,
Engineering and Public Policy,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: jmichalek@cmu.edu

W. Ross Morrow
Assistant Professor
Mechanical Engineering,
Iowa State University,
Ames, IA 50011
e-mail: wrmorrow@iastate.edu

Yimin Liu
Technical Expert
Ford Motor Company,
Dearborn, MI 48121
e-mail: Yliu59@ford.com

When design decisions are informed by consumer choice models, uncertainty in choice model predictions creates uncertainty for the designer. We investigate the variation and accuracy of market share predictions by characterizing fit and forecast accuracy of discrete choice models for the US light duty new vehicle market. Specifically, we estimate multinomial logit models for 9000 utility functions representative of a large literature in vehicle choice modeling using sales data for years 2004–2006. Each model predicts shares for the 2007 and 2010 markets, and we compare several quantitative measures of model fit and predictive accuracy. We find that (1) our accuracy measures are concordant: model specifications that perform well on one measure tend to also perform well on other measures for both fit and prediction. (2) Even the best discrete choice models exhibit substantial prediction error, stemming largely from limited model fit due to unobserved attributes. A naïve "static" model, assuming share for each vehicle design in the forecast year = share in the last available year, outperforms all 9000 attribute-based models when predicting the full market one year forward, but attribute-based models can predict better for four year forward forecasts or new vehicle designs. (3) Share predictions are sensitive to the presence of utility covariates but less sensitive to covariate form (e.g., miles per gallons versus gallons per mile), and nested and mixed logit specifications do not produce significantly more accurate forecasts. This suggests ambiguity in identifying a unique model form best for design. Furthermore, the models with best predictions do not necessarily have expected coefficient signs, and biased coefficients could misguide design efforts even when overall prediction accuracy for existing markets is maximized.
[DOI: 10.1115/1.4028282]

1 Introduction

Design researchers have proposed a variety of methods to predict the influence of design decisions on firm profit as part of a broader effort to base design decisions explicitly on predictions of downstream consequences for the firm [1]. The majority of these methods apply discrete choice methods [2] to predict consumer choice as a function of product attributes and price. Such predictions are proposed as a way to guide or even optimize design decisions [3–11]. Application of choice models within design implicitly relies on accurate choice predictions [5,12]. Given the many sources of uncertainty in such models, however, Frischknecht et al. [8] question the suitability of using choice models in a design context. At a minimum, researchers must be aware of the degree of prediction error and uncertainty when employing market models in design.

Prediction error can arise from many sources, including noisy data, finite data, omitted variables, changes in preferences or market conditions between estimation and prediction, and misspecification of the choice process [13]. Recent design research has modeled some aspects of model uncertainty by posing distributions over model coefficients [5,12]. Following standard asymptotic results, coefficient distributions are most often assumed to be normal with mean vector and covariance matrix determined by properties of the log-likelihood function. However, model misspecification is virtually guaranteed in most revealed preference contexts, given the complexity of human choice behavior for difficult decisions [14], and standard statistical results do not apply in such settings, nor are they comprehensive. Moreover, few applications of choice modeling in any field carefully analyze sensitivity of model fit or forecast accuracy using alternative utility specifications or error structures that might imply different design decisions. A realistic portrait of these aspects of predictive error cannot be captured in a fully generalizable way across product domains or contexts but can nevertheless be better understood via data-driven examination in the specific market of interest.

We focus on the effect of model specification and characterize share prediction accuracy of multinomial logit models in an empirical study of recent new vehicle markets using revealed preference sales data. The automotive sector is among the most popular product domains for application of choice modeling in general [4,7–9,11,15–44] and in the design literature specifically [4,7–9,15,21,24,27,28,32,35]. Logit models, along with variants including nested and mixed logit models, represent the most popular modeling approach by far. While stated choice methods fit to conjoint survey data are common [3,9,24,27,39–41], they measure hypothetical choices and generally must be calibrated to achieve a match with market sales data [25,45]. We focus here on choice models fit to aggregate market sales data [4,7,8,15–20,22,28,29,32–38,40,43,46].

Given the importance of the vehicle choice application in the design literature and beyond, a better understanding and characterization of prediction accuracy in this domain and its implications for design is needed. We aim to address this need with an automotive case study by fitting a set of models representative of those in the literature to past vehicle sales data, using the resulting models to predict sales in later years, and assessing prediction accuracy.

Our analysis is focused on the following research questions:

(Q1) How should we measure prediction accuracy, and do different measures lead to different conclusions about which models predict best?
(Q2) How widely do predictions vary for alternative model specifications? Which specifications have the best predictions, and how good are they?
(Q3) What are the implications for using choice models in design, particularly of new products?

Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received May 28, 2013; final manuscript received June 23, 2014; published online October 20, 2014. Assoc. Editor: Bernard Yannou.

The design literature has not yet investigated what measures of forecast accuracy exist or compared these measures to understand how they differ in characterizing accuracy, thus Q1. Q2 applies appropriate measures to the specific task in our case study. Q3 focuses on prediction accuracy for new vehicle designs, and we examine the relationship between accurate prediction in existing markets versus potential to predict response to new designs that deviate from market patterns (e.g., correlations with unobserved attributes). We view design as primarily interested in the introduction of new products or (large) changes to product features, motivating a focus on new vehicles.

2 Literature Review

Broadly, there are two schools of research in the vehicle demand literature. The first is concerned foremost with predicting future vehicle demand shares, usually at an aggregate level like vehicle class or powertrain type, and often without transparency about the assumptions and models used to make the forecast. We henceforth refer to this type of literature as "forecasting". The second school is interested in model construction and in vehicle and consumer attributes coefficient estimation especially as it pertains to willingness-to-pay and demand elasticity in past markets. We henceforth refer to this type of literature as "explanatory." Appendix A compares publications of each type.[1]

Forecasting studies are conducted by private or government research entities or issued in report format from an academic research institute (see Appendix A). Reports are typically not peer reviewed and rarely contain a full mathematical description of the model, making it impossible to reproduce the model without additional information. Some reports include sensitivity cases formed with variations on model assumptions; for example, the Energy Information Administration Annual Energy Outlook [47] contains base, low, and high alternative vehicle future market share as a result of base, low, and high future oil prices. This type of sensitivity only captures uncertainty about model input parameters and assumes that model specification and estimated coefficients are known. In practice, model specifications for choice contexts as complex as automotive purchases are always uncertain, and the relevant question is whether or not the model is sufficient for its intended function. The forecasting literature is typically not used in engineering design models due to lack of transparency and documentation of data and modeling assumptions and lack of models that make predictions as a function of design variables. Rather, models from the explanatory literature are applied in a predictive context.

The bulk of the new vehicle purchase demand literature is explanatory, conducted by academic researchers and published in peer-reviewed academic journals (see Appendix A). This literature extensively discusses model estimation and to a lesser degree model selection, including potential sources of error from model misspecification. Usually researchers compare the goodness-of-fit across several specifications in order to determine which model best represents a known, current reality. However, most of this literature does not attempt to make predictions about future vehicle market share penetration or evaluate models with predictive capabilities in mind (Frischknecht et al. [8] is a rare exception). In general, models that fit the existing data best may not necessarily be the best at predicting counterfactuals: statistical models may be misspecified, containing systematic difference in prediction from true process ("bias"), or may be sensitive to overfitting noise in the data instead of signal ("variance") [48].

The earliest applications of economic models for overall automotive demand focused on macroeconomic variables and, as Train [49] highlights, only included price. These studies are referred to as aggregate studies because the level of granularity of predictions is at the whole market or vehicle class level as opposed to individual vehicle designs[2]. Disaggregate studies evolved to predict the number of vehicles an individual household would choose to own [49]. For example, Lave and Train [44] advanced this work by proposing a disaggregate model of vehicle class purchase choice based on consumer characteristics and additional vehicle characteristics, such as fuel economy, weight, size, number of seats, and horsepower. A wide variety of models followed over the next three decades: Boyd and Mellman [43], who propose a random coefficient logit model adopted by others [11,28,35,50,51]; Berry et al. [42], who include an alternative-specific constant (ACS) in the utility function of a random coefficient demand model adopted by others [16,52–54]; Brownstone and Train [39], who propose several choice model specifications using the results of a California conjoint study described in Bunch et al. [55] and adopted by others [56,57]; and Whitefoot and Skerlos [11], who investigate the effect of fuel economy standards on vehicle size and employ a logit model with coefficients drawn directly from the literature. Other new-vehicle purchase models include [23,26,29,31,34,40,41,58].

We use the preceding literature to inform comparison models of our creation; we do not recreate prior models exactly due to limited availability of data or specifics about estimation methods. Instead, we form a combinatorial set of utility specifications using covariate forms from these prior models, fit them all to a common data set, and test them all on a common prediction set. Appendix B summarizes the covariates used in past models and those adopted for our tests.

3 Methods

Our overall goals are to examine the robustness of multinomial logit model predictions over various utility function specifications and to compare the predictions across the structural specifications of logit, mixed logit, and nested logit (for brevity we refer to the multinomial logit model as "logit"). We identify a universe of covariates informed by the literature and form combinations of them such that we have defined all possible linear utility function specifications from these covariates. We then estimate the logit coefficients on US consumer vehicle purchase data from 2004 to 2006 and predict market share for each of the vehicles in the US purchase data from 2007 and 2010.

Using the measures described in Sec. 3.4, we rank the predictive accuracy across utility function specification for each of the measures.

3.1 The Data Set. Our data set draws vehicle attribute information from Ward's Automotive Index [59] and aggregate US sales data from Polk [60] for vehicle sales during 2004–2007 and 2010. Other studies have used a variety of data sources (including these) as well as stated preference surveys. We use 2004–2006 data for estimation because we expect three years of data to be sufficient to predict a successive year, and we predict 2007 and 2010 sales to examine the effects of different time horizons. We implicitly assume that all individuals who purchased a vehicle considered all of the other vehicles available in the same year and made a compensatory decision based on vehicle attributes.

Our models consider only new vehicle buyers, thus there is no outside good (option to not purchase any vehicle). Inclusion of an outside good allows a choice model to endogenously determine market size. Excluding it models only share among the vehicles purchased, which is likely less sensitive to macroeconomic factors. There are many factors that drive share and are not included in our models, but we are interested in how well a modeler can predict when relying primarily on available vehicle attribute data.

3.2 Model Specification. Each model uses the utility function

[1]An electronic companion to this paper containing the appendices referenced herein can be found at http://repository.cmu.edu/meche/70/

[2]We use the term "vehicle design" to refer to vehicle make-model.

$$u_{ij} = \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_{ij} \quad (1)$$

where u_{ij} is the utility of vehicle design j for consumer i, $\mathbf{x}_j$ is the attribute vector of vehicle j, $\boldsymbol{\beta}$ is the vector of model parameters to be estimated, and ε_{ij} is an error term. Following standard assumptions, if ε_{ij} is independently identically distributed (iid) and follows a type I extreme value distribution, then the probability P_j that a randomly selected consumer will choose vehicle j can be expressed as

$$P_j = \frac{\exp\left(\mathbf{x}_j'\boldsymbol{\beta}\right)}{\sum_{k=1}^{J}\exp\left(\mathbf{x}_k'\boldsymbol{\beta}\right)} \quad (2)$$

where J is the number of vehicle design options. This is the (multinomial) logit formula.

While any choice of covariates $\mathbf{x}$ is possible in principle, we focus on combinations of covariates used in the prior literature. We survey the automotive demand literature to identify the universe of independent variables historically used in automotive discrete choice models (Appendix B). From this list of candidate covariates, we select a subset to define a manageable set of models. Many of the models in the literature include demographic or consumer usage covariates, but because Polk sales data [60] does not include individual-level choices, we ignore demographics. For some demographic information like gender or income an aggregate distribution over the US population is available, but because we do not know which consumers selected which vehicles, sampled consumer attributes are unlikely to accurately determine specific individuals' sensitivity to vehicle attributes. We omit several variables because they are not available in our data sources:

- *Indirect vehicle attributes* like consumer reports ratings for handling and safety—These would be unknown at the time of prediction.
- *Vehicle and battery maintenance costs*—These covariates are used primarily when predicting alternative vehicle share, and they will not vary substantially across conventional and hybrid powertrains.
- *Acceleration time* (*seconds*)—We indirectly test inclusion of acceleration through functions of horsepower and weight. Note that horsepower/weight correlates well with 0–60 mph acceleration time for cars well but poorly for trucks.
- *Range*—This covariate is used primarily when predicting alternative vehicle share and will not vary substantially across conventional and hybrid powertrains. A related fuel economy covariate is included.
- *Top speed*—We use an alternative measure of performance through horsepower and weight.
- *Number of seats*—We use vehicle class, which is closely related to seating.
- *2-year retained value*—Like the consumer rating data this would not be known at the time of prediction.
- *Attributes specific to alternative-vehicles* (e.g., dummies for hybrid or electric power trains)—These are not relevant to our data set, which includes conventional vehicles and only a limited number of hybrid powertrains.

The highlighted covariates in Appendix B are those which remain after omitting demographic, usage, indirect, and unavailable attributes. Some studies group price and fuel economy variables into discrete levels of each rather than treating them as continuous variables. We consider all covariates (except for class and brand dummies) to be continuous variables because, unlike controlled conjoint experiments, the market data do not fit well into a small number of discrete levels. Price is always included as a covariate and can take any of the forms listed in Table 1; vehicle class dummies are also always included. The other highlighted covariates in Appendix B can take one of the forms listed in Table 1 or can be excluded from the utility function entirely ("excluded" option). Given these covariate options, there are 9000 possible utility specifications for the logit model outlined in Table 1. Operating cost includes the macroeconomic variable of retail gas price. Though we aim to exclude nonvehicle attributes, this covariate was particularly prevalent in the literature. Furthermore, while having more covariates cannot decrease best model fit on a given data set, that does not imply that more covariates will improve model forecast accuracy. In general, introducing more covariates introduces the risk of overfitting the estimation data.

From the selected covariates, we assume that the utility function is linear in parameters (a standard assumption in the vast majority of logit model applications because it ensures that the log-likelihood function is concave [2]) and construct models using all possible linear combinations of covariates.

Many of these covariates are correlated. Such correlations can induce bias in the estimated coefficients if not corrected [63]. However, while this presents difficulties in drawing inferences from the coefficients (e.g., willingness-to-pay) it does not necessarily affect the ability to make predictions from the model so

Table 1 Covariate forms tested in utility function specifications

	Functional form options				
Covariate	Option 0	Option 1	Option 2	Option 3	Option 4
Price		Price ($)	Price + op cost	ln(price)	
Operating cost[a]	Excluded	Fuel cost/mile	Miles/fuel cost	Miles/gallon	Gallons/mile
Acceleration[b]	Excluded	Horsepower/weight (hp/wt)	wt/hp	$\exp(c_1 \times (\text{hp/wt})\, c_2)$	hp
Size	Excluded	Length	Width	Length-width	Length × width
Style	Excluded	(Length × width)/height			
Air conditioning	Excluded	Dummy if air-conditioning is standard			
Transmission	Excluded	Dummy if auto. transmission is standard			
Brand	Excluded	Dummy for country of origin[c]	Dummy for brand[d]		
Vehicle class		Dummies for vehicle class[e]			

[a]Fuel cost is average annual gas price [61] in 2004 dollars, adjustment based on the consumer price index [62].
[b]$c_1 = -0.00275$ and $c_2 = -0.776$ as in the EIA Annual Energy Outlook [47].
[c]Country of origin includes: United States, Europe, and Asia; excludes United States dummy for identification.
[d]Brand includes: Acura, Audi, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ford, GMC, Honda, Hummer, Hyundai, Infiniti, Isuzu, Jaguar, Jeep, Kia, Land Rover, Lexus, Lincoln, Mazda, Mercedes, Mercury, Mitsubishi, Nissan, Oldsmobile, Pontiac, Porsche, Saab, Saturn, Scion, Subaru, Suzuki, Toyota, Volkswagen, Volvo; excludes Acura dummy for identification.
[e]Class includes: Compact, midsize sedan, full size sedan, luxury sedan, SUV, luxury SUV, pickup, minivan, van, and sports; van is excluded for identification.

long as the correlations in the training data would also be present in the prediction set. For vehicle markets, this is likely to hold for near-term predictions, though it may not hold for new designs that do not follow prior patterns in the marketplace.

For illustration of this concept, suppose the true choice generator uses the utility function $u(\mathbf{x}|\boldsymbol{\beta}_0) = \boldsymbol{\beta}_0'\mathbf{x} + \varepsilon$, and the designs in the market follow a pattern: $\mathbf{x} = \mathbf{A}'\mathbf{y}$ for $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, m < n$. Then for any coefficient vector $\{\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \boldsymbol{\Delta} : \mathbf{A}\boldsymbol{\Delta} = 0\}$, $u(\mathbf{x}|\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{A}'\mathbf{y} + \varepsilon = (\mathbf{A}\boldsymbol{\beta}_0 + \mathbf{A}\boldsymbol{\Delta})'\mathbf{y} + \varepsilon = \boldsymbol{\beta}_0'\mathbf{x} + \varepsilon = u(\mathbf{x}|\boldsymbol{\beta}_0)$. Therefore, choice probabilities are identical for any $\boldsymbol{\Delta}$ in the null space of $\mathbf{A}$, and $\boldsymbol{\beta}_0$ is not identifiable: coefficient estimates $\boldsymbol{\beta}$ could be arbitrarily far from their true value $\boldsymbol{\beta}_0$. Nevertheless, $u(\mathbf{x}|\boldsymbol{\beta}) = u(\mathbf{x}|\boldsymbol{\beta}_0)$, so utility estimates (and therefore choice probabilities) can be correct even for arbitrarily biased coefficients as long as the new designs follow the pattern in the marketplace $\mathbf{x} = \mathbf{A}'\mathbf{y}$. If a new design deviates from the prior pattern $\tilde{\mathbf{x}} = \mathbf{A}'\mathbf{y} + \mathbf{z}$, utility (and therefore choice probabilities) may be biased: $u(\tilde{\mathbf{x}}|\boldsymbol{\beta}) = (\boldsymbol{\beta}_0+\boldsymbol{\Delta})'(\mathbf{A}'\mathbf{y}+\mathbf{z})+\varepsilon = (\mathbf{A}\boldsymbol{\beta}_0+\mathbf{A}\boldsymbol{\Delta})'\mathbf{y} + (\boldsymbol{\beta}_0 + \boldsymbol{\Delta})'\mathbf{z} + \varepsilon = \boldsymbol{\beta}_0'\mathbf{A}'\mathbf{y} + \boldsymbol{\beta}_0'\mathbf{z} + \boldsymbol{\Delta}'z + \varepsilon = u(\tilde{\mathbf{x}}|\boldsymbol{\beta}_0) + \boldsymbol{\Delta}'z$. Therefore, models that predict well overall may nevertheless have biased coefficients that predict poorly for new designs that deviate from the market pattern. We assess predictive accuracy for products in the marketplace and also examine variation in implications of coefficient estimates for new designs.

3.3 Model Estimation.

The likelihood of the estimated parameters L is defined as the probability of generating the observed data given the estimated parameter values

$$L\left(\hat{\boldsymbol{\beta}}|\mathbf{x}\right) = \prod_{j=1}^{J} \left(P_j\right)^{n_j} \tag{3}$$

where n_j is the sales of vehicle j. The maximum likelihood estimator of the parameters $\hat{\boldsymbol{\beta}}$ is the value of the vector that maximizes L. The monotonic transformation $\ln(L)$ is typically used as the objective function for computational benefit. For more detail on logit models and their estimation see Train [2].

The mixed logit, or random coefficients logit, model is similar to the logit model except the individual β's are allowed to vary over the population to represent heterogeneous consumer preferences. In our case we assume that they are independently normally distributed

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix, and the maximum likelihood procedure estimates the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using numerical integration [2]. This specification relaxes the independence from irrelevant alternatives (IIA) restriction for substitution patterns [2].

Our nested logit specification divides the vehicles into groups or nests by vehicle class and fits a logit model to each of the nests. We assume that the utility functional form is the same for each nest, but coefficients may differ across nests. For example, the β for price will be different for midsize cars than it is for pickups. However, within a nest β is fixed. A nested logit exhibits the IIA property for products within a nest, but relaxes the IIA restriction for products in different nests.

As generalizations of the logit model, nested and mixed logit models will necessarily fit any set of *estimation* data at least as well as the logit. The mixed logit generalization of the logit model is even flexible enough to represent most random utility maximization models, given enough flexibility over the coefficient distribution [2]. However, nested and mixed logit models need not *predict* as well as logit models due to the potential for overfitting.

3.4 Evaluation Measures.

After fitting each of the model specifications, we evaluate prediction error using likelihood measures, the Kullback–Leibler divergence (KL) [64], a cumulative distribution of error tolerance (CDFET), and the average share error (ASE), and we compare the goodness-of-fit using the above measures as well as the Akaike information criterion (AIC) [65], and the Bayesian information criterion (BIC) [66]. Each of these measures is described below. We compare models selected as best by these measures to one another and to literature-informed benchmark models.

Likelihood: Likelihood, defined in Eq. (3), and monotonic transformations of likelihood, such as log-likelihood $\ln(L)$ and average likelihood (AL) ($L^{1/N}$, where N is the number of choices observed) measure the probability that the model would generate the data observed. When comparing two models for the same data set, the model with larger L is more likely to generate the data observed.

KL divergence: The KL divergence measures the difference between a predicted distribution and the true distribution [67].

$$KL\left(s_j||P_j\right) = \sum_{j=1}^{J} \ln\left(\frac{s_j}{P_j}\right) s_j \tag{5}$$

where $s_j = n_j/J$ is the market share of vehicle design j. The KL measure is also a monotonic transformation of L, thus L and KL will rank models identically, and maximizing likelihood is equivalent to minimizing KL (see Appendix C for proof).

ASE: ASE measures the average error in share predictions across the vehicle designs.

$$\mathrm{ASE} = \frac{1}{J}\sum_{j=1}^{J} \left|s_j - P_j\right| \tag{6}$$

We report ASE as a summary statistic in Appendix D but do not use it as a basis for model selection because it does not holistically capture distribution divergence: It will not distinguish between models with large error for one vehicle alternative vs. the same degree of error spread out among many vehicle alternatives.

Error tolerance cumulative distribution function (CDF): The CDFET graphs the fraction of vehicles with absolute share prediction error, $|s_j - P_j|$ for vehicle design j, less than a specified value. This measure, to our knowledge proposed here, evaluates a model in terms of error tolerance levels. We use absolute share error rather than relative error because relative error overemphasizes small prediction errors for vehicles with small market share. A CDFET is a more comprehensive description of model prediction error than likelihood measures because it characterizes the distribution of accuracy across the vehicle share predictions, rather than just how well a model predicts "on average".

Two additional measures apply only to assess fit with estimation data, not predictive accuracy [68].

AIC: AIC is a variation of likelihood that attempts to penalize overfitting.

$$\mathrm{AIC} = 2\ln(L) - 2k \tag{7}$$

where k is the number of model parameters.

BIC: BIC is similar to AIC but with a stronger penalty for an increasing number of covariates.

$$\mathrm{BIC} = 2\ln(L) - \ln(J)k \tag{8}$$

AIC and BIC can take on the value of any negative real number, have no standalone meaning, and are only useful as compared to other candidate models fit to the same data set. Larger values are preferred. Derivations and consistency proofs for the KL, AIC, and BIC measures can be found in Ref. [68].

4 Results

Of the 9000 tested utility function specifications, for 8993 (99.9%) the Knitro optimization algorithm for MATLAB converged

Table 2 RAL calculated on the prediction data set for select model specifications and data sets

Scenario	1	2	3	4	5	6
Estimation data	2004–2006	2006	2007	2004–2006	2004–2006	2004–2006
Prediction data	2007	2007	2007	2010	2007	2007
Market	Full market	Full market	Full market	Full market	Luxury sedan	New designs
AL of ideal model (predicted shares = actual shares)	*0.0076*	*0.0076*	*0.0076*	*0.0080*	*0.0384*	*0.4610*
RAL of no info model	55.3%	55.3%	55.3%	43.6%	63.6%	93.2%
RAL of static model	**88.3**%	**88.3**%	**88.3**%	23.7%	73.3%	95.9%
RAL of class dummies only logit	65.9%	65.9%	65.9%	53.6%	NA	95.0%
RAL of best fit logit model for L/AIC/BIC of estimation data	76.4%	77.7%	81.7%	67.3%	73.3%	96.5%
RAL of logit model with greatest likelihood for prediction data	79.0%	79.0%	81.7%	**68.5**%	87.9%	**97.5**%
RAL of mixed logit with best logit estimation fit covariates	79.0%	79.0%	85.6%	67.3%	**89.2**%	97.0%
RAL of nested logit with best logit estimation fit covariates	73.8%	72.4%	80.3%	64.8%	NA	95.3%

Note: in Scenario 5 luxury sedan vehicles are used for estimation and prediction; in scenario 6 the full market is used for estimation, but evaluation measures are assessed for prediction of new vehicles only. Italicized numbers emphasize that the number is an AL as opposed to an RAL. Bold numbers indicate the greatest RAL in a given column (most accurate model for a given scenario).

to likelihood-maximizing coefficients, and the other seven failed to converge. Only the 8993 models that successfully converged were considered as candidate models. The candidate models were ranked from best to worst on each measure. There were no two models with identical values for any measure (no ties). In the following results "best models" refer to the models ranked as number one for a given measure.

4.1 Q1: Model and Evaluation Measure Comparison. We refer to a model that most accurately predicts the in-sample estimation data according to a given measure as the "best estimative model", and we refer to a model that most accurately predicts the out-of-sample prediction data as the "best predictive model." The traditional goodness-of-fit measures—likelihood/KL and AIC/BIC—select the same best estimative model, and they also agree upon the specification of the best predictive model. The CDFET goodness-of-prediction measure selects distinct model specifications as the best predictive models dependent upon the desired error tolerance level (we test error tolerance levels of 25%, 50%, and 75%). The three CDFET best predictive models are also distinct from the best estimative and predictive models under the AIC, BIC, and likelihood criteria. See Appendix D for selected model measure comparisons and coefficient estimates.

Though the best likelihood/AIC/BIC estimative model is distinct from the best predictive model, the difference in form is small. They include the same covariates but in different forms (e.g., operating cost as miles/dollar as opposed to gallons/mile) with the exception of luxury and transmission which contribute little to utility relative to the contribution of the other attributes.

4.2 Q2: Model Accuracy. Table 2 summarizes the AL calculated on the prediction data set for select combinations of model specification (rows) and estimation/prediction data set scenarios (columns). We report the *relative* average likelihood (RAL) in Table 2 defined as the AL of the model divided by the AL of an ideal aggregate model that predicts shares perfectly. The reason we report RAL instead of simply AL is because choice diversity in the data necessarily lowers the maximum attainable value of AL with any model. Thus RAL describes the amount of predictive power obtained by a particular model relative to the best possible predictive power that could be obtained with any aggregate model.

The rows compare the predictive performance of the model that has the best predictions and the model that fit the estimation data best. Using each of the utility functions from the best estimative logit models, we fit additional mixed and nested logit models. Due to computational limitations, we did not run all 9000 utility form combinations for the mixed and nested logit structural specifications. Rather we used the results from the logit model output to inform the selection of covariate form for the mixed and nested logit models. The "no info" row is calculated by assigning an equal share to all vehicles. The static model row assumes that shares in the prediction year are identical to the most recent share of the vehicle design available in the estimation data for all vehicle designs present in both the estimation and prediction data, and all new vehicle designs receive an equal proportion of the remaining share.

Scenario 1 is our base case, where models are fit to sales in years 2004–2006 and used to predict 2007 sales. Scenario 2 uses only 2006 data to predict 2007, assessing sensitivity of predictions to the amount of data used for estimation. Scenario 3 fits the models directly to 2007 data, helping to identify the portion of prediction error that stems from model fit, rather than from changes over time. Scenario 4 uses 2004–2006 data to predict 2010 sales, assessing differences when predictions are made farther into the future. Scenario 5 assesses predicative accuracy for a single vehicle class,[3] rather than the entire market, and scenario 6 assesses only the predictive accuracy for new vehicle designs introduced in 2007. Comparisons can be made within each column to evaluate the prediction accuracy across model specifications for a given estimation/prediction data set.

In scenarios 1–3, which predict the full 2007 market, the best predictive logit model predicts better than the best estimative model, the class dummies model does not predict as well as the models which contain vehicle attributes, and the no info model predicts worst, as expected. Nested logit predictions have lower AL than logit, but mixed logit predictions have higher AL.[4] That the nested logit does not predict better than the logit suggests that the relaxation of the IIA property among the nests selected does not improve prediction. Model predictions could potentially be improved further by exploring alternative parameter distributional forms such as multivariate normal with a full covariance matrix [69], although that introduces more potential for overfitting with aggregate sales data. We leave such explorations for future work. See Appendix E for mixed and nested logit coefficient estimates and Appendix F for actual versus predicted shares.

In all three scenarios the static model outperforms all other models. Additionally, we see little difference in prediction quality between scenarios 1 and 3 when using the same model (compare across columns) compared to the difference due to model specification (compare down rows), even though the prediction set and

[3]This is distinct from the "class dummies only logit" which includes data for the entire market but uses only dummies representing each class as covariates.

[4]A likelihood ratio test of the best logit and mixed logit models calculated on 2007 data suggests that there is sufficient evidence to reject the null hypothesis that the mixed logit model predicts significantly better at the $\alpha = 0.1$ level.

estimation set are identical in scenario 3. Together, these results indicate that residual error in model fit is a major source of prediction error, and there is too much missing data or model misspecification in the attribute-based models to fit or predict the full market as well as the static model. Without data on missing covariates that influence choice, such as vehicle aesthetics, it is difficult to fully explain choice behavior at the vehicle design level with only the available covariates.

However, scenario 4 examines a longer time horizon and reveals that the static model has poor predictive capability when forecasting farther into the future. The attribute-based models attempt to capture consumer choice as a function of observable attributes plus random noise, but since not all attributes are observed, share is not fit perfectly. In contrast, the static model does not attempt to explain the reason for consumers' choices but instead simply assumes consumers will make the same choices year after year. The static model does well for the 2007 forecasts because share for each vehicle model changes little from year to year, but over a longer time horizon vehicle designs change and new designs are added to the market (~37% of the vehicle designs sold in 2010 did not appear in the 2004–2006 data). The static model has no information about these new designs, so it loses predictive capability, and over a longer prediction horizon the attribute-based models perform substantially better than the static model.

Scenario 5 indicates that the attribute-based models also perform better than the static model in the luxury sedan class. The best class model is distinct for each class, though all class models include some form of all covariates with the exception of style and automatic transmission as standard. The AL of 2007 class predictions increases when the best estimative class level model is fit to class data as opposed to the best estimative full market model fit to the full market data with the exception of midsize and sports cars (see Appendix G for table of class model specifications and model AL comparison by class).

Figure 1(*a*) shows the CDFET for selected models of scenario 1. The *x*-axis is the absolute difference between the predicted share and the actual share, and the *y*-axis is the proportion of vehicle designs whose share prediction error is less than the corresponding value on the *x*-axis. For example, in Fig. 1(*a*) point (0.25%, 0.7) indicates that 70% of the share predictions made by the best AIC/BIC/KL models deviate from the observed share by less than 0.25% (the average vehicle design share in this market is 0.42%).

The worst models all perform similarly to one another in scenario 1 and lie on top of the class-only curve in Fig. 1(*a*) (and are thus omitted for readability). While a model could plausibly be posed that predicts worse than the no info model, we do not observe it in our utility specifications. The best models and worst models differ most noticeably in their omission of covariates. The best models include some form of almost every covariate, whereas the worst models omit covariates entirely. For example, the worst model as selected by the likelihood and AIC measures applied to the estimated data only contains the covariates price and class. Conversely, if we compare only models that contain some form of price, operating cost, acceleration, size covariates, and class and brand dummies (style, luxury, and automatic transmission dummies could be excluded), then we see no practical difference in the predictive power of the best and worst models. No one covariate in isolation sets the best models apart from the worst models. A model's predictive power thus appears to be robust to covariate form but sensitive to the exclusion of attributes.

4.3 Q3: Implications for Design.

Scenario 6 compares the best-predictive logit model for all vehicles to the model that best predicts the shares of the new vehicle designs introduced in 2007. The best new vehicle model is determined similarly to the best predictive logit model of scenarios 1–3 by ranking the models on each of the measures; however, the measures in this case were calculated by treating each of the new vehicles individually and the holdover vehicles as an aggregated "other" share. (The "other" share is calculated as the sum of all holdover vehicle shares.) In contrast to scenario 1, the attribute-driven logit models of scenario 6 have a higher likelihood than the static model, since the static model has no information about new designs.

The CDFET of Fig. 1(*b*) shows that at lower values of error tolerance the attribute-driven models are superior to the static model and that there is some difference in prediction quality between models that predict best for the whole market versus the new vehicle market. Overall, while the static model outperforms attribute-based models for near-term predictions, attribute-based models are needed for predicting the performance of new vehicle designs and for making longer-term predictions. Still, the degree of uncertainty and error in predictions for new designs may be too large to guide design choices appropriately in some contexts.

Appendix D summarizes model coefficients for several specifications including those representative of models in the literature as well as best estimative and best predictive models. It is clear that different specifications lead to different inferences about the effect of attribute changes on choice. For instance, the utility function specifications based on Boyd and Mellman [43], Berry et al.

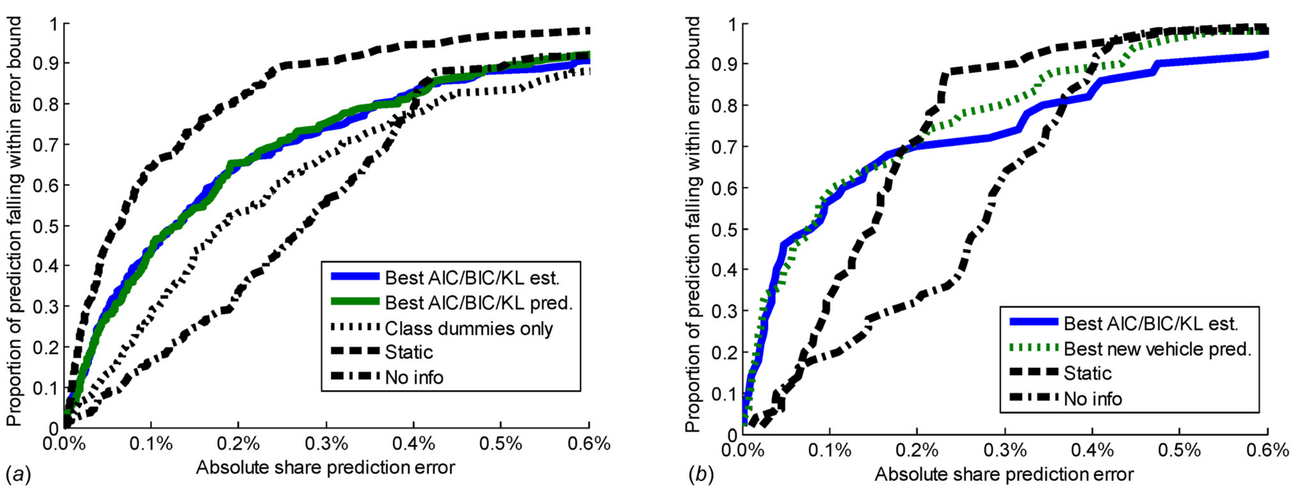

Fig. 1 CDF of error tolerance for the best logit model specifications as measured by likelihood/KL and AIC/BIC measures on 2004–2006 sales estimation data and 2007 sales prediction data compared to alternative models (*a*) full market and (*b*) new vehicle designs only

[42], and Whitefoot and Skerlos [11] result in a coefficient for operating cost that suggests consumers prefer higher efficiency (longer range per unit cost or lower fuel consumption per unit distance) all else being equal, as expected. But the best estimative and best predictive models suggest that consumers prefer lower efficiency. This can happen because efficiency may serve as a proxy for unobserved variables (e.g., size, performance, or styling variables not captured in the data). While the latter models make better predictions for existing vehicle markets that follow established patterns (attribute correlations), they could misguide design efforts that divert from established market patterns.

5 Limitations

Our investigation is a first step in a larger goal of characterizing the design impacts of choice prediction uncertainty. All of our models have error resulting from misspecification and missing information (as do all similar models in the literature that are based on market sales data rather than controlled experiments). For example, we do not have information on attributes that are important in some vehicle classes (like towing capacity for trucks), and we lack information and quantification of some key purchase drivers, such as esthetics. We lack individual-level choice data with consumer covariates, such as demographics or usage variables [9], which can help explain choice behavior and improve predictions when predictions of future population covariates are available. Nevertheless, such limitations are common in choice models used to assess the vehicle market or guide design choices. Our study suggests that if models lack transparent quantifications of important determinants of product choices, designers should be cautious about basing design decisions on choice models.

More research is needed to assess a wider scope of modeling alternatives. We did not consider ASCs—product-specific factors that can proxy for omitted variables—and their use in prediction or design. ASCs can generate models that match estimation data shares exactly; however, they contain no information about specific unobserved product features, and they are unknown for any new product designs. We also ignore a major component of the new vehicle modeling literature: covariate endogeneity—a correlation between model covariates and the unobserved terms like error. Endogeneity implies that coefficients are biased and inconsistent if not properly estimated, typically requiring instrumental variables techniques [2]. We also did not consider alternative estimation methods (e.g., Bayesian methods) and alternative heterogeneity specifications (e.g., latent class models, a mixed logit model with joint parameter distributions, mixture models, and generalized logit models that account for scale and coefficient heterogeneity [69]).

Our study uses random utility discrete choice models that treat consumers as observant rational utility maximizers with consistent preferences. While this is a popular approach to modeling consumer choice, important criticisms exist. For instance, preferences can evolve over time [25], changing with cultural symbolism [70] and/or social interactions [71]. The theory of construction of preference adapted to design by MacDonald et al. [14] suggests that consumers' preferences for attributes do not exist a priori but are rather evaluated on a case-by-case basis [14]. Morrow and MacDonald [10] suggest that vehicle choice behavior may be better represented by a "consider-then-choose" model [72] where consumers first screen out most alternatives using simple rules, subsequently maximizing utility over a smaller "consideration set" [73]. The potential value of this type of model is suggested here by the better performance of class-only models, a special case of the consider-then-choose model. More broadly, the Lucas critique warns against use of aggregated historical data to predict outcomes in counterfactual future scenarios [74].

6 Conclusions

While the topic of uncertainty associated with choice predictions is widely discussed in the design community (e.g., Refs. [3–5,12,14,28]), there is no current consensus as to what processes and measures best quantify model uncertainty. This gap motivated our first research question, Q1. We investigated several well-known measures of model performance evaluated on a prediction set. For the automotive case study examined, likelihood measures (and the rank-equivalent KL divergence measure) tend to identify the same top-ranked model as the penalized likelihood measures AIC and BIC do. While CDFET measures identify different top-ranked models, depending on the error tolerance selected, the resulting models share most covariates. Models that perform well on one measure tend to perform well on the other measures, and models that perform poorly on one measure also tend to perform poorly on the other measures. In other words, determination of the best models in our study did not depend strongly on potentially arbitrary selection of the measure used to evaluate predictive accuracy.

Overall our results confirm several intuitive features of this application: attribute-based models predict better than models with no information; models of a particular vehicle class typically make better predictions than models of the full market; including more covariates generally improves predictive accuracy; and better model fit correlates well with better predictive accuracy. The match between fit and predictive accuracy, suggesting no major overfitting issues, is particularly encouraging, since the modeler has access to choice data for estimation but not choice data in the counterfactual predictive context. These findings would have to be validated in other product domains on a case-by-case basis.

We also observe a number of less intuitive results that are relevant to design. First, the models we construct are fairly poor predictors of future shares. In our base scenario, our best predictive model has an average error of 0.24% (the average share of a vehicle design is 0.42%), which translates to an error of approximately 37,500 vehicles sold for the 2007 market. The limited predictive power of standard models on real data in a canonical product category suggests designers should apply discrete choice models cautiously, though predictions may be substantially better in domains with fewer unobserved attributes or with conjoint data (where all attributes are observed).

Second, we find that attribute-based models do not furnish the best predictions for short-run forecasts in stable market conditions; attribute-based models estimated on 2004–2006 data were outperformed in predicting 2007 shares by the "static" model that assumes no changes in shares. However, attribute-based models are superior to the static model when predicting new vehicles only, since the static model lacks information about new entrants. There are some intuitive reasons why the static model might perform better than attribute-based models for short term predictions of existing designs given relatively stable market conditions. First, the static model may implicitly capture effects related to omitted vehicle attributes neglected by attribute-based models. Second, the static model may predict well in the short-run simply because of "inertial" conditions specific to the automotive market, particularly multiperiod production schedules and inventory buildup that must ultimately be cleared over the short run using unobserved advertising and/or purchasing incentives.

Third, while including an appropriate set of product attributes as model covariates is important to improving predictive accuracy, the form those covariates take in the utility function is less important in this application. This implies that it may be less important to test many variations of utility function covariate form when constructing a model, but it also means that any design decisions (e.g., design optimization results) that are not robust to variation in utility function covariate form may not be justified given the near equivalence of alternative covariate form in fit and prediction error with market data. If different utility specifications lead to different design decisions but the data cannot discern which form best represents choices, then design decisions cannot be reliably based on any single specification.

Finally, we observe that some of the models with the best predictive accuracy have coefficients with unexpected signs—likely biased due to correlation with unobserved attributes. Despite good prediction accuracy in existing markets, where attribute correlations are similar from year to year, these models may misguide design efforts if the designer makes changes that do not follow correlations in the marketplace. For example, the sign of the coefficient for the gallons per mile (gpm) attribute of the best predictive logit model is negative,[5] suggesting that consumers prefer lower fuel economy, all other attributes being equal. In fact, consumers may purchase vehicles with lower fuel economy because of other features of those vehicles unobserved by the modeler (e.g., size, performance, or styling attributes not captured in the model). The model predicts well if the new market retains such correlations, but a designer who lowers fuel economy alone is not likely to obtain the outcome predicted by the model. Thus, accuracy of predictions in existing markets is not a sufficient condition for use in design.

To verify that our results are not specific to the 2004–2006 timeframe, we conducted a similar analysis with estimation data from years 1971–1973 and 1981–1983 with prediction data from the respective one and four year forward markets. We find that our conclusions are generally robust to alternate timeframes: our accuracy measures are concordant; the best models exhibit substantial prediction error stemming from limited model fit; the static model outperforms the attribute-based models when predicting the full market one year forward but attribute-based models can predict better for four year forward forecasts or new vehicle designs; share predictions are sensitive to the presence of utility covariates but less sensitive to covariate form; nested and mixed logit specifications do not produce significantly more accurate forecasts; and the 1971–1973 models with best predictions do not necessarily have expected coefficient signs (though 1981–1983 models do). See Appendix H of the supplemental material for additional detail.

References

[1] Hazelrigg, G. A., 1998, "A Framework for Decision-Based Engineering Design,"ASME J. Mech. Des., **120**(4), pp. 653–658.

[2] Train, K. E., 2009, *Discrete Choice Models With Simulation*, Cambridge University, New York.

[3] Wassenaar, H. J., and Chen, W., 2003, "An Approach to Decision-Based Design With Discrete Choice Analysis for Demand Modeling," ASME J. Mech. Des., **125**(3), pp. 490–497.

[4] Wassenaar, H. J., Chen, W., Cheng, J., and Sudjianto, A., 2005, "Enhancing Discrete Choice Demand Modeling for Decision-Based Design," ASME J. Mech. Des., **127**(4), pp. 514–523.

[5] Besharati, B., Luo, L., Azarm, S., and Kannan, P. K., 2006, "Multi-Objective Single Product Robust Optimization: An Integrated Design and Marketing Approach," ASME J. Mech. Des., **128**(4), pp. 884–892.

[6] Williams, N., Azarm, S., and Kannan, P. K., 2008, "Engineering Product Design Optimization for Retail Channel Acceptance," ASME J. Mech. Des., **130**(6), p. 061402.

[7] Shiau, C.-S. N., and Michalek, J. J., 2009, "Optimal Product Design Under Price Competition," ASME J. Mech. Des., **131**(7), p. 071003.

[8] Frischknecht, B. D., Whitefoot, K., and Papalambros, P. Y., 2010, "On the Suitability of Econometric Demand Models in Design for Market Systems," ASME J. Mech. Des., **132**(12), p. 121007.

[9] He, L., Chen, W., Hoyle, C., and Yannou, B., 2012, "Choice Modeling for Usage Context-Based Design," ASME J. Mech. Des., **134**(3), p. 031007.

[10] Morrow, W. R., Long, M., and MacDonald, E. F., 2012, "Models in Decision Based Design Optimization," ASME DETC2012-71176.

[11] Whitefoot, K. S., and Skerlos, S. J., 2012, "Design Incentives to Increase Vehicle Size Created From the U.S. Footprint-Based Fuel Economy Standards," Energy Policy, **41**(1), pp. 402–411.

[12] Resende, C. B., Grace Heckmann, C., and Michalek, J. J., 2012, "Robust Design for Profit Maximization With Aversion to Downside Risk From Parametric Uncertainty in Consumer Choice Models," ASME J. Mech. Des., **134**(10), p. 100901.

[13] Payne, J. W., 1976, "Task Complexity and Contingent Processing in Decision Making: An Information Search," Organ. Behav. Hum. Perform., **16**(2), pp. 366–387.

[14] MacDonald, E. F., Gonzalez, R., and Papalambros, P. Y., 2009, "Preference Inconsistency in Multidisciplinary Design Decision Making," ASME J. Mech. Des., **131**(3), p. 031009.

[15] Morrow, W. R., Long, M., and MacDonald, E. F., 2014, "Market-System Design Optimization With Consider-Then-Choose Models," ASME J. Mech. Des., **136**(3). p. 031003.

[16] Knittel, C. R., and Metaxoglou, K., 2012, "Estimation of Random Coefficient Demand Models: Two Empiricists' Perspective," Rev. Econ. Stat., **96**(1), pp. 34–59.

[17] Allcott, H., and Wozny, N., 2012, "Gasoline Prices, Fuel Economy, and the Energy Paradox," Rev. Econ. Stat. (to be published).

[18] Dubé, J.-P., Fox, J. T., and Su, C.-L., 2012, "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," Econometrica, **80**(5), pp. 2231–2267.

[19] Beresteanu, A., and Li, S., 2011, "Gasoline Prices, Government Support, and the Demand for Hybrid Vehicles in the United States," Int. Econ. Rev. (Philadelphia), **52**(1), pp. 161–182.

[20] Copeland, A., Dunn, W., and Hall, G., 2011, "Inventories and the Automobile Market," RAND J. Econ., **42**(1), pp. 121–149.

[21] Hoyle, C., Chen, W., Wang, N., and Gomez-Levi, G., 2011, "Understanding and Modelling Heterogeneity of Human Preferences for Engineering Design," ASME J. Eng. Des., **22**(8), pp. 583–601.

[22] Li, S., Liu, Y., and Zhang, J., 2011, "Lose Some, Save Some: Obesity, Automobile Demand, and Gasoline Consumption in the United States," J. Environ. Econ. Manage., **61**(1), pp. 52–66.

[23] Musti, S., and Kockelman, K. M., 2011, "Evolution of the Household Vehicle Fleet: Anticipating Fleet Composition, PHEV Adoption and GHG Emissions in Austin, Texas," Transp. Res. Part A Policy Pract., **45**(8), pp. 707–720.

[24] Zhang, T., Gensler, S., and Garcia, R., 2011, "A Study of the Diffusion of Alternative Fuel Vehicles: An Agent-Based Modeling Approach," J. Prod. Innov. Manag., **28**(2), pp. 152–168.

[25] Axsen, J., Mountain, D. C., and Jaccard, M., 2009, "Combining Stated and Revealed Choice Research to Simulate the Neighbor Effect: The Case of Hybrid-Electric Vehicles," Resour. Energy Econ., **31**(3), pp. 221–238.

[26] Dagsvik, J. K., and Liu, G., 2009, "A Framework fo Analyzing Rank-Ordered Data With Application to Automobile Demand," Transp. Res. Part A Gen., **43**(1), pp. 1–12.

[27] Hoyle, C., Chen, W., Ankenman, B., and Wang, N., 2009, "Optimal Experimental Design of Human Appraisals for Modeling Consumer Preferences in Engineering Design," ASME J. Mech. Des., **131**(7), p. 071008.

[28] Shiau, C.-S. N., and Michalek, J. J., 2009, "Should Designers Worry About Market Systems?," ASME J. Mech. Des., **131**(1), p. 011011.

[29] Vance, C., and Mehlin, M., 2009, *Tax Policy and CO_2 Emissions—An Econometric Analysis of the German Automobile Market*, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Ruhr-Universität Bochum, Universität Dortmund, Universität Duisburg-Essen, Bochum, Dortmund, Duisburg, Essen, Germany.

[30] Dasgupta, S., Siddarth, S., and Silva-Risso, J., 2007, "To Lease or to Buy? A Structural Model of a Consumer's Vehicle and Contract Choice Decisions," J. Mark. Res., **44**(3), pp. 490–502.

[31] Train, K., and Winston, C., 2007, "Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers," Int. Econ. Rev. (Philadelphia), **48**(4), pp. 1469–1497.

[32] Kemper, L., 2006, "Multilevel Optimization for Enterprise-Driven Decision-Based Product Design," *Decision Making in Engineering*, L. Kemper, W. Chen, and L. Schmidt, eds., ASME Press, NY, pp. 203–214.

[33] Berry, S., Levinsohn, J., and Pakes, A., 2004, "Differentiated Products Demand Systems From a Combination of Micro and Macro Data: The New Car Market," J. Polit. Econ., **112**(1), pp. 68–105.

[34] Greene, D. L., Patterson, P. D., Singh, M., and Li, J., 2005, "Feebates, Rebates and Gas-Guzzler Taxes: A Study of Incentives for Increased Fuel Economy," Energy Policy, **33**(6), pp. 757–775.

[35] Michalek, J. J., Papalambros, P. Y., and Skerlos, S. J., 2004, "A Study of Fuel Efficiency and Emission Policy Impact on Optimal Vehicle Design Decisions," ASME J. Mech. Des., **126**(6), pp. 1062–1070.

[36] Petrin, A., 2002, "Quantifying the Benefits of New Products: The Case of the Minivan," J. Polit. Econ., **110**(4), pp. 705–729.

[37] Sudhir, K., 2001, "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," Mark. Sci., **20**(3), pp. 42–60.

[38] Berry, S., Levinsohn, J., and Pakes, A., 1999, "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," Am. Econ. Rev., **89**(3), pp. 400–430.

[39] Brownstone, D., and Train, K., 1999, "Forecasting New Product Penetration With Flexible Substitution Patterns," J. Econ., **89**(1–2), pp. 109–129.

[40] Dagsvik, J. K., Wennemo, T., Wetterwald, D. G., and Aaberge, R., 2002, "Potential Demand for Alternative Fuel Vehicles," Transp. Res. Part B: Methodol., **36**(4), pp. 361–384.

[41] McCarthy, P. S., 1996, "Market Price and Income Elasticities of New Vehicle Demands," Rev. Econ. Stat., **78**(3), pp. 543–547.

[42] Berry, S., Levinsohn, J., and Pakes, A., 1995, "Automobile Prices in Market Equilibrium," Econometrica, **63**(4), pp. 841–890.

[43] Boyd, J. H., and Mellman, R. E., 1980, "The Effect of Fuel Economy Standards on the US Automotive Market: A Hedonic Demand Analysis," Transp. Res. Part A Gen., **14**, pp. 367–378.

[44] Lave, C., and Train, K., 1979, "A Disaggregate Model of Auto-Type Choice," Transp. Res. Part A Gen., **13**(1), pp. 1–9.

[45] Feit, E. M., Beltramo, M. A., and Feinberg, F. M., 2010, "Reality Check: Combining Choice Experiments with Market Data to Estimate the Importance of Product Attributes," Manage. Sci., **56**(5), pp. 785–800.

[46] Bunch, D., Greene, D., Lipman, T., and Shaheen, S., 2011, *Potential Design, Implementation, and Benefits of a Feebate Program for New Passenger Vehicles in California*, University of California, Davis, Institute of Transportation Studies, Davis, CA.

[5]We reject the null hypothesis that the coefficient is equal to zero at the $\alpha = 0.01$ level for a two-sided t-test.

[47] U.S. Energy Information Administration, 2011, *Annual Energy Outlook 2011 With Projections to 2035*, U.S. Energy Information Administration, Washington, DC.
[48] Gigerenzer, G., and Brighton, H., 2009, "Homo Heuristicus: Why Biased Minds Make Better Inferences," Top. Cogn. Sci., **1**(1), pp. 107–143.
[49] Train, K., 1979, "Consumers' Responses to Fuel-Efficient Vehicles: A Critical Review of Econometric Studies," Transportation (Amsterdam), **8**(3), pp. 237–258.
[50] Chen, C., 2001, "Design for the Environment: A Quality-Based Model for Green Product Development," Manage. Sci., **47**(2), pp. 250–263.
[51] Shiau, C.-S. N., Michalek, J. J., and Hendrickson, C. T., 2009, "A Structural Analysis of Vehicle Design Responses to Corporate Average Fuel Economy Policy," Transp. Res. Part A: Policy Pract., **43**(9–10), pp. 814–828.
[52] Nevo A., 2000, "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," J. Econ. Manage. Strategies, **9**(4), pp. 513–548.
[53] Dubé, J., Fox, J., and Su, C., 2012, "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," Econometrica **80**(5), pp. 2231–2267.
[54] Skrainka, B., and Judd, K., 2011, "High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation," (published online). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1870703
[55] Bunch, D. S., Bradley, M., and Golob, T. F., 1993, "Demand for Clean Fuel Vehicles in California A Discrete Choice Stated Preference Pilot Project," Transp. Res. Part A: Policy Pract., **27A**(3), pp. 237–253.
[56] Brownstone, D., Bunch, D. S., and Train, K., 2000, "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles," Transp. Res. Part B: Methodol., **34**(5), pp. 315–338.
[57] McFadden, D., and Train, K., 2000, "Mixed MNL Models for Discrete Response," J. Appl. Econ., **15**(5), pp. 447–470.
[58] Selby, B., and Kockelman, K., 2012, "Microsimulating Automobile Markets: Evolution of Vehicle Holdings and Vehicle Pricing Dynamics," Transp. Res. Forum, **51**(2), pp. 83–96.
[59] Ward's Automotive Group, 2008, "Ward's Yearly Vehicle Specifications." Available at: http://wardsauto.com/data-center
[60] Polk, 2008, "IHS Polk Insight Yearly Volume Data." Available at: http://blog.polk.com/blog/new-vehicle-sales
[61] U.S. Energy Information Administration, 2013, "U.S. All Grades All Formulations Retail Gasoline Prices" (accessed: Aug. 1, 2013) http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0_PTE_NUS_DPG&f=A
[62] "Annual Average Indexex," Bur. Labor Statitstics Data Online, (accessed: Aug. 1, 2013), http://www.bls.gov/cpi/cpi_dr.htm#2007
[63] Wooldridge, J. M., 2006, *Introductory Econometrics: A Modern Approach*, Thomas Higher Education, Mason, OH.
[64] Kullback, S., and Leibler, R. A., 1951, "On Information and Sufficiency," Ann. Math. Stat., **22**(1), pp. 79–86.
[65] Akaike, H., 1973, "Information Theory and an Extension of the Maximum Likelihood Principle," Proceedings of the Second International Symposium on Information Theory, B. N. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, Hungary, pp. 267–281.
[66] Schwarz, G., 1978, "Estimating the Dimension of a Model," Ann. Stat., **6**(2), pp. 461–464.
[67] Burnham, K. P., and Anderson, D. R., 2002, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, Springer, New York.
[68] Claeskens, G., and Hjort, N., 2008, *Model Selection and Model Averaging*, Cambridge University Press, NY.
[69] Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N., 2009, "The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity," Mark. Sci., **29**(3), pp. 393–421.
[70] Heffner, R. R., Kurani, K. S., and Turrentine, T. S., 2007, "Symbolism in California's Early Market for Hybrid Electric Vehicles," Transp. Res. Part D: Transp. Environ., **12**(6), pp. 396–413.
[71] Axsen, J., and Kurani, K. S., 2011, "Interpersonal Influence in the Early Plug-in Hybrid Market: Observing Social Interactions With an Exploratory Multi-Method Approach," Transp. Res. Part D: Transp. Environ., **16**(2), pp. 150–159.
[72] Hauser, J., and Wernerfelt, B., 1990, "An Evaluation Cost Model of Consideration Sets," J. Consum. Res., **16**(4), pp. 393–408.
[73] Gilbride, T. J., and Allenby, G. M., 2004, "A Choice Model With Conjunctive, Disjunctive, and Compensatory Screening Rules," Mark. Sci., **23**(3), pp. 391–406.
[74] Lucas, R. E., 1976, "Econometric Policy Evaluation: A Critique," Carnegie-Rochester Conference Series on Public Policy, **1**, pp. 19–46.